

Geomagnetic D_{st} index forecast based on IMF data only

G. Pallochia, E. Amata, G. Consolini, M. F. Marcucci, and I. Bertello

Istituto di Fisica dello Spazio Interplanetario, Istituto Nazionale di Astrofisica, Rome, Italy

Received: 21 July 2005 – Revised: 8 February 2006 – Accepted: 14 February 2006 – Published: 19 May 2006

Abstract. In the past years several operational D_{st} forecasting algorithms, based on both IMF and solar wind plasma parameters, have been developed and used. We describe an Artificial Neural Network (ANN) algorithm which calculates the D_{st} index on the basis of IMF data only and discuss its performance for several individual storms. Moreover, we briefly comment on the physical grounds which allow the D_{st} forecasting based on IMF only.

Keywords. Interplanetary physics (Interplanetary magnetic fields) – Magnetospheric physics (Solar wind-magnetosphere interactions)

1 Introduction

It has been known for decades that solar activity influences the near-Earth environment through the solar wind variable flow and energetic particles emissions. The description of such influences and the development of tools for their now-casting and forecasting is the subject of space weather. It is now widely accepted that space weather effects may damage critical equipment, such as communication satellites or power lines and pipelines on the ground, and disrupt HF communications and GPS links, etc. As such, the prediction of space weather effects has both scientific and economical reasons.

In the framework of space weather an important role is played by geomagnetic storms, which are comprised of processes occurring in near-Earth space. During geomagnetic storms very intense fluctuations of the horizontal component of the ground magnetic field are observed (Gonzalez et al., 1994), due to variations in the equatorial ring current. A measure of these variations is provided by the Disturbance Storm Time index (D_{st}), which is calculated on an hourly

basis from measurements made by a network of four ground magnetometer stations at low and middle latitudes.

After the pioneering paper by Gosling (1993), CMEs are now widely recognized as the dominant interplanetary phenomenon responsible for intense magnetic storms. In the past years, many studies have been devoted to the relation between D_{st} and solar wind conditions. Among them we recall Gonzalez et al. (1999), who studied extensively the storm time profile and intensity in relation with the solar wind structures associated with CMEs and Corotating Interaction Regions (CIRs), Kane (2005), who critically investigated the relationship of solar and interplanetary plasma parameters with geomagnetic storms, and Gonzalez and Echer (2005), who studied the relationship between peak D_{st} and peak negative B_z during intense geomagnetic storms; moreover, we recall the recent review by Yermolaev et al. (2005) and all references therein. A complementary approach to these themes has been taken by several authors who tried to forecast D_{st} from measurements of Interplanetary Magnetic Field (IMF) and solar wind plasma parameters. As a result, a number of empirical models has been developed, based on differential equations (see, e.g. Burton et al., 1975; Fenrich and Luhmann, 1998; O'Brien and McPherron, 2000; Temerin and Li, 2002; Wang et al., 2003), and on Artificial Neural Networks (see, e.g. Wu and Lundstedt, 1997; Lundstedt et al., 2002). All of them share the characteristics of having, as their inputs, both magnetic and plasma parameters of the solar wind, so that when, for some reason, some of the inputs are missing, the model predictions cannot be relied upon. This is more likely to occur with regard to the plasma parameters, as they are provided by instruments which can be affected by enhanced solar X-ray and energetic particle fluxes to a greater extent than magnetometers; moreover, at times the solar wind speed can exceed the upper instrumental limit.

Starting from the last consideration, we describe here a new algorithm, hereafter called EDDA (Empirical D_{st} Data

Correspondence to: G. Pallochia
(giuseppe.pallochia@ifsi-roma.inaf.it)

Algorithm), based on an Artificial Neural Network (ANN), which computes D_{st} from IMF data only. The aim of this work was not to obtain the best possible algorithm for the D_{st} prediction, but to build an operational service which can reliably forecast D_{st} on the basis of IMF data only.

2 D_{st} forecasting through Artificial Neural Networks

2.1 Why choose Artificial Neural Networks?

The solar wind shapes the magnetosphere and transfers to it mass, energy and momentum through various processes at the magnetopause boundary. In the magnetosphere itself various processes occur on different temporal and spatial scales, involving particle acceleration, magnetic reconnection, particle injection along magnetic field lines, wave particle interactions and so forth. Many efforts have been made over the years to model the magnetospheric response to solar wind variations, but the problem is far from being solved. Several studies have addressed single aspects of the solar wind-magnetosphere interaction. Such is the problem of predicting geomagnetic storms through D_{st} forecasting. In this respect, a preliminary consideration to be made is that the magnetosphere may react in different ways depending on its current state, which, in turn, may be determined by the recent “history” of its interaction with the solar wind. In particular this is true for the D_{st} index, which can react with a delay of a few hours to the solar wind stimuli.

It has been suggested by various authors in recent works that the magnetospheric response to solar wind changes is highly organised and complex in nature (e.g. Klimas et al., 1996; Consolini and Chang, 2001). This is evidence of nonlinear dynamics related to the energy storage, transport and release, and of the inherent out-of-equilibrium configuration of the magnetosphere-ionosphere system. The modeling of such a complex and nonlinear dynamics could benefit from the use of new approaches. This is the case of Artificial Neural Networks (ANN), which can capture the hidden parallel interactions of an input-output system and forecast its response on the basis of the input only. Indeed, in the last years the ANN technique has been extensively used (see Lundstedt et al., 2002, and references therein). In the following we shall make use of this same technique.

2.2 The ANN architecture

The aforementioned considerations suggest the use of an ANN architecture which includes some “memory” of the system evolution. This can be accomplished in different ways. A possible solution is to make use of a perceptron (i.e. a pure feedforward network) and to feed it with the input parameters at the current time t and at N preceding times: $t-1, \dots, t-N$. However, this approach has two practical drawbacks: 1) the resulting perceptron is hardly scalable; 2) it is difficult to determine the correct N , namely, the optimal correlation time

between each input and D_{st} . A different solution to the problem is to use a so-called Elman network, i.e. a two-layer recurrent network where the output of each neuron in the hidden layer is replicated as an additional input, called a context unit (Elman, 1990). Thus, at time t the context units contain information coming from the network state at time $(t-1)$ and set a context for processing at time t , so that the state of the whole network at a particular time depends on an aggregate of the previous states, as well as on the current inputs.

We made several preliminary tests using both Elman and perceptron networks and decided eventually to use an Elman network. The reasons for this choice have already been fully described by Wu and Lundstedt (1997) and Lundstedt et al. (2002). Wu and Lundstedt (1997) made detailed comparisons of the results obtained for several Elman networks; in particular, they discussed in detail networks with different numbers of hidden layers, all based on the following four inputs: IMF intensity, B , solar wind density, n , and speed, V , and the B_s parameter (defined through B_z , the IMF GSM z component, as follows: $B_s = |B_z|$, for $B_z < 0$; $B_s = 0$, otherwise). For such networks they quoted correlation coefficients between predicted and observed D_{st} ranging from 0.87 for one hidden neuron, to 0.89 for 4 hidden neurons and to 0.90 for 24 hidden neurons, and RMSE ranging from 17.1 nT, to 16.5 nT and to 15.7 nT, respectively. We also made several similar tests and concluded that a fair trade-off between the network complexity and its performance is to use an Elman network with one hidden layer containing four neurons, as already done by Lundstedt et al. (2002), whose algorithm was proven to perform better than three algorithms based on differential equations (Burton et al., 1975; Fenrich and Luhmann, 1998; O’Brien and McPherron, 2000), all based on both plasma and IMF data. Therefore, in the following, we will use the Lundstedt et al. (2002) algorithm for comparative purposes and we will call it “Lund” for short. Our algorithm differs from the “Lund” one in the choice of the input parameters, as described hereafter.

2.3 The choice to develop an IMF-based network

To our knowledge, the inputs of all published algorithms for D_{st} forecasting comprise of both solar wind plasma and IMF data (see Lundstedt et al., 2002, and Temerin and Li, 2002, and all references therein). We take a different point of view and note that the dependence of past models on plasma parameters poses a serious operational constraint. The reason is that plasma instruments fail more often than magnetometers, and the expected operational life is far longer for magnetometers than for plasma instruments. To illustrate the different performance of magnetometers and plasma instruments, we considered an extended time interval for which verified data were available for both ACE and SOHO: 3 February 1999, 22:00 UT–11 July 2005, 23:00 UT for a total of 56 402 h. For this period we examined hourly averages of ACE IMF (from the MAG instrument)

Table 1. ACE/SWEPAM failures.

Start (Year Doy h)	Stop (Year Doy h)	D_{st} min (nT)
2000 194 12	2000 198 00	-301
2000 314 15	2000 315 21	-096
2001 268 20	2001 269 15	-102
2001 310 01	2001 311 01	-292
2001 328 05	2001 328 16	-221
2003 301 13	2003 303 23	-401
2003 307 03	2003 307 16	-036
2005 015 14	2005 018 21	-121
2005 020 07	2005 021 03	-051

and solar wind dynamic pressure (from the SWEPAM instrument) and of SOHO solar wind dynamic pressure (from the CELIAS/MTOF proton monitor – the SOHO payload does not include a magnetometer) looking for data gaps. The result of this analysis is the following:

- a) ACE IMF: 17 missing hourly averages;
- b) ACE solar wind pressure: 6804 missing hourly averages;
- c) SOHO solar wind pressure: 1858 missing hourly averages.

In 194 cases the SOHO and ACE missing hourly averages occurred at the same time. On these grounds it is reasonable to assume that an operational service which forecasts D_{st} on the basis of ACE data only will probably have a failure rate of 12 percent, caused by the missing solar wind data. This can be mitigated by using SOHO solar wind data and ACE IMF data, as in this case the failure rate would be 3.3 percent. The effect on the service of the plasma data gaps could be further mitigated by using either SOHO or ACE data, depending on their availability. However, although this can easily be done for post-event analysis, it cannot be simply and reliably implemented for the real time operation of the service. Further considerations can be made with regard to the plasma data gaps. The point is that such gaps often occur in conjunction with large emissions of particles and radiation from the Sun, as the instruments devoted to the measurement of solar wind plasma parameters can saturate in these occasions for hours or even days. Table 1 displays a list of nine extended periods of missing ACE plasma parameters which occurred in the above defined time interval. For each event, the first six columns from the left give the start and stop times in year, day, hour, while the last column on the right displays the minimum value of the Kyoto D_{st} recorded during the given time period. All such data gaps occurred after a solar flare and in seven out of nine cases were followed by a geomagnetic storm. This list includes the well-known storms of October 2003, which will be further discussed in Sect. 3.4, and the well-known Bastille event, i.e. 14 July 2000, which we use here to illustrate the relation of SEP (Solar Ener-

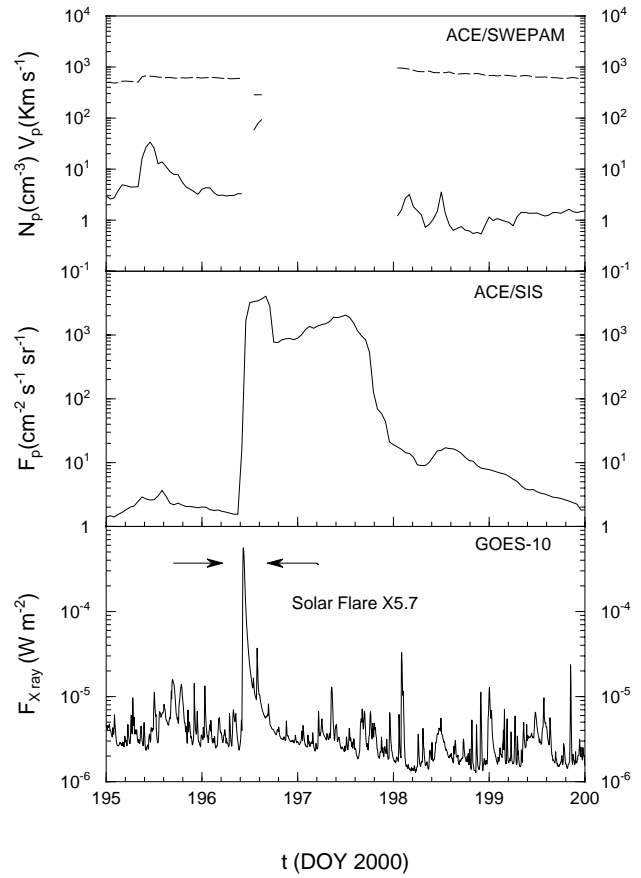


Fig. 1. Upper panel: ACE/SWEPAM proton density (solid line) and bulk speed (dashed line) hourly averages. Middle panel: ACE/SIS integral proton flux ($E > 10$ MeV) hourly averages. Lower panel: GOES-10 X ray flux five-minute averages ($1.0 < \lambda < 8.0$ Å wavelength band).

getic Protons) and solar X ray fluxes with such data gaps. Figure 1 shows, from top to bottom, hourly averages of the ACE/SWEPAM solar wind verified speed and number density, the ACE/SIS integral proton flux for $E > 10$ MeV, the GOES-10 X ray flux in the $1.0 < \lambda < 8.0$ Å wavelength band, from day 195 to day 200, 2000. As this event has already been the subject of numerous papers (e.g. Watari et al., 2001), we do not describe it in detail. Instead, it is enough to notice, for our purposes, that the ACE/SWEPAM data display a long data gap between 12:00 UT on day 196 and 00:00 UT on day 198. Exactly in that time interval, the proton flux at ACE/SIS is higher by a 10^2 – 10^3 factor with respect to the level displayed before and after the event. The increase in the proton flux occurs very sharply at 12:00 UT on day 196, in coincidence with a large peak of the X-ray flux (probably related to an X5.7 flare on the Sun) a 10^2 factor larger than the background flux. No gaps were observed during the same time for the ACE/MAG magnetometer data (not shown in the figure). From this discussion we conclude that indeed

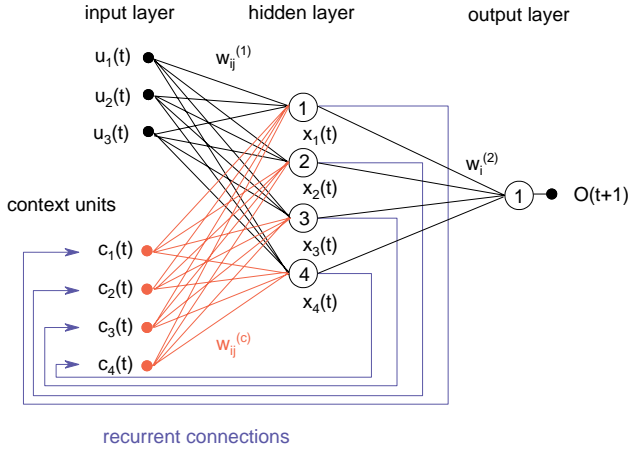


Fig. 2. The Elman scheme used for EDDA. u_1 , u_2 and u_3 are the inputs (normalized B_z , B^2 and B_y^2); w 's are the network weights. The blue lines indicate copying of each hidden layer output x into the corresponding context unit c , so that $c_k(t) = x_k(t-1)$ (see details in text).

magnetometers are more efficient than plasma instruments in producing continuous time series of data at L1, which are essential for running an operational service devoted to D_{st} forecasting. On these grounds, it is reasonable to verify the feasibility of a D_{st} forecasting algorithm based on IMF only.

3 The EDDA Model

3.1 The selected network

On the basis of the considerations made in the preceding section, we trained and tested several networks with different combinations of IMF components. In conclusion, we selected for the development of the EDDA algorithm an Elman network (Elman, 1990) with the following structure (Fig. 2):

1. 3 input lines: B_z , B^2 , B_y^2 ,
2. 4 context units,
3. 1 hidden layer with 4 neurons,
4. 1 linear output neuron.

In the discussion section we will comment on the significance of the input choice. The input parameters are hourly averages calculated from L1 IMF data in GSM coordinates. Before being fed into the network, the inputs were normalised (see the Appendix for details). The output of the i -th hidden layer neuron is:

$$x_i(t) = \tanh\left(\sum_{j=1}^3 w_{ij}^{(1)} u_j(t) + \sum_{k=1}^4 w_{ik}^{(c)} c_k(t)\right), \quad (1)$$

Table 2. The EDDA training set.

SC	Start (Year Doy h)	End (Year Doy h)
WIND	1995 080 03	1995 109 06
WIND	1995 263 11	1995 305 02
WIND	1997 094 23	1997 119 22
WIND	1997 316 05	1998 001 03
ACE	2000 141 15	2000 158 06
ACE	2000 299 03	2000 324 02
ACE	2001 083 23	2001 125 10
ACE	2001 292 17	2001 307 01

where t is the time in integer hours, the first sum is made over the three normalized external inputs $u_j(t)$, and the second one is made over the four context units of content $c_k(t)$. $w_{ij}^{(1)}$ and $w_{ik}^{(c)}$ are the weights of the connections between the i -th hidden layer neuron and, respectively, the j -th input and the k -th context unit. The context units are defined as $c_k(t) = x_k(t-1)$; in other words, each context unit is connected to the corresponding hidden layer neuron by a recurrent non trainable connection, whose weight is set constantly to 1, and acts as a memory bank, by receiving at a given time t , the output of the corresponding hidden layer neuron at the preceding time $t-1$, i.e. one hour earlier. The normalized output of the network is obtained from a linear combination of the 4 hidden layer outputs:

$$O(t+1) = \sum_{i=1}^4 w_i^{(2)} x_i(t), \quad (2)$$

where $w_i^{(2)}$ is the weight of the connection between the output unit and the i -th hidden neuron. The output O is assigned the time $t+1$ for inputs at the time t , to account for the 1-h average travel time of the solar wind from L1 to the Earth's magnetopause.

3.2 The training

The database for developing the EDDA algorithm consists of WIND and ACE IMF hourly averages and of final Kyoto D_{st} values from 1995 to 2002 (after 2002 final D_{st} values are not available yet), amounting overall to $\sim 64\,000$ h. The IMF data were collected close to the L1 libration point. From such a database a training set was built, comprising of about 6000 hourly averages from the periods listed in Table 2. The network weights \mathbf{w} were then determined by minimizing the cost function:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^{N_{tr}} (T^{(i)} - O^{(i)})^2, \quad (3)$$

where N_{tr} is the number of data points in the training set, $T^{(i)}$ is the i -th observed Kyoto D_{st} (i.e. the “right answer”)

and $O^{(i)}$ the corresponding network output. The minimization of E is performed in two steps. Firstly, the weights, initially set as small random values, are modified for a prefixed number of iterations by the error backpropagation method (Rumelhart et al., 1986). Then they are perturbed, independently from one another, by adding a random number extracted from a normal distribution. If this decreases the cost function, the perturbed weights become the new network coefficients and the process is repeated until the E decrease rate drops below a given threshold. This second step can be viewed as a random walk on the hyper-surface E in the space of network weights, where each step is constrained downhill. As the number of local minima of the cost function is not “a priori” known, it is useful to repeat the training many times, in order to try to explore different attraction basins and hopefully to determine the best local minimum. This hybrid scheme (a sort of Monte Carlo method) frees the training phase from the boring research of the best values for the parameters in the backpropagation (i.e. the learning rate and momentum). Several distinct algorithms were trained and some were retained for further testing, all having very similar performances over the training set.

3.3 The testing

A test set was built by merging together all the periods contained in the 1995–2002 database and not previously used during the training. The selected algorithms were run over the whole set and, for each of them, the forecast D_{st} s values were binned in 25 nT intervals according to the corresponding Kyoto D_{st} values. On this basis, for each algorithm and for each bin, a percent root mean square error was calculated as

$$R^{(n)} = 100 \frac{\sqrt{\frac{1}{N_{\text{bin}}^{(n)}} \sum_{j=1}^{N_{\text{bin}}^{(n)}} (T^{(j)} - O^{(j)})^2}}{T_c^{(n)}}, \quad (4)$$

where n specifies the 25 nT interval, $T_c^{(n)}$ is the corresponding central D_{st} value, $T^{(j)}$ and $O^{(j)}$ are the j -th Kyoto D_{st} and algorithm output, respectively, $N_{\text{bin}}^{(n)}$ is the number of data points of the test set whose $T^{(n)}$ falls in the n -th interval. The R parameter provides a quantitative tool to assess the algorithm performance for different geomagnetic conditions. To this regard, we use the following storm classification: “small” for $-50 \text{ nT} < D_{st} < -30 \text{ nT}$, “moderate” for $-100 \text{ nT} < D_{st} < -50 \text{ nT}$, “intense” for $-200 \text{ nT} < D_{st} < -100 \text{ nT}$, “severe” for $D_{st} < -200 \text{ nT}$. The algorithm having the smallest R values for $T < -25 \text{ nT}$ was chosen as the final EDDA algorithm. For the selected algorithm Fig. 3 displays R and N_{bin} as a function of T_c , in the upper and lower panels respectively. The N_{bin} histogram approximates the distribution of the Kyoto D_{st} for the test set: 94 per cent of the points are concentrated in the $-50 \text{ nT} < T_c < 25 \text{ nT}$ interval, but an extended asymmetrical

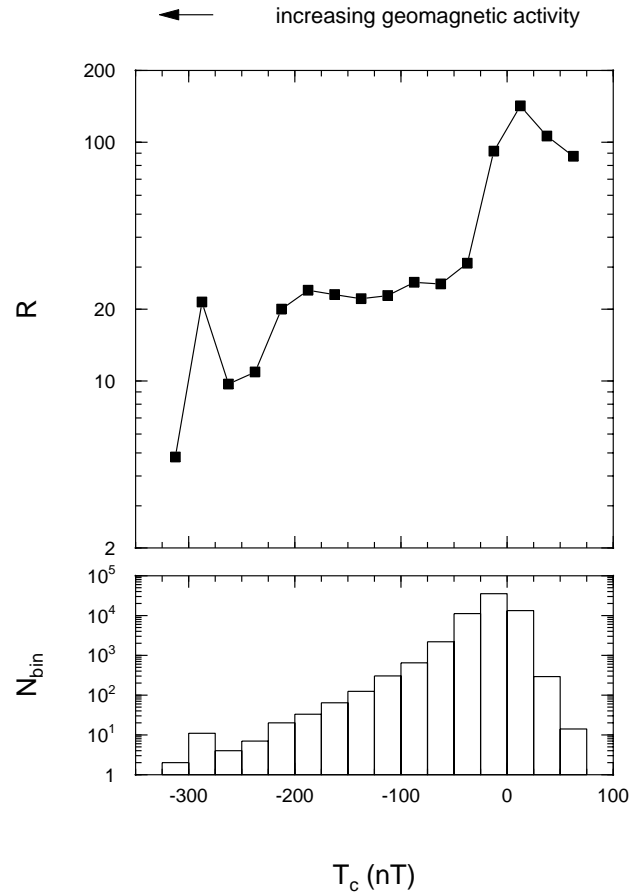


Fig. 3. Upper panel: relative error R for the EDDA algorithm, as defined in Eq. (4), calculated over the whole test set, and plotted against T_c (see details in text). Lower panel: number of points over which each R is calculated.

tail extends to negative D_{st} values. N_{bin} drops below 10 for $T_c < -225 \text{ nT}$, which implies that the last four R values on the left have little statistical relevance.

Let us first examine the behaviour of R for quiet time values of D_{st} . We see that R has a maximum of 142 in the $(0, 25 \text{ nT})$ bin R , and is close to 100 in the two nearby bins and in the $(50, 75 \text{ nT})$ bin. As far as we can guess, such large relative errors can result from two different contributions: errors in the D_{st} baseline, and a failure to reproduce the initial compression prior to the storm main phase (see also the Discussion section). Moving towards more negative D_{st} values, we see that R drops to about 30 in the $(-50 \text{ nT}, -25 \text{ nT})$ bin, roughly corresponding to “small” storms; and is fairly constant, between 20 and 25, for $-225 \text{ nT} < T_c < -50 \text{ nT}$, i.e. for “moderate” and “intense” storms. For $T_c < -225 \text{ nT}$, i.e. for “severe” storms, R decreases further, but, as N_{bin} drops below 10, such data points have a reduced statistical relevance.

The data we have just described can be used to compare the EDDA performance with that of the “Lund” algorithm. For that purpose, we calculated the “Lund” R from

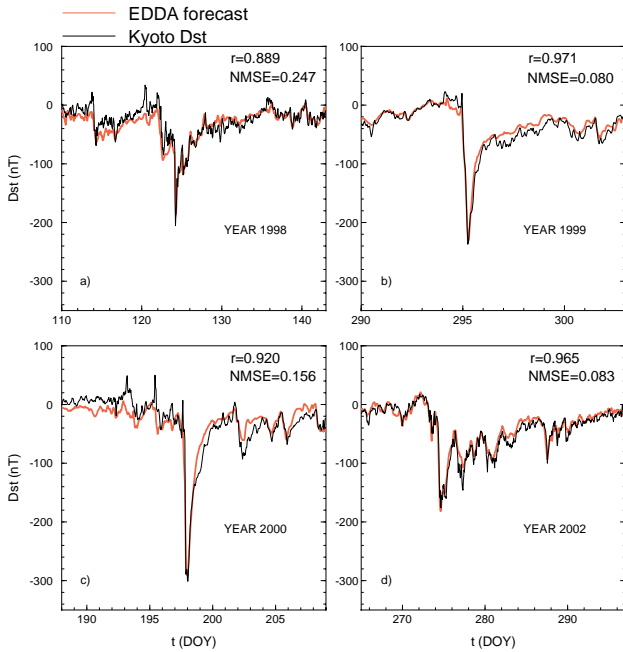


Fig. 4. Comparison of the EDDA D_{st} (red lines) with the Kyoto D_{st} (black lines) for geomagnetically disturbed periods. r is the linear correlation coefficient and NMSE the normalised mean square error (see text for details).

the data displayed in Fig. 3 of Lundstedt et al. (2002). For $-25 \text{ nT} < T_c < 25 \text{ nT}$ we found smaller values, in the order of 83. For $-50 \text{ nT} < T_c < -25 \text{ nT}$, i.e. for “small” storms, we obtained 32, similar to the EDDA value. Finally, for $-125 \text{ nT} < T_c < -50 \text{ nT}$, i.e. for “moderate” storms, we found values between 22 and 25, again similar to the EDDA values. This comparison is based on different data sets for EDDA and “Lund”: as we described earlier, the EDDA test set comprises of about 58 000 hourly averages, while the data set used for Fig. 3 of Lundstedt et al. (2002) comprises of 40 000 hourly averages. We do not know to what extent the two data sets overlap. As regards to “intense” and “severe” storms, we cannot use R to compare the two algorithms, because: a) Fig. 3 of Lundstedt et al. (2002) stops at $T_c = -125 \text{ nT}$; b) we cannot build an extended common test set for both EDDA and “Lund” in the 1995–2002 period, as we do not know the exact training periods used for the “Lund” algorithm. Therefore, for $D_{st} < -125 \text{ nT}$, we have to use more recent data, from 2003 onwards. Unfortunately, after 2002 final Kyoto D_{st} values are not available yet and it is not reasonable to base a statistical comparison on provisional D_{st} data. Therefore, for such periods, we will only compare the two algorithms for some individual storms (see Sect. 3.4).

Before closing this discussion on the EDDA testing, we notice that in addition to R , it is interesting to consider the EDDA linear correlation coefficient, which is 0.83, and the

total root mean square error

$$RMSE = \sqrt{\frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} (T^{(j)} - O^{(j)})^2}, \quad (5)$$

which amounts to 13.9 nT over the whole test set (N_{test} is the number of test set data points). These values can be directly compared with those quoted by Wu and Lundstedt (1997) for a similar network with the four inputs B , n , V and B_s : 0.89 and 16.5. Although the training and testing sets are rather different, also this comparison suggests that the performances of the two algorithms are comparable. This and the preceding discussion on the R parameter suggest that, when both plasma and IMF data are available, EDDA produces forecasts at least comparable to those of other “normal” algorithms and, as such, provides a reliable operational forecasting tool.

To show typical examples of the performance of the EDDA algorithm for individual large storms, Fig. 4 compares its forecasted D_{st} with the Kyoto D_{st} index for four cases of high geomagnetic activity, all pertaining to the test set, i.e. not used in the network training. In each panel the linear correlation coefficient r between forecast and observed D_{st} , and the normalised mean square error are reported. The latter is defined as

$$NMSE = \frac{\frac{1}{N_{ev}} \sum_{j=1}^{N_{ev}} (T^{(j)} - O^{(j)})^2}{Var_{ev}}, \quad (6)$$

where, for each event, N_{ev} is the number of points and Var_{ev} the variance of the Kyoto D_{st} . In the top left hand panel D_{st} displays a decrease by $\sim 60 \text{ nT}$ on day 113, 1998, preceded by an increase by $\sim 20 \text{ nT}$, lasting on the order of 12 h; two further such increases occur on day 121 and on day 122, by $\sim 40 \text{ nT}$ and $\sim 20 \text{ nT}$, respectively; immediately after that, a negative sharp jump, by $\sim 100 \text{ nT}$, occurs on day 122 over a few hours. This is followed by an increase lasting about two days by $\sim 60 \text{ nT}$, followed by the main storm dip to a minimum value of $\sim -210 \text{ nT}$ on day 124. The recovery phase lasts about 4 days. The EDDA D_{st} follows closely the Kyoto D_{st} , both during the dips and in the recovery phase; we notice, however, some minor disagreements, on the order of $\sim 10\text{--}20 \text{ nT}$ and the fact that EDDA fails to reproduce the four short-lived increases on days 113, 121 and 122. Such increases can be interpreted as due to solar wind compressions prior to the storm main phase. In the top right-hand panel D_{st} decreases to $\sim -240 \text{ nT}$ on day 295, 1999, to recover over 4 days, while in the bottom left-hand panel D_{st} displays several peaks before a decrease to $\sim -300 \text{ nT}$ on day 198, 2000, and a recovery lasting 2.5 days. In both cases the D_{st} behaviour is reproduced very well by EDDA, with an exception made again for the short-lived increases just prior to the storm main phase. In the last panel, the Kyoto D_{st} displays a broad peak by $\sim 40 \text{ nT}$, lasting from day 270 through day 273, 2002, a decrease to $\sim -50 \text{ nT}$, a short-lived peak by $\sim 40 \text{ nT}$, a main

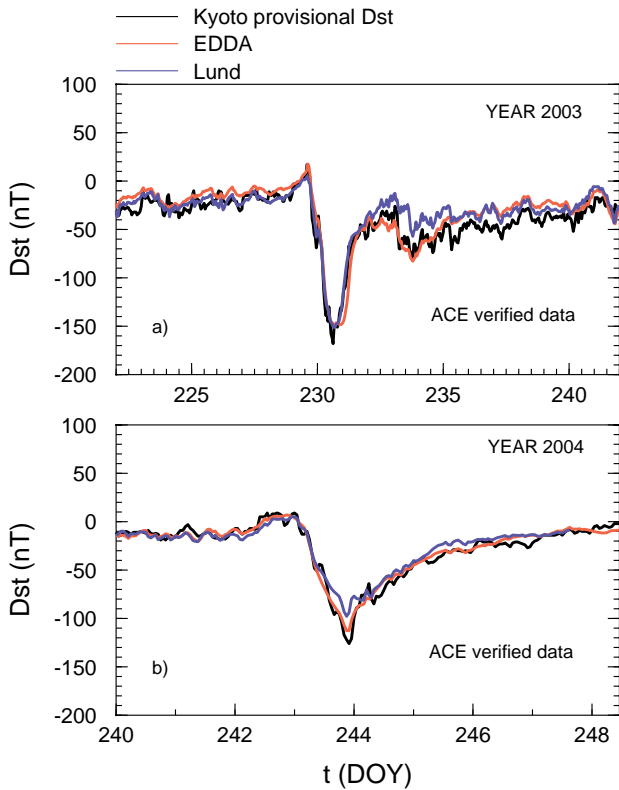


Fig. 5. Comparison of the EDDA (red line) and “Lund” D_{st} (blue line) with the Kyoto D_{st} (black line) for two “intense” storms.

dip to ~ -195 nT on day 274, 2002, and various oscillations during a 10–15 day recovery phase. EDDA reproduces very well all the D_{st} variations, including the compression prior to the main phase, with an exception made for a difference of ~ 20 nT observed during day 283 in the recovery phase. As a final remark, we notice that the four cases we have discussed are all rather different from each other, as regarding the strength of the storms and their time scales. In the discussion section we will comment on the compressions prior to the main phase and on the deviations of the order of 20 nT often observed when $D_{st} > -50$ nT.

In the Appendix we list all the EDDA weights and normalisation factors.

3.4 EDDA performance for recent storms

We now apply the EDDA and “Lund” algorithms to more recent data, from 2003 onwards, pertaining neither to the training set nor to the test set of either algorithm. We recall that the comparison is made with the provisional Kyoto D_{st} as the final D_{st} is not available yet.

Figure 5 shows two “intense” storms: that of day 230, 2003, in the top panel, and that of day 243, 2004, in the bottom panel. In both cases the ACE/SWEPAM data do not display any gap, so that the “Lund” algorithm could be

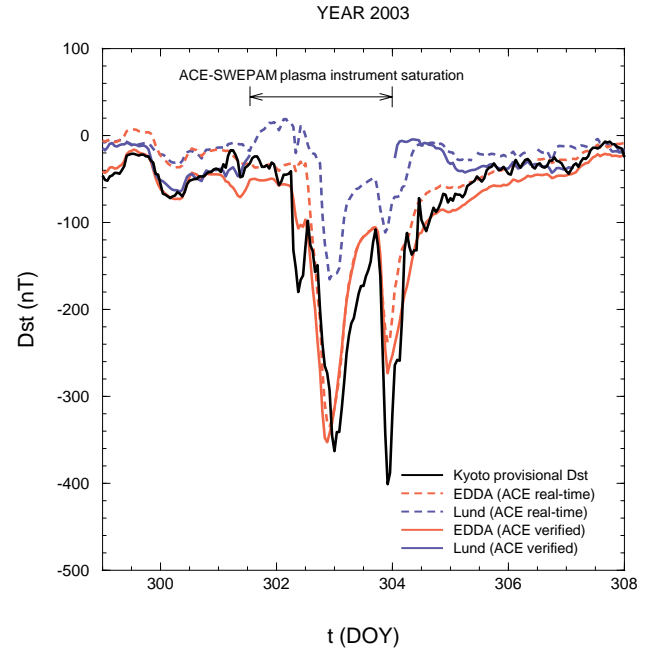


Fig. 6. Performance of the EDDA and “Lund models” (inputs: B_z , N_p , V_p) for two storms in October 2003 when the ACE plasma instrument malfunctioned (see details in the text).

run without problems. The Kyoto index is plotted in black, while the EDDA and “Lund” indices are plotted in red and blue, respectively. We see that both EDDA and “Lund” reproduce rather well the Kyoto D_{st} for both storms in all their phases: the initial compression, the main phase (with minima of ~ -170 nT and ~ -130 nT in the top and bottom panels, respectively) and the recovery phase. We recall here that in Sect. 3.3 the R parameter showed that “Lund” and EDDA should perform similarly for “small” and “moderate” storms. Figure 5 suggests that this is also true for “intense” storms.

We now turn to discuss a disturbed period when the ACE/SWEPAM instrument malfunctioned for a long time. For that purpose, we consider the well-known “severe” storms observed at the end of October 2003, which have been widely attributed to two CMEs connected to two solar flares of magnitudes X17 and X11, occurring one after another. Figure 6 displays the Kyoto provisional D_{st} index (black line) and two EDDA (red lines) and “Lund” (blue lines) indices from day 299 through day 308, 2003. For both the models the solid line refers to predictions made by using as inputs ACE verified data, while the dashed line refers to ACE real time data. At the top of the figure, a horizontal arrowed line marks the period, from 13:00 UT on day 301 to 23:00 UT on day 303, during which the ACE/SWEPAM data were heavily affected by an instrument malfunction, probably caused by enhanced fluxes of solar energetic particles. This lasted for more than two days, during which ACE transmitted to Earth on real time the incorrect values of

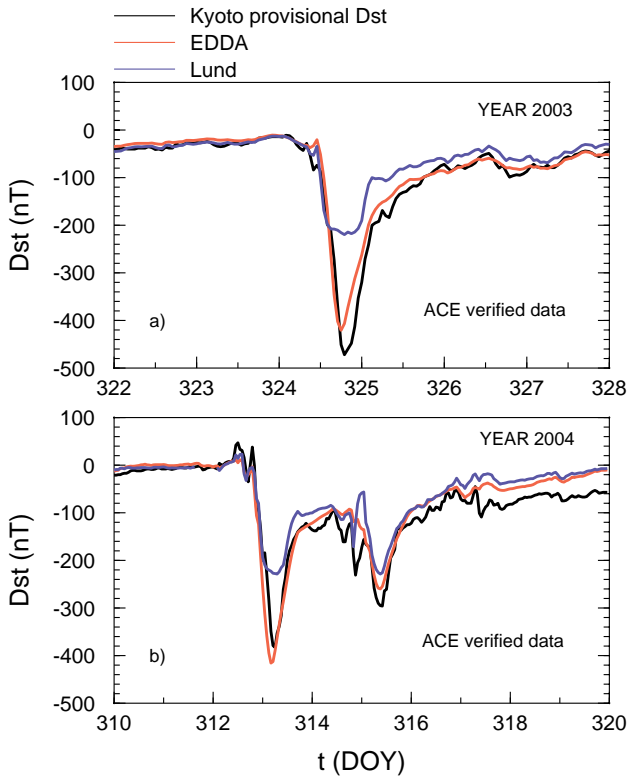


Fig. 7. Comparison of the EDDA (red lines) and “Lund” (blue lines) D_{st} predictions with the Kyoto provisional D_{st} index (black line) for two recent storms. In both panels the model D_{st} is calculated from ACE verified data.

proton density and bulk flow speed. During the same period the ACE/MAG magnetometer continued to produce reliable data. The incorrect plasma data have been excluded by the ACE/SWEPAM team from the verified final data. For that same time interval, the SOHO CELIAS/MTOF proton monitor measured a solar wind speed close to 1100 km/s (i.e. the upper instrumental limit), while the ACE Solar Wind Ion Composition Spectrometer (SWICS) indicated alpha particle speeds in excess of 1900 km/s. This also suggests that the SOHO verified plasma data are not reliable. In this situation, it does not make sense to run the “Lund” algorithm for verified data from 13:00 UT on day 301 to 23:00 UT on day 303. As a consequence, the solid blue line in Fig. 6 has a gap for that time period. On the other hand, we show the D_{st} forecasted by the “Lund” model for real time ACE data, to show the effect of the use of incorrect plasma data inputs on the D_{st} forecast. The “Lund” forecast misses altogether the first D_{st} decrease by ~ 120 nT on day 302; for the first storm it reaches a minimum of -180 nT, to be compared to the -360 nT Kyoto minimum, while for the second storm it displays a minimum of -110 nT, to be compared with the -400 nT Kyoto minimum. It is reasonable to argue that the residual correlation between the real-time “Lund” forecast and the Kyoto

D_{st} is due to its B_z input. As regards to the “Lund” forecast based on verified data, we notice that it compares reasonably well with the Kyoto D_{st} before the data gap. After the gap, it takes the algorithm almost two days to match again the Kyoto D_{st} . As regards to EDDA, we notice that its forecast based on verified ACE IMF data reproduces rather well the first storm, i.e. the large decrease to ~ -360 nT at 00:00 UT on day 303, with an exception made for a 3-h advance. On the other hand, the initial D_{st} decrease by ~ 120 nT on day 302 is only partially reproduced by EDDA. Moreover, EDDA fails to correctly forecast the minimum value of D_{st} for the second storm by 140 nT. In this case the EDDA forecast departs by about 35% from the “real” value, clearly more than the average value of 23 for the R parameter shown in Fig. 3, for $T_c < -50$ nT. Finally, we notice that the EDDA “real time” and the EDDA “verified” forecasts differ by $\sim 20 - 40$ nT for small D_{st} values. This is due to the fact that real time IMF data are affected by some errors which are corrected later on by the ACE/MAG team. However, the effects of such errors on the forecasted D_{st} are well below the expected average relative errors which can be expected from EDDA (see above the discussion over Fig. 3) and are negligible during the main phase of the first large storm.

In Fig. 7 we show the EDDA (red lines) and “Lund” (blue lines) performance for two “severe” storms which occurred after October 2003: the November 2003 storm (upper panel) and the November 2004 storm (lower panel). The ACE plasma data appear to be reliable and continuous, with an exception made for one missing data point in the first interval. To this regard, we comment that the missing plasma point would have produced a spurious transient in the “Lund” forecast. Actually, we avoided that by interpolating the ACE plasma data before feeding them to the network. In both occasions the Kyoto provisional D_{st} (black line) displays a very large negative excursion, below -400 nT. In the November 2003 case the Kyoto D_{st} rises slowly from ~ -40 nT to ~ -30 nT from day 322 to 12:00 UT on day 323, rises to a maximum value of ~ -10 nT around 06:00 UT on day 324, undergoes two successive decreases by ~ 40 nT over several hours, and then decreases to reach a minimum of ~ -470 nT at 19:00 UT on day 324, followed by a 2–2.5-day recovery. Both “Lund” and EDDA reproduce very well the Kyoto D_{st} until 10:00 UT on day 324. After that, “Lund” displays a broad minimum around ~ -210 nT, while EDDA drops to ~ -420 nT, 50 nT (i.e. $\sim 10\%$), short of the Kyoto minimum. EDDA follows the Kyoto index during the recovery phase, while “Lund” displays values larger than the Kyoto D_{st} by an amount which decreases from ~ 100 nT to ~ 10 nT over three days. In the November 2004 case an initial compression is observed on day 312; after the main storm with ~ -380 nT minimum, a further D_{st} negative excursion occurs, with ~ -300 nT minimum, during the recovery phase. Both “Lund” and EDDA reproduce very well the Kyoto D_{st} until the initial compression, which they both underestimate. Then “Lund” goes through two minima

of ~ -210 nT, corresponding to the two storms, short of ~ 170 nT and ~ 80 nT, respectively, from the Kyoto minima. On the contrary, EDDA (red lines) reproduces both the first and the second storms, in the first case overestimating the minimum by ~ 40 nT, and in the second case underestimating it by ~ 40 nT. However, three smaller dips on day 314 are missed by EDDA.

4 Discussion

Before discussing our findings, first of all, we recall that Wu and Lundstedt (1997) performed a very extensive study of the linear cross-correlation coefficients with different time lags between IMF and plasma parameters, on one hand, and D_{st} , on the other hand. They considered 9554 h storm time periods, and 1002 h quiet periods and calculated the linear correlation with D_{st} for single solar wind parameters and various functions based on them. We only briefly recall here that they found strong correlations with D_{st} for several functions and for B , B_s , B_z , and V , while they found a poor correlation for B_y . Such correlations were stronger for time delays ranging from one to several hours, depending on the single parameter. These results guided Wu and Lundstedt (1997) and, later on, Lundstedt et al. (2002) in selecting their ANN input parameters, including both IMF and solar wind parameters. We have not presented a similar analysis in this work, as we consider these results to be firmly established. However, we remark that the magnetospheric dynamics is thought to be nonlinearly related to the solar wind input, as we already recalled in Sect. 2.1. Actually, this is one of the reasons why ANN is useful for predicting the D_{st} index.

We have taken a different approach to the problem and have chosen to develop an IMF based ANN algorithm. To this regard, we have presented a reasonable argument for our choice in Sect. 2.3, arguing that magnetometers are more reliable than plasma instruments in terms of continuous temporal coverage. As we stated in the Introduction, the aim of this work was to develop an operational ANN D_{st} algorithm based on IMF only, putting a particular emphasis on the reproduction of the D_{st} behaviour for “severe” storms. The preceding sections show that that task has been accomplished, as the EDDA algorithm has been extensively and successfully tested over 58 000 h between 1995 and 2002 and over 26 000 h between 2003 and 2005.

We now try to explain how EDDA can predict D_{st} on the basis of IMF only. In this respect, we recall that magnetic reconnection at the magnetopause and in the tail is, by many authors, considered as an essential process for the transfer of energy, mass and momentum from the solar wind to the magnetosphere and for energy conversion in the magnetosphere itself (Dungey, 1961). Several parameters have been defined and tested over the course of years, to describe the transfer of energy from the solar wind to the magnetosphere. Among them is the Akasofu ϵ parameter (Perreault and Akasofu,

1978; Akasofu, 1981), recently reviewed by Koskinen and Tanskanen (2002). This is defined as $\epsilon = VB^2 \sin^4(\theta/2) l_0^2$, where B is the magnitude of IMF, V is the solar wind speed, θ is the IMF clock angle and $l_0 \simeq 7$ Earth radii. We notice that ϵ contains the square of the magnetic field intensity, i.e. the second EDDA input. Moreover, ϵ contains the θ clock angle, which is determined by B_z and B_y . B_z is EDDA’s first input. B_y^2 was taken as the third EDDA input, because B_y is known to play a role in the transfer of mass, energy and momentum transfer from the solar wind to the magnetosphere (e.g. Cravens, 1997); since its sign should not be important for such processes, it appears reasonable to have it squared. To comment on the fact that the ϵ parameter also depends on V , we notice that this dependence is linear, whereas on B this dependence is quadratic. This suggests that changes in V have a smaller impact on the magnetic energy flux carried by the solar wind. However, this point deserves probably further investigation which could be the subject of a future work.

In Sect. 3.3 we compared the EDDA performance with that of the “Lund” algorithm, which, as we already noted, was found in the past to perform better (see Lundstedt et al., 2002) than several algorithms based on differential equations (Burton et al., 1975; Fenrich and Luhmann, 1998; O’Brien and McPherron, 2000). First of all, we made the comparison from a statistical point of view, finding that “Lund” performs somewhat better than EDDA for quiet time D_{st} values, while the two algorithms are expected to perform in a similar way for “small” and “moderate” storm time D_{st} values. This result was extended to “intense” storms by the examination of two cases (see Fig. 5). However, when we compared the two algorithms for four “severe” recent storms (which are a fair sample of the “severe” storms observed between 2003 and 2005), we concluded that EDDA performs better: for the October 2003 two storms, the “Lund” forecasts, based on real time data, fail because the plasma data provided by ACE/SWEPAM are wrong (Fig. 6); for the November 2003 and 2004 storms “Lund” largely fails to match the minima of both storms (Fig. 7). With regard to such different performances we can make the hypothesis, that they depend on the different training sets used for the two algorithms, but cannot test this hypothesis as we do not know the exact time periods used for the “Lund” training.

In the last part of this section we wish to comment on minor limitations of the EDDA algorithm, which we have already pointed out in the preceding sections. First of all, we recall the second of the October 2003 storms (see Fig. 6), which occurred during the recovery phase of a preceding storm. In this case we remarked a relative error of 35 percent on the D_{st} minimum. The probable reason for this poor performance is that EDDA was not properly trained for such a situation, where a new solar wind stimulus hit the magnetosphere which was in a very disturbed state already. Such a situation has been observed for other storms in our data set. On the other hand, in apparently similar cases, the second storm is correctly reproduced by EDDA (e.g. see day 315,

2004 in Fig. 7). The reason for such different performances is a matter for future investigation.

We now turn to the differences often observed between the Kyoto index and the forecasted index for small D_{st} values (cf. Sect. 3.3). As quiet periods account for the great majority of data points, the ultimate effect of such differences is the increase in the total RMSE and of the high values of the R parameter for $D_{st} > -25$ nT, as already discussed with reference to Fig. 3. Actually, high values for such parameters, in the order of 100, are commonly quoted for all past algorithms, as it can be seen in Fig. 3 of Lundstedt et al. (2002) and in Table 1 of Wang et al. (2003). In this respect, several considerations can be made. First of all, we recall that the production of the D_{st} index requires subtracting out a fairly big Sq ionospheric current system contribution. This subtraction is imperfect and leaves a residual signal (Temerin and Li, 2002). Moreover, it must be considered that the unperturbed value of the horizontal component of the geomagnetic field is not exactly known, but it is only estimated through the annual average of the D_{st} values calculated from the 5 quietest days observed at each station during every month. It is plausible, then, that the determination of the baseline can be, in certain circumstances, an error source for the prediction of small D_{st} absolute values. In fact, we observe in the test data set periods of very low geomagnetic activity, also lasting ten days, when the predicted D_{st} is constantly lower or higher, on the order of 20 nT, than the observed one. Actually, the uncertainties in the determination of D_{st} are one of the major concerns of all authors attempting to develop algorithms for its forecast (e.g. see the discussion made by Temerin and Li, 2002).

To conclude this section, we briefly comment on the EDDA performance during the initial compression of the magnetosphere prior to the main storm phase. It is accepted that such a compression be produced by the solar wind dynamic pressure (mostly density) increase. Therefore, it is reasonable that the EDDA algorithm fails to reproduce such a compression, as it does not include among its inputs neither the solar wind density nor its speed. We have noted that this seems to happen, to different extents, for the majority of the storms described in this paper. However, this does not occur for the storms of day 274, 2002, day 229, 2003, and day 243, 2004. We have already noticed in Sect. 3.3 that the large value of the parameter R for small values of D_{st} could be due in part to this effect. We note, however, as we recalled already, that large values of the RMS and of the R parameter for small D_{st} values are also quoted for algorithms which include the plasma parameters among their inputs (see, e.g. Fig. 3 of Lundstedt et al., 2002) and (Table 1 of Wang et al., 2003). Moreover, “Lund” and EDDA behaved in the same way during the initial compression for the three examples discussed in Sect. 3.4. A more detailed discussion of this issue goes beyond the purpose of this paper.

5 Summary

We described EDDA, a new Elman Artificial Neural Network, trained over ~ 6000 hourly averages of WIND and ACE data. EDDA calculates the D_{st} index on the basis of IMF data only and is, therefore, capable of issuing operational forecasts based on the current real time ACE L1 IMF data. The EDDA performance was carefully examined over an extended test period from 1995 to 2002. Moreover, its performance was checked for the recent “severe” geomagnetic storms which occurred between 2003 and 2005, for which provisional, but not final, D_{st} data are available. Moreover, it was shown that the EDDA performance is in general comparable to that of similar algorithms which also make use of plasma data, although a failure to fully reproduce the initial compression prior to the storm main phase may occur. It was also shown that the EDDA ability to forecast the D_{st} index based on IMF only has an undoubted operational advantage in all circumstances (very interesting in the framework of space weather) when the predictions of algorithms based on both IMF and plasma parameters fail because the solar wind speed exceeds the upper limit of the L1 plasma instruments or large radiation, and SEP fluxes cause temporary faults in such instruments.

Finally, we suggested that the three magnetic inputs of the EDDA model, namely B_z , B^2 , B_y^2 , which closely recall the information contained in the Akasofu ϵ parameter, can catch, especially in enhanced geomagnetic activity conditions, the large majority of the relevant information necessary to describe the relationship between the solar wind trigger and the D_{st} index.

Appendix A

The Empirical D_{st} Data Algorithm

The EDDA algorithm can be implemented through a computer programme performing the following steps.

- 1) set $\mathbf{c} = (0.0, 0.0, 0.0, 0.0)^T$
- 2) get first or next input data B_z, B^2, B_y^2
- 3) compute $\mathbf{u} = (u_1, u_2, u_3)^T = \left(\frac{B_z}{N_{r1}}, \frac{B^2}{N_{r2}}, \frac{B_y^2}{N_{r3}} \right)^T$
- 4) compute $\mathbf{x} = \tanh(w^{(1)}\mathbf{u} + w^{(c)}\mathbf{c})$
- 5) compute $D_{st} = N_{r4}w^{(2)}\mathbf{x}$
- 6) set $\mathbf{c} = \mathbf{x}$
- 7) goto 2

At the beginning the context units are unknown and are set to 0. Depending on solar wind conditions, it takes from a few hours to 1–2 days for EDDA to reach normal operation. The same occurs in the case of data gaps.

The inputs should be calculated as follows: $B_z = \langle b_z \rangle$, $B^2 = \langle b^2 \rangle$, $B_y^2 = \langle b_y^2 \rangle$. Here the averages are made over

1 h and the b symbol denotes high time resolution IMF GSM data.

In step 3 the inputs are divided by the normalisation factors N_{r1} , N_{r2} and N_{r3} , so that the network operates on adimensional inputs. Similarly, in step 5 the network output is multiplied by the normalisation factor N_{r4} . This is common practice with neural networks (see, e.g. Wu and Lundstedt, 1997). The four normalisation factors were obtained by calculating the maxima of the absolute values of the corresponding parameters over the period 1995–2002 for the OMNI data set.

The EDDA weights and normalization factors are:

$$w^{(1)} = \begin{pmatrix} -0.00863 & 0.25861 & -0.03929 \\ 0.00387 & -0.04102 & -0.01309 \\ -0.05323 & 0.01495 & -0.03294 \\ -0.00418 & -0.03920 & -0.01922 \end{pmatrix}$$

$$w^{(c)} = \begin{pmatrix} 0.29238 & 0.07855 & 0.05065 & 0.15604 \\ 0.10092 & 0.88728 & 0.03804 & 0.11780 \\ 0.28768 & 0.01891 & 0.76351 & 0.07686 \\ 0.56835 & 0.15990 & 0.05034 & 0.11086 \end{pmatrix}$$

$$w^{(2)} = (-0.08588 \quad -0.00006 \quad -2.02734 \quad -0.07205)$$

$$N_{r1} = 44.7 \text{ nT}, \quad N_{r2} = (48.2)^2 (\text{nT})^2,$$

$$N_{r3} = (34.3)^2 (\text{nT})^2, \quad N_{r4} = 387.0 \text{ nT}.$$

Acknowledgements. The authors would like to thank the ACE spacecraft team and SEC for making the real time and verified data available, and the WIND, GOES, SOHO spacecraft teams and NSSDC for making their datasets available, too. This work was partially supported by the ESA contract N. 17032/03/NL/LvH, in the framework of the Pilot Project on Space Weather Applications.

Topical Editor T. Pulkkinen thanks M. Temerin and another referee for their help in evaluating this paper.

References

- Akasofu, S. I.: Prediction of development of geomagnetic storms using the solar wind-magnetosphere energy coupling function ϵ , *Planet. Space Sci.*, 29, 1151–1158, 1981.
- Burton, R. K., McPherron, R. L., and Russell, C. T.: An empirical relationship between interplanetary conditions and Dst, *J. Geophys. Res.*, 80, 4204–4214, 1975.
- Consolini, G. and Chang, T. S.: Magnetic Field Topology and Criticality in Geotail Dynamics: Relevance to Substorm Phenomena, *Space Sci. Rev.*, 95, 309–321, 2001.
- Cravens, T. E.: *Physics of Solar System Plasmas*, Cambridge University Press, Cambridge, UK, 1997.
- Dungey, J. W.: Interplanetary Magnetic Field and the Auroral Zones, *Phys. Rev. Lett.*, 6, 47–48, 1961.
- Elman, J. L.: Finding structure in time, *Cognitive Sci.*, 14, 179–211, 1990.
- Fenrich, F. R. and Luhmann, J. G.: Geomagnetic response to magnetic clouds of different polarity, *Geophys. Res. Lett.*, 25, 2999–3002, 1998.
- Gonzalez, W. D. and Echer, E.: A study on the peak Dst and peak negative Bz relationship during intense geomagnetic storms, *Geophys. Res. Lett.*, 32, 18 103–18 106, 2005.
- Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rosotoker, G., Tsurutani, B. T., and Vasyliunas, V. M.: What is a geomagnetic storm?, *J. Geophys. Res.*, 99, 5771–5792, 1994.
- Gonzalez, W. D., Tsurutani, B. T., and Clúa de Gonzalez, A. L.: Interplanetary origin of geomagnetic storms, *Space Sci. Rev.*, 88, 529–562, 1999.
- Gosling, J. T.: The solar flare myth, *J. Geophys. Res.*, 98, 18 937–18 950, 1993.
- Kane, R. P.: How good is the relationship of solar and interplanetary plasma parameters with geomagnetic storms?, *J. Geophys. Res.*, 110, 2213–2215, 2005.
- Klimas, A. J., Vassiliadis, D., Baker, D. N., and Roberts, D. A.: The organized nonlinear dynamics of the magnetosphere, *J. Geophys. Res.*, 101, 13 089–13 114, 1996.
- Koskinen, H. E. J. and Tanskanen, E. I.: Magnetospheric energy budget and the epsilon parameter, *J. Geophys. Res.*, 107, A11, doi:10.1029/2002JA009283, 2002.
- Lundstedt, H., Gleisner, H., and Wintoft, P.: Operational forecasts of the geomagnetic index D_{st} index, *Geophys. Res. Lett.*, 29(24), 2181, doi:10.1029/2002GL016151, 2002.
- O'Brien, T. P. and McPherron, R. L.: Forecasting the ring current index Dst in real time, *J. Atmos. Terr. Phys.*, 62, 1295–1299, 2000.
- Perreault, P. and Akasofu, S.-I.: A study of geomagnetic storms, *Geophys. J.*, 54, 547–573, 1978.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, *Nature*, 323, 533–536, 1986.
- Temerin, M. and Li, X.: A new model for the prediction of Dst on the basis of the solar wind, *J. Geophys. Res.*, 107, A12, 1472, doi:10.1029/2001JA007532, 2002.
- Wang, C. B., Chao, J. K., and Lin, C. H.: Influence of the solar wind dynamic pressure on the decay and injection of the ring current, *J. Geophys. Res.*, 108, A9, 1341, doi:10.1029/2003JA009851, 2003.
- Watari, S., Kunitake, M., and Watanabe, T.: The Bastille Day (14 July 2000) Event in Historical Large Sun-Earth Connection Events, *Solar Phys.*, 204, 425–438, 2001.
- Wu, J. and Lundstedt, H.: Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks, *J. Geophys. Res.*, 102, 14 255–14 268, 1997.
- Yermolaev, Y. I., Yermolaev, M. Y., Zastenker, G. N., Zelenyi, L. M., Petrukovich, A. A., and Sauvaud, J. A.: Statistical studies of geomagnetic storm dependencies on solar and interplanetary events: a review, *Planet. Space Sci.*, 53, 189–196, 2005.