

Prediction of SYM-H index during large storms by NARX neural network from IMF and solar wind data

L. Cai, S. Y. Ma, and Y. L. Zhou

Dept. of Space Physics, College of Electronic Information, Wuhan University, Wuhan 430079, China

Received: 26 December 2008 – Revised: 16 November 2009 – Accepted: 22 January 2010 – Published: 2 February 2010

Abstract. Similar to the Dst index, the SYM-H index may also serve as an indicator of magnetic storm intensity, but having distinct advantage of higher time-resolution. In this study the NARX neural network has been used for the first time to predict SYM-H index from solar wind (SW) and IMF parameters. In total 73 time intervals of great storm events with IMF/SW data available from ACE satellite during 1998 to 2006 are used to establish the ANN model. Out of them, 67 are used to train the network and the other 6 samples for test. Additionally, the NARX prediction model is also validated using IMF/SW data from WIND satellite for 7 great storms during 1995–1997 and 2005, as well as for the July 2000 Bastille day storm and November 2001 superstorm using Geotail and OMNI data at 1 AU, respectively. Five interplanetary parameters of IMF B_z , B_y and total B components along with proton density and velocity of solar wind are used as the original external inputs of the neural network to predict the SYM-H index about one hour ahead. For the 6 test storms registered by ACE including two super-storms of min. SYM-H < -200 nT, the correlation coefficient between observed and NARX network predicted SYM-H is 0.95 as a whole, even as high as 0.95 and 0.98 with average relative variance of 13.2% and 7.4%, respectively, for the two super-storms. The prediction for the 7 storms with WIND data is also satisfactory, showing averaged correlation coefficient about 0.91 and RMSE of 14.2 nT. The newly developed NARX model shows much better capability than Elman network for SYM-H prediction, which can partly be attributed to a key feedback to the input layer from the output neuron with a suitable length (about 120 min). This feedback means that nearly real information of the ring current status is effectively directed to take part in the prediction of SYM-H index by ANN. The proper history length of the output-feedback may

mainly reflect on average the characteristic time of ring current decay which involves various decay mechanisms with ion lifetimes from tens of minutes to tens of hours. The Elman network makes feedback from hidden layer to input only one step, which is of 5 min for SYM-H index in this work and thus insufficient to catch the characteristic time length.

Keywords. Magnetospheric physics (Solar wind-magnetosphere interactions; Storms and substorms; Instruments and techniques)

1 Introduction

Magnetic storm is a multi-faceted phenomenon that originates from the sun and interplanetary disturbances and manifests in the geospace of magnetosphere-ionosphere-thermosphere (M-I-T) coupling system. Long-lasting large southward IMF (Interplanetary Magnetic Field) enables energy transfer from the solar wind to the magnetosphere to enhance greatly through magnetic reconnection at dayside magnetopause. Large interplanetary electric fields (IEF) break into geospace, leading to dramatically enhanced and earthward shifted ring currents which cause a rapid decrease in the horizontal component of the geomagnetic field on the ground at low latitudes and characterize a geomagnetic storm development (Gonzalez et al., 1994; Gosling et al., 1991; Tsurutani et al., 1988; Daglis, 2006). Satellite observations show that storm time ring current mainly consists of ions with energy from tens to hundreds keV which drift westward over the equatorial area within $2-7 R_E$ height. During the main phase of large storm, the content of heavy ionospheric ions (especially O^+) in the ring current ($L < 4$) is extremely increased. The larger the magnetic storm is, the greater the O^+ contributes to the ring current (Daglis, 1997). When southward IMF decreases or disappears, the rate of magnetic reconnection reduces, convection electric field turns weak,



Correspondence to: S. Y. Ma
(syyma@whu.edu.cn)

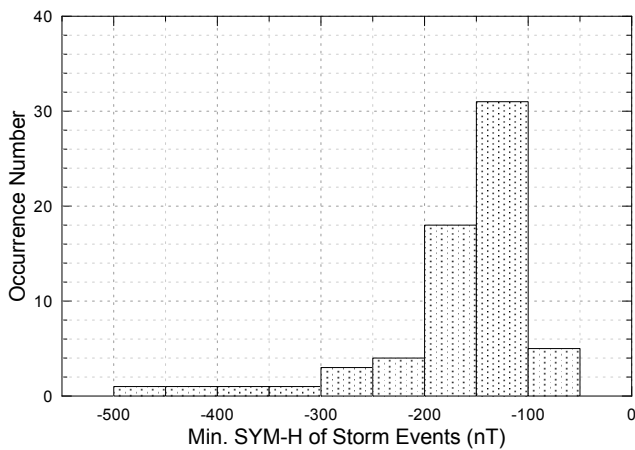


Fig. 1. Histogram of minimum SYM-H index for the events used in ANN model.

particles newly input into ring current decreases, the convection boundary moves outside, cold ionospheric plasma refills the exhausted inner magnetosphere, then encounters with energy particles, and thus gives rise to plasma waves. Ring current ions loss is caused through 4 kinds of processes such as charge exchanges with neutral atoms, Coulomb collisions with plasma, pitch angle diffusion caused by wave-particle interaction, as well as drifting out of the dayside magnetopause (Daglis, 2006), and thus the ring current decays. This is the recovery phase of a magnetic storm. Among the loss processes cited above, the charge exchange is dominant for mono-charged ions (e.g., H^+ , O^+) with moderate energy (10–100 keV). For ions with higher energy, the loss caused by the wave-particle interaction becomes more significant due to the decrease of charge exchange section. For low energy heavy ions (less than 10 keV), the loss caused by Coulomb collision cannot be neglected. The fast decay of ring current during the recovery phase of intense storm is caused by the rapid loss of O^+ and He^+ .

The hourly Dst index is a measure of the total energy content of the ring current, being used the most widely so far to quantify the geomagnetic storm activity (Sugiura and Poros, 1964; Mayaud, 1980; Rangarajan, 1989). SYM-H is another geomagnetic storm index to measure the intensity of the storm-time ring current proposed in recent more than two decades. In contrast to 1 h time-resolution of Dst, SYM-H has the advantage of much higher time resolution of 1 min. Essentially, the SYM-H index is the same as Dst (Wanliss and Showalter, 2006) except that it is calculated from a different set of stations and in a slightly different coordinate system.

In the past several decades, many studies have been devoted to the relationship between the solar wind and the geomagnetic storm (e.g., Rostoker and Fälthammar, 1967; Aka-sofu, 1981; McPherron et al., 1988). Various methods have

been applied to predict geomagnetic disturbances as indicated by Dst index from solar wind and IMF observations, such as differential equations (Burton et al., 1975; Wang et al., 2003), statistical correlative analysis (Baker, 1986; Temerin and Li, 2002; Yermolaev et al., 2005), and artificial neural networks (ANNs) (Wu and Lundstedt, 1996; Gleisner et al., 1996; Wu and Lundstedt, 1997; Lundstedt et al., 2002; Pallocchia et al., 2006; Amata et al., 2008). Of those, artificial neural networks have shown their great ability of non-linear mapping in Dst prediction. However, there are not any reports yet on the prediction of SYM-H index by ANN in the publications. The prediction of SYM-H index can be applied not only in magnetic storm forecasting, but also in predicting thermosphere and ionosphere parameters through empirical models where the SYM-H or Dst index plays a dominant role (e.g. Zhou et al., 2009; Ridley and Liemohn, 2002). The higher time-resolution of SYM-H index makes its prediction somewhat different from and harder than the Dst index.

In this study, an artificial neural network (ANN) of Non-linear Auto Regressive with eXogenous inputs (NARX) has been developed for the first time to predict SYM-H index about one hour in advance from solar wind and IMF parameters. The newly developed NARX model shows much better capability than Elman network in SYM-H prediction. In the following sections we will firstly present the data used to establish the SYM-H prediction model of NARX NN and the parameters selected as the external inputs. Then a brief description is given on the architectures and training methods for Elman recurrent neural network (ERNN) that we have ever tried, as well as the newly developed NARX network. Afterwards, the performance and prediction results of the two ANN models are shown and compared with each other. Finally, a brief discussion and summary is given along with ideas for further worthy studies.

2 Data sets and external inputs

The solar wind and IMF data observed from ACE satellite at Lagrange point L1 are used to establish the neural network model to predict SYM-H index. The ACE data are obtained from CDAWeb site of NASA and the SYM-H index data are provided by World Data Center for Geomagnetism, Kyoto. The solar wind and IMF data are sampled at rates of every 64 and 16 s, respectively, while SYM-H index at 1 min. To unify the temporal resolution, they are all averaged every 5 min. Therefore, all the data we used to establish the ANN model have the same time resolution of 5 min. In this study, we collected 73 time intervals that contain storm events with the minimum SYM-H value less than -85 nT during 1998–2006, of which 67 intervals comprising 66 800 samples are used to train the neural network and the other 6 intervals are used for testing. The selection of the interval is constrained by the data availability and quality of ACE observations in addition to the threshold of -80 nT for minimum

Table 1. Storm intervals used to develop the ANN prediction model.

No.	Start date	End date	Min. SYM-H (nT)	No.	Start date	End date	Min. SYM-H (nT)
1	1998-02-17	1998-02-19	-119	35	2001-10-21	2001-10-24	-210
2	1998-03-10	1998-03-12	-119	36	2001-10-27	2001-10-30	-150
3	1998-05-03	1998-05-06	-266	37	2002-03-23	2002-03-25	-114
4	1998-06-25	1998-06-26	-120	38	2002-04-17	2002-04-21	-183
5	1998-08-06	1998-08-08	-168	39	2002-05-11	2002-05-13	-109
6	1998-08-26	1998-08-28	-171	40	2002-05-23	2002-05-25	-113
7	1998-09-24	1998-09-26	-213	41	2002-08-01	2002-08-04	-114
8	1998-10-19	1998-10-20	-120	42	2002-08-19	2002-08-22	-119
9	1998-11-07	1998-11-10	-179	43	2002-09-03	2002-09-07	-108
10	1998-11-12	1998-11-15	-123	44	2002-09-07	2002-09-12	-167
11	1999-01-13	1999-01-14	-111	45	2002-10-01	2002-10-06	-153
12	1999-02-17	1999-02-20	-127	46	2002-10-14	2002-10-15	-88
13	1999-02-28	1999-03-02	-93	47	2002-10-23	2002-10-26	-87
14	1999-04-16	1999-04-17	-122	48	2002-11-20	2002-11-22	-126
15	1999-09-22	1999-09-24	-160	49	2003-05-29	2003-06-03	-162
16	1999-11-13	1999-11-15	-106	50	2003-06-16	2003-06-19	-162
17	2000-01-22	2000-01-24	-101	51	2003-07-10	2003-07-17	-125
18	2000-02-12	2000-02-13	-164	52	2003-08-17	2003-08-20	-138
19	2000-05-23	2000-05-26	-159	53	2003-11-04	2003-11-05	-87
20	2000-06-08	2000-06-09	-90	54	2003-11-19	2003-11-21	-488
21	2000-07-16	2000-07-21	-294	55	2003-11-21	2003-11-23	-488
22	2000-08-10	2000-08-13	-235	56	2004-01-21	2004-01-27	-137
23	2000-09-17	2000-09-19	-196	57	2004-02-11	2004-02-12	-107
24	2000-10-02	2000-10-07	-183	58	2004-07-22	2004-07-24	-122
25	2000-10-12	2000-10-15	-100	59	2004-07-24	2004-07-27	-166
26	2000-10-28	2000-10-30	-120	60	2004-07-27	2004-07-29	-208
27	2000-11-06	2000-11-07	-174	61	2004-08-29	2004-09-01	-128
28	2000-11-10	2000-11-11	-126	62	2004-11-07	2004-11-13	-393
29	2000-11-26	2000-11-30	-165	63	2005-01-21	2005-01-23	-101
30	2001-03-19	2001-03-21	-434	64	2005-05-15	2005-05-22	-302
31	2001-03-27	2001-04-03	-275	65	2005-06-12	2005-06-14	-112
32	2001-04-11	2001-04-14	-275	66	2005-06-23	2005-06-24	-101
33	2001-08-17	2001-08-19	-130	67	2005-09-11	2005-09-16	-127
34	2001-09-30	2001-10-05	-187				

SYM-H. Generally, each time interval comprises 3 phases, that is, 1 days before the storm commencement, the whole storm (including initial, main and recovery phases), and ~ 1 day after the storm. All the storm events for network training are listed in Table 1, giving the time interval, the value of minimum SYM-H for each storm event. Figure 1 displays the histogram of the minimum SYM-H for the storm events used for neural network training. It can be seen from Fig. 1 and Table 1 that the training samples contain 11 super-storms with min. SYM-H less than -200 nT, occupying a fraction of 16.4%, while great storms with minimum SYM-H value from -200 nT to -100 nT hold a dominant percentage of 73%. The testing samples using ACE SW/IMF data are all great storms (minimum SYM-H < -100 nT), containing 2 super-storms (see Table 2, the 3 columns on the left). In addition, to validate the trained ANN model, another 9 storm events registered in rather than ACE measurements are also

used for test. Of them, 6 great storms occurring in 1995–1997 and one super-storm in 2005 are tested using SW/IMF data from WIND satellite (see Table 3, the 3 columns on the left). The other two ones are superstorms occurring on July 2000 Bastille day and 24 November 2001, which are tested using Geotail and OMNI data at 1 AU with no time ahead, respectively.

The selection of inputs has an important impact on the result of ANN prediction. In previous studies of Dst prediction by ANNs, different combinations of solar wind and IMF parameters with different time-length (i.e., the correlation time between each input and Dst) have ever been used as the ANNs' inputs, such as group of n , V and IMF B_z with 1 h correlation length (Lundstedt et al., 2002) where n is the proton number density of solar wind, and V the solar wind velocity, and group of only IMF of B_z , B_y^2 and $B^2 = B_x^2 + B_y^2 + B_z^2$ with the same correlation time length

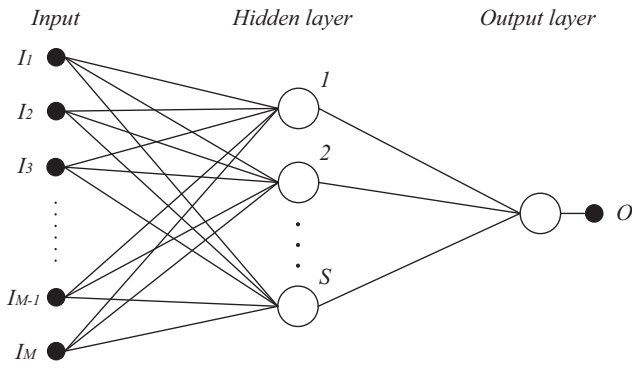


Fig. 2. Architecture of two-layer back-propagation neural network.

(Pallochia et al., 2006) and others. In the present study we use the parameter combination of n , V , B_z , B_y and B with 90-min history length for each parameter as input to predict SYM-H index 60 min ahead. Thus, our prediction model has totally 90 input nodes, each 18 for B_y , B_z , B , n and V . The input \mathbf{I} at moment t and the corresponding output \mathbf{O} are:

$$\mathbf{I}(t) = \{B_y(t), B_z(t), B(t), n(t), V(t), \dots, B_y(t-17), B_z(t-17), B(t-17), n(t-17), V(t-17)\}, \quad (1)$$

$$\mathbf{O}(t) = \{\text{SYM-H}(t+12)\}. \quad (2)$$

In order to make the neural network converge more easily, all the input parameters are normalized to $[-0.8, 0.8]$ before used in ANN.

3 ANN models

In this study, we introduce for the first time the NARX neuron network to the prediction of magnetic storm index SYM-H with time resolution of a few minutes and compare it with usual Elman network which enable already very good prediction for Dst index with one hour resolution. In this section we present first the architectures of the two kinds of neural networks.

3.1 Architectures

Both of the Elman and NARX artificial neural network are based on the two-layer BP network, having different layout for global feedback. So, we begin with a brief presentation of the so-called BP network.

3.1.1 BP neural network

BP (back-propagation) neural network is a kind of classic feed-forward network using error back-propagating algorithm for its training. It is usually a two-layer network with one hidden layer when used in the Dst prediction. Figure 2 shows its architecture. For prediction purpose, there

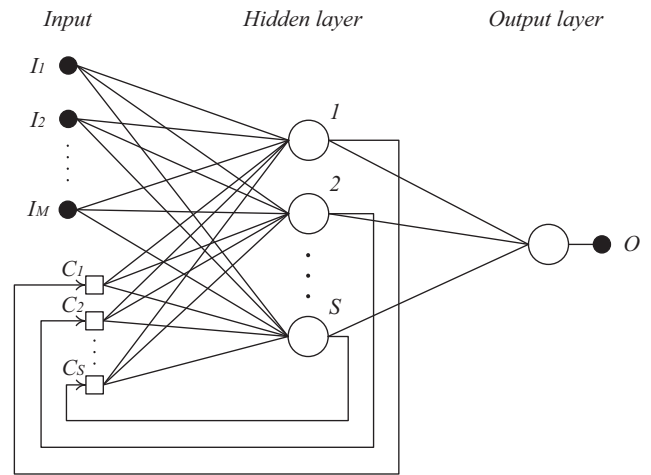


Fig. 3. Architecture of Elman neural network.

is only one neural in output layer. The activation function is the hyperbolic tangent for the hidden layer and linear one for the output layer (Haykin, 1999). For the input $\mathbf{I} = \{I_1, I_2, \dots, I_M\}$, the output of the j -th hidden unit, H_j , is

$$H_j = \tanh \left(\sum_{i=1}^M w_{ji} I_i + b_j \right), \quad (3)$$

where I_i is the value of input node i , M is the number of input nodes, w_{ji} is the connecting weight between the input node i and the hidden neuron j , and b_j is bias of the hidden neuron j . And the output O is

$$O = \sum_{j=1}^S w_{oj} H_j + b_o, \quad (4)$$

where S is the number of hidden neurons, w_{oj} is the connecting weight between the hidden neuron j and the output neuron, and b_o is bias of the output neuron.

3.1.2 Elman neural network

An Elman neural network is a two-layer BP network with feedbacks from hidden neurons to input (Elman, 1990), being called also Elman recurrent network (ERN). Figure 3 shows its architecture. The input of an ERN consists of two parts, one is true external input $\mathbf{I} = \{I_1, I_2, \dots, I_M\}$, and the other is feedback input $\mathbf{C} = \{C_1, C_2, \dots, C_S\}$. The feedback input nodes are called context nodes or context units. The value of context unit l at time step t is simply a copy of the hidden output at time step $t-1$, i.e.

$$C_l(t) = H_l(t-1) \quad l = 1, 2, \dots, S. \quad (5)$$

For the Elman network, the j -th hidden unit output H_j is

$$H_j = \tanh \left(\sum_{i=1}^M w_{ji} I_i + \sum_{l=1}^S w_{jl} C_l + b_j \right). \quad (6)$$

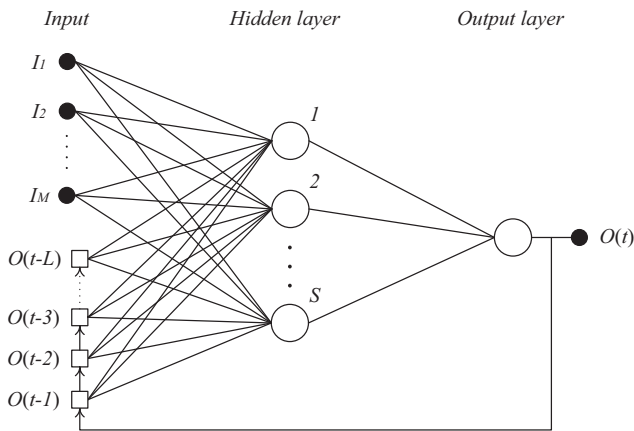


Fig. 4. Architecture of Elman neural network.

Context units are considered containing history information of the network status. And it is proved that the Elman neural network can give very good performance of Dst prediction (Wu and Lundstedt, 1997).

The Elman ANN used in this study has various numbers of original external input nodes, 12 hidden neurons and 1 output neuron.

3.1.3 NARX neural network

We introduce a new kind of feedback neural network of Nonlinear Auto Regressive with eXogenous inputs (NARX) model (Haykin, 1999) to predict the SYM-H index from IMF and solar wind. The NARX network is also a two-layer BP network with a time-delay feedback connection as shown in Fig. 4. The input of network consists of two parts as well: true input and context input. In contrast to ERN, the context input for NARX is not from the hidden layer but from the output layer with some certain time-delay. Assuming that $\mathbf{I}(t)$ and $\mathbf{O}(t)$ are the true input and output at time step t and the length of feedback time delay line is L , the context input is $\mathbf{C}(t) = \{O(t-1), O(t-2), \dots, O(t-L)\}$. The j th hidden unit output is

$$H_j = \tanh \left(\sum_{i=1}^M w_{ji} I_i + \sum_{l=1}^L w_{jl} C_l + b_j \right). \quad (7)$$

The context nodes fed back from the output contain the being predicted parameter's information with much more steps of history than the Elman network. The determination of the optimum history length for feedback, i.e., the proper number of the context units, will be investigated in detail in Sect. 4.

The NARX ANN used in this study has 90 original external input nodes, and 12 hidden neurons as well as 1 output neuron.

3.2 Establish of ANN prediction model

3.2.1 The training method

The Momentum Back-Propagation algorithm is used to train the 2 kinds of networks. The course of learning all the samples for once is called an epoch. For each epoch, the cost function is defined as

$$E = \frac{1}{2} \sum_{k=1}^N (T^k - O^k)^2, \quad (8)$$

where T^k and O^k are, respectively, the target output and the actual output of the output neuron, N is the number of training samples. For each connecting weight w at the n -th epoch, it is updated according to

$$\Delta w(n) = -\eta \frac{\partial E}{\partial w} + \alpha \Delta w(n-1), \quad (9)$$

where η is the learning rate and α is the momentum parameter. For connecting weights between the hidden and output neurons, they are adjusted following

$$\Delta w_{oj}(n) = \eta \sum_{k=1}^N \delta_o^k H_j^k + \alpha \Delta w_{oj}(n-1), \quad (10)$$

$$\delta_o^k = T^k - O^k. \quad (11)$$

For connecting weights between the input nodes and the hidden neurons, the adjusting is as follows

$$\Delta w_{ji}(n) = \eta \sum_{k=1}^N \delta_j^k I_i^k + \alpha \Delta w_{ji}(n-1), \quad (12)$$

$$\delta_j^k = \left(1 - (H_j^k)^2 \right) \quad j = 1, 2, \dots, S, \quad (13)$$

where the input $\mathbf{I} = \{I_i\}$ includes the context input for Elman and NARX network.

While learning, η and α are adjusted dynamically. According to the change of cost function, ΔE , the adjusting rule is

$$\eta = \begin{cases} a\eta_0 & \Delta E < 0 \\ b\eta_0 & \Delta E \geq 0 \ \& \ \Delta E/E < \varepsilon \\ \eta_0 & \Delta E \geq 0 \ \& \ \Delta E/E \geq \varepsilon \end{cases}, \quad (14)$$

$$\alpha = \begin{cases} \alpha_0 & \Delta E < 0 \\ 0 & \Delta E \geq 0 \end{cases}, \quad (15)$$

where a and b are positive constants and $a > 1$ and $b < 1$, α_0 is the original momentum parameter, ε is a pre-defined scale value that is usually less than 5%. In this study, we set the training parameters as: $a = 1.04$, $b = 0.7$, $\eta_0 = 1$, $\alpha_0 = 0.7$ and $\varepsilon = 0.03$.

When the neural networks are training, the accuracy parameters of correlation coefficient ρ and root mean square error (RMSE) as well as the averaged relative variance (ARV) are calculated for the training samples at the end of every training epoch. The accuracy parameters of ρ , RMSE and ARV are defined as (Wu and Lundstedt, 1997)

$$\rho = \frac{\sum_{k=1}^N (T^k - \bar{T})(O^k - \bar{O})}{\sqrt{\sum_{k=1}^N (T^k - \bar{T})^2} \sqrt{\sum_{k=1}^N (O^k - \bar{O})^2}}, \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (T^k - O^k)^2}, \quad (17)$$

$$\text{ARV} = \frac{\sum_{k=1}^N (T^k - O^k)^2}{\sum_{k=1}^N (T^k - \bar{T})^2}, \quad (18)$$

where N is the number of examples, \bar{T} and \bar{O} are the average values of the observed and the predicted SYM-H, respectively.

At the beginning of training, RMSE decreases quickly. With the epoch number increasing, the error decreases more and more slowly, converging gradually to an asymptotic value. Assuming that $\text{RMSE}_{\min}(n)$ is the minimum of RMSE till epoch n , we define a parameter of γ as

$$\gamma(\tau) = \frac{\text{RMSE}_{\min}(n - \tau) - \text{RMSE}_{\min}(n)}{\text{RMSE}_{\min}(n)}, \quad (19)$$

to judge if the training should finish at epoch n . When $\gamma(1000) < 0.001$, the training is thought to be close to the limit and should be stopped.

3.2.2 NARX ANN training with different output-feedback lengths

A sequence of NARX networks with different length L of the output feedback time-delay line (i.e. number of the context nodes) has been trained using the 67 storm intervals in order to find the most proper value of L and its relationship with the time scale of the ring current decay. The value of L ranges from one (corresponding to 5 min) to 1152 (4 days) with different time resolution, being taken as: 1, 2, 4, 6, 8, 10, ..., 24, 26, 64, 144, 288, 576, 1152.

3.2.3 Elman ANN training with different history length of input

All the 67 storm intervals are also trained using a series of Elman network with different history length of the external input to see the performance of the Elman network for SYM-H prediction. The history length of the input was taken as 90 min, 240 min, 480 min, and 720 min, separately.

4 Prediction testing results

We use the parameters of ρ , RMSE and ARV cited in Sect. 3.2.1 to characterize the network prediction performance.

4.1 SYM-H prediction by NARX neural network

4.1.1 The optimum length of the output feedback time-delay

As described in Sect. 3.2.2, a sequence of NARX networks with different length L of the output feedback time-delay line (i.e. number of the context nodes) has been trained in order to find the most proper value of L . All the trained NARX networks with different L are tested for the 6 storm events using IMF/SW data from ACE satellite. The test results are shown in Fig. 5, which gives the variation of the prediction performance (characterized by the three parameters of ρ , RMSE and ARV) with output-feedback delay time length of L .

It can be seen from Fig. 5 that the optimum value of L is 24 (about 120 min) when the prediction performance gets a maximum correlation coefficient of 0.95 between the observed and predicted SYM-H and a minimum RMSE of 17.12 nT for all the 6 testing events as a whole. It seems that the delay time length of the output feedback would be at least larger than 60 min so as to get fair prediction performance. If the length is less than 40 min, the prediction performance is poor. On the other hand, larger length about 1–2 days seems still to provide not so bad prediction performance.

4.1.2 Prediction results by using ACE data for 6 testing events

Table 2 lists on its right 3 columns the three parameters of prediction performance for each test event by the NARX neural network with a context node number of 24. The correlation coefficient between the observed and predicted SYM-H for each individual test event is more than 0.92 and the best one reaches about 0.98 with a RMSE as small as 7.2 nT. Especially, the two super-storms with minimum SYM-H less than -200 and -300 nT respectively are also accurately predicted, having correlation coefficients equal or larger than 0.95 and averaged relative variances less than 13.5%. Figure 6 displays a comparison of the observed SYM-H indices with predicted ones for each test event.

4.1.3 Validation of NARX model using more data (WIND, Geotail/OMNI)

In order to verify the prediction capability of NARX model, we have made further tests using IMF/SW data available from WIND satellite for another 7 great storms occurring in 1995–1997 and 2005, in which the time ahead of prediction is adjusted according to the position of WIND relative to ACE for each event.

Table 2. Prediction performance by using ACE data for 6 test events.

No.	Time period		Min. SYM-H (nT)	Prediction result		
	Start date	End date		ρ	RMSE (nT)	ARV
1	1999-10-21	1999-10-23	-221	0.948	21.84	0.132
2	2000-04-06	2000-04-09	-315	0.976	19.34	0.074
3	2005-05-07	2005-05-10	-115	0.927	13.52	0.159
4	2005-05-28	2005-05-31	-126	0.954	17.97	0.198
5	2005-08-31	2005-09-01	-116	0.979	7.22	0.047
6	2006-04-13	2006-04-15	-110	0.931	15.72	0.255

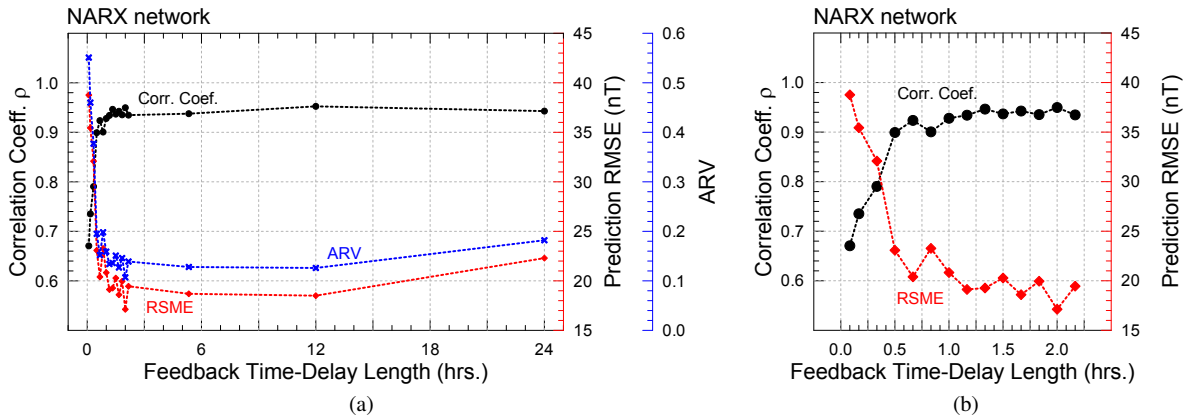


Fig. 5. Prediction performance by NARX network versus feedback delay time length of L . (a) For L -value from 0 to 24 h; (b) A zoom-in view of (a) for L -value from 5 to 130 min.

The prediction model is still the NARX network with $L = 24$ which is trained using IMF/SW data from ACE satellite.

Because WIND is nearer to the earth than ACE and the distance varies in a large range, the preact time of prediction is corrected for each event according to the distance between WIND and the earth at the event occurrence. Table 3 lists the prediction results of the 7 events. These test results show again that the NARX model has good performance in prediction of SYM-H. As a whole of the 7 events, the three performance parameters have the values of $\rho = 0.91$, RMSE= 14.6 nT, ARV= 18.0%. The best one has the performance parameters of $\rho = 0.966$, RMSE= 8.58 nT and ARV= 7.3%, as seen in Table 3. Figure 7 shows the SYM-H index predicted by NARX neural network with 24 context units in comparison with the observed SYM-H for the 7 testing storm events whose IMF/SW data are from WIND satellite measurements.

Besides, in order to validate our NARX model, we have also tried to predict the SYM-H index for the superstorms of July 2000 and 24 November 2001 using IMF/SW data from Geotail and OMNI at 1 AU, respectively. Data from Geotail and OMNI can be treated as from dayside magnetopause, so the prediction using these data should be of short time ahead. In the testing, the prediction is one

point (5 min) ahead. Figure 8 gives the prediction results, which also show acceptable good performance as: $\rho = 0.89$, RMSE= 37.3 nT, ARV= 23.5% for July 2000 storm, while $\rho = 0.91$, RMSE= 30.2 nT, ARV= 30.2% for 24 November 2001 storm.

The prediction results for the 9 events using IMF/SW data from rather than ACE satellite indicate that the NARX prediction model we established has potential capability to predict SYM-H index using input data of IMF/SW from other sources.

4.2 Comparison with Elman neural network

To make clear if the Elman network is also effective in SYM-H prediction as it is in Dst prediction, we have made a series training of the Elman networks with different input time delay lines of 90, 240, 480, 720 min (i.e., 1.5 h, 4.0 h, 8.0 h, 12 h), using ACE IMF/SW data for the 67 storms. And, a series of tests are then carried out for the 6 great storms.

Table 4 gives the prediction performance of Elman network with different input time delay lines. When using 90 min length of IMF/SW wind data as input, the testing result shows that the correlation coefficient between the predicted and the observed SYM-H is 0.682, RMSE is 38.54 nT

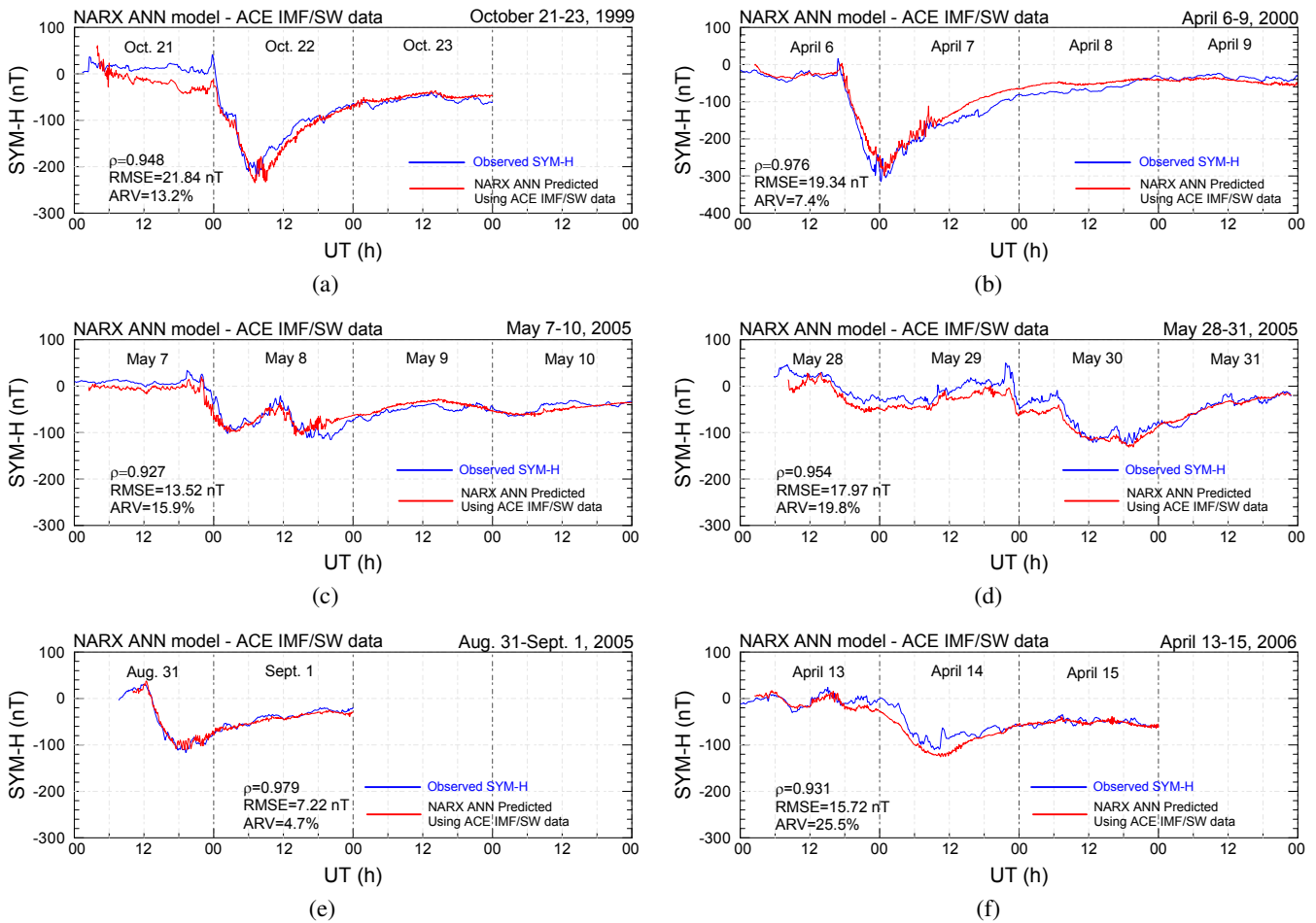


Fig. 6. Comparison of observed SYM-H index with that predicted by NARX neural network with output-feedback delay time length of 120 min for testing storm events using ACE data.

Table 3. Prediction result of each test event whose IMF/SW data are from WIND.

No.	Time period		Min. SYM-H (nT)	Prediction result		
	Start date	End date		ρ	RMSE (nT)	ARV
1	1995-03-25	1995-03-29	-99	0.886	12.96	0.223
2	1995-04-06	1995-04-10	-163	0.943	14.65	0.157
3	1995-10-18	1995-10-20	-125	0.963	12.79	0.096
4	1997-04-20	1997-04-23	-100	0.882	13.95	0.366
5	1997-11-06	1997-11-09	-125	0.874	13.57	0.273
6	1997-11-21	1997-11-25	-121	0.966	8.58	0.073
7	2005-08-23	2005-08-27	-174	0.877	21.45	0.262
As a whole				0.91	14.6	0.18

and AVR is 54.4% for 6 test examples as a whole, which is much worse than the prediction result of NARX model.

It is also shown from Table 4 that when the input time length increases from 90 min to 240 min, the SYM-H prediction performance of the Elman ANN improves obviously

from $\rho = 0.68$, RMSE= 38.5 nT, ARV= 54.4% to $\rho = 0.86$, RMSE= 26.6 nT, ARV= 26.9%. If the input time length continues to increase the SYM-H prediction performance changes very slowly. When it reaches 720 min long, the SYM-H prediction performance parameters for the 6 storms

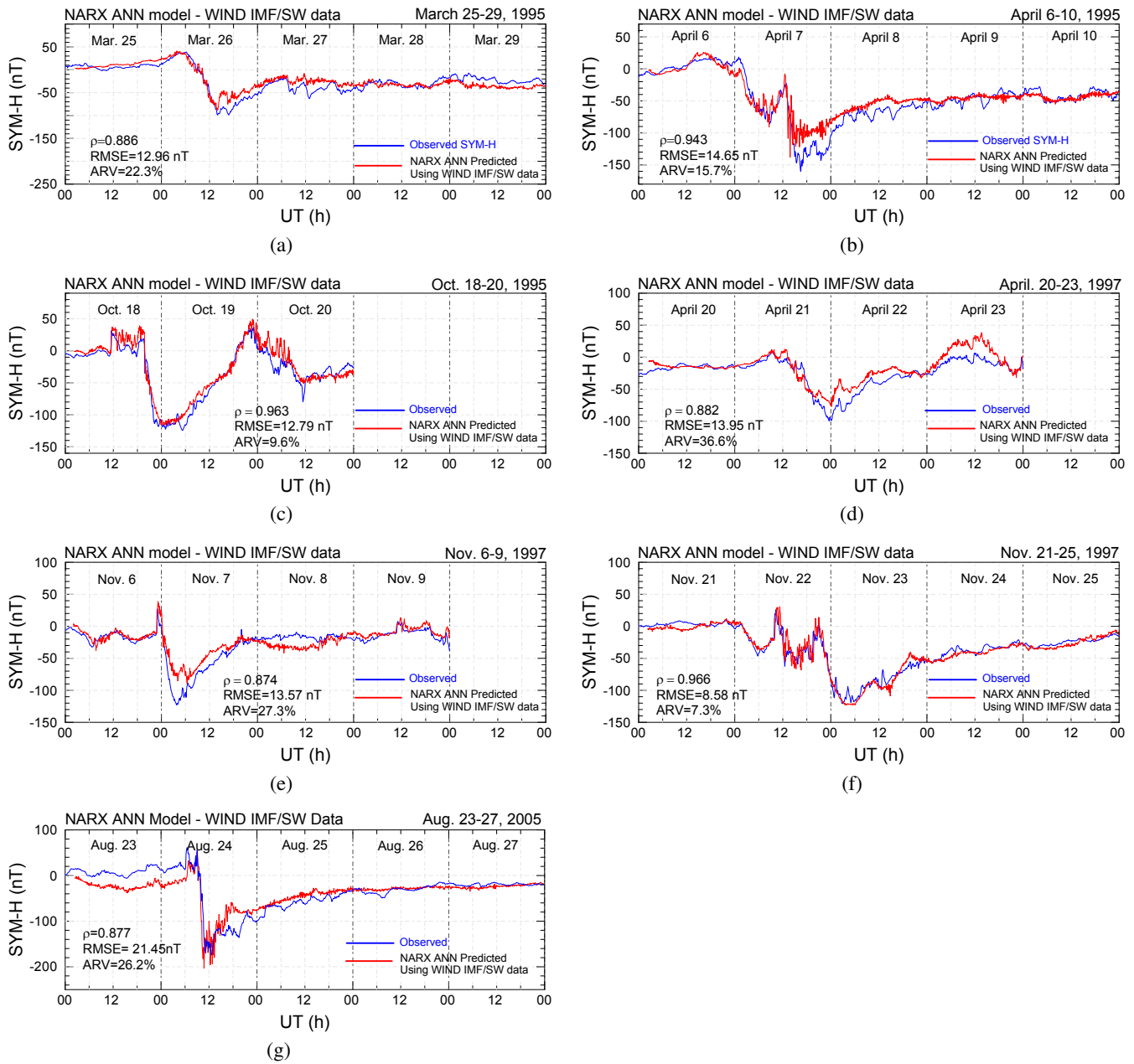


Fig. 7. Comparison of observed SYM-H index with that predicted by NARX neural network with 24 context units for each testing storm event whose IMF/SW data are from WIND.

as a whole increased up to: $\rho = 0.86$, RMSE= 26.2 nT, ARV= 26.2%, which is still not better than NARX model ($\rho = 0.95$, RMSE= 17.1 nT, ARV= 10.7%). Besides, longer input time length degraded greatly the applicability when only short time qualified IMF/SW data are available (see Fig. 9e).

Figure 9 displays a comparison of observed SYM-H index with that predicted by Elman neural network with 720 min length of IMF/SW input data for each testing storm. The

SYM-H indices predicted by NARX model with 90 min length of IMF/SW input data are also displayed by green curves for a comparison of the two models. Figure 10 gives the spot plots of the model predicted SYM-H versus observed ones for NARX model and two Elman models with different input time lengths. Figures 9 and 10 show clearly that the NARX model does exceed the Elman network in SYM-H prediction.

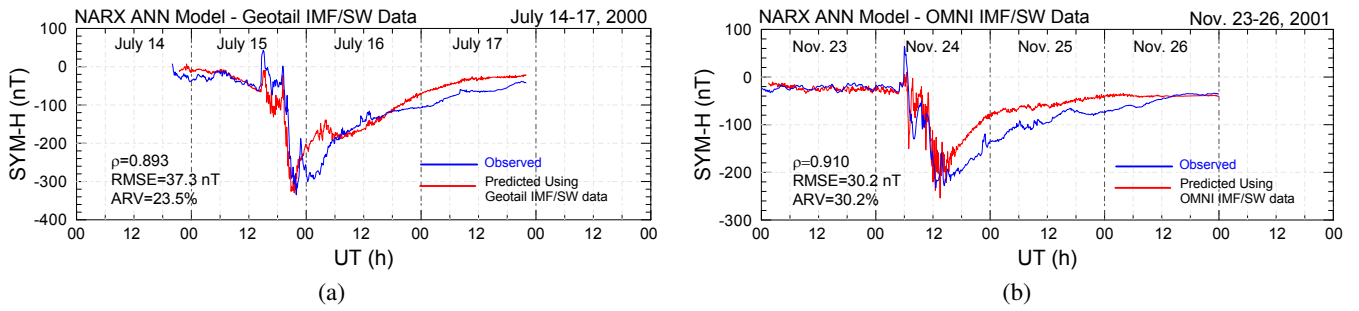


Fig. 8. Comparison of observed SYM-H index with that predicted by NARX neural network with 24 context units for testing storm events whose IMF/SW data are from Geotail and OMNI.

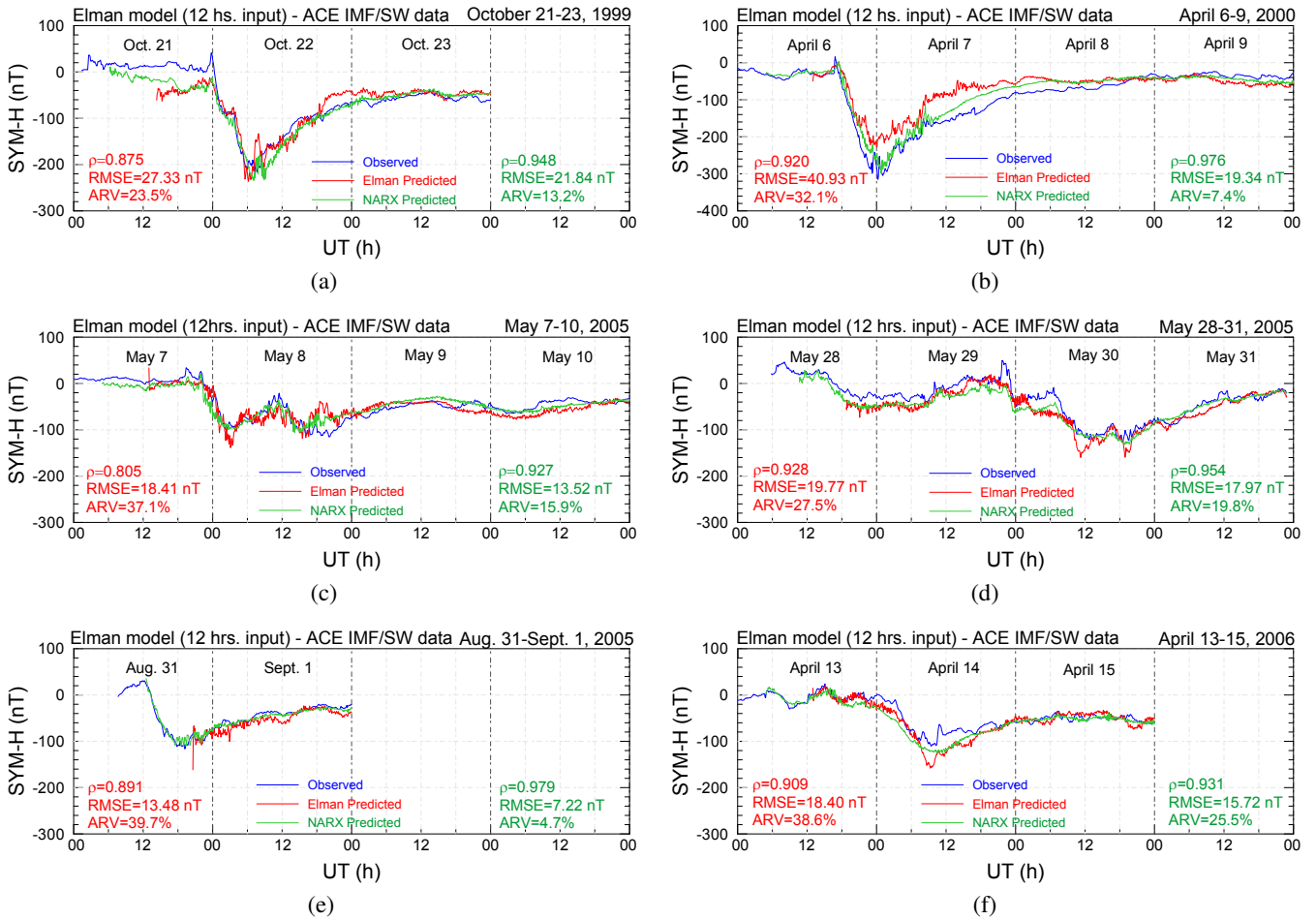


Fig. 9. Comparison of observed SYM-H index with that predicted by Elman neural network with 720 min length history IMF/SW parameters as input for each testing storm event whose IMF/SW data are from ACE. The SYM-H indices predicted by NARX model are also plotted in the corresponding figures by green lines for a comparison of the two models.

Thus, the classical Elman ANN is not effective in SYM-H prediction as it is in Dst prediction. The cause of it could be attributed to that as below. In general, it is not wrong that the Elman network can catch more complex processes

than the NARX due to the connections from the hidden layer. However, the classical Elman network makes feedback from hidden layer to input only one step, which is of 5 min scale for SYM-H index in this work and thus surely insufficient

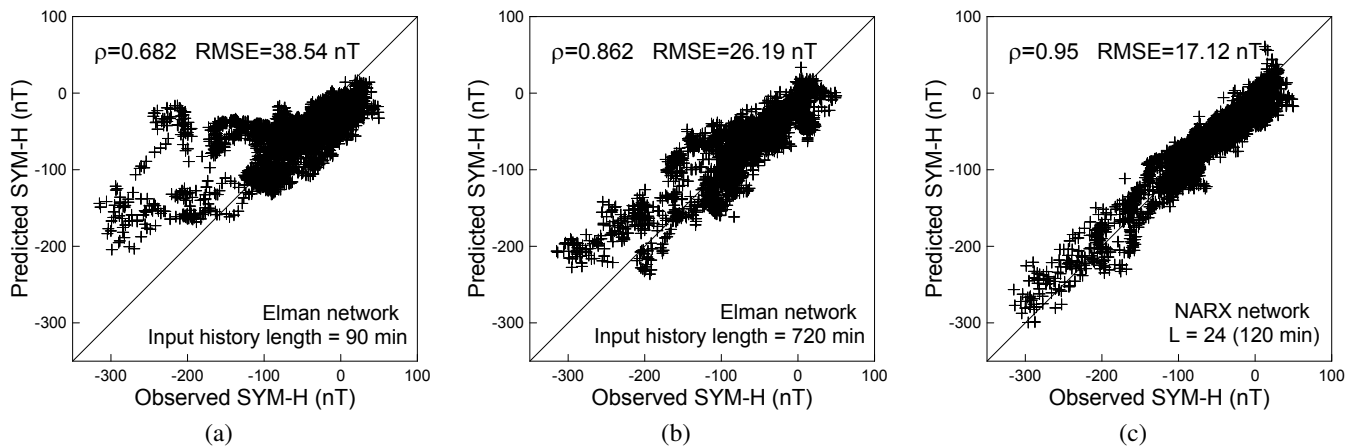


Fig. 10. Correlation plots of the SYM-H indices observed and predicted by Elman and NARX neural networks. **(a)** For the Elman model with input history length of 90 min; **(b)** For the Elman model with input history length of 720 min; **(c)** For NARX model with feedback time delay length of 120 min.

Table 4. Prediction result by Elman ANN with different length of input history.

Input history (Mins)	Prediction result		
	ρ	RMSE (nT)	ARV
90	0.682	38.54	0.544
240	0.812	30.20	0.340
480	0.855	26.60	0.269
720	0.862	26.19	0.263

to catch the characteristic RC decay time that is at least tens minutes. For Dst prediction by Elman network this feedback of one step means a time scale of 60 min which is some extent enough to match the characteristic RC decay time. It should be noted also that the RC decay processes associated with RC ion losses persistently occur during both of the storm main and recovery phases rather than only the recovery phase. So, the insufficiency of classical Elman ANN in catching the characteristic RC decay time would deteriorate the prediction performance of SYM-H for both of main and recovery phases of the storms.

However, the Elman neural network has potential capability to catch more complex processes due to the connections from the hidden layer to the output. It is possible that a modified Elman ANN may become powerful as NARX by using new techniques like pruning method etc.

4.3 Comparison with previous Dst prediction by ANN

A practical comparison between the SYM-H prediction performances of our NARX model with previous ones is indeed hard to be made, since the prediction of SYM-H index has

ever not been reported before. However, studies of Dst prediction by neural network has been conducted for more than ten years, among which, a detailed study on the prediction of Dst by Elman network is made by Wu and Lundstedt (1997). A comparison between their Dst prediction results and our SYM-H prediction results by NARX model will be made as follows.

In the study by Wu and Lundstedt (1997), IMF/SW data with history length of 1, 4, 8 and 12 h are taken as input to predict the variation of Dst index and a quite accurate prediction is achieved. Table 5 shows the comparison between their Dst prediction result and our SYM-H prediction result by NARX model, from which we can find that the accuracy of SYM-H index prediction by NARX model is comparable with that of the prediction of Dst by Elman neural network.

5 Discussion and conclusions

In this study, a new kind of neural network has been introduced for the first time to predict storm time SYM-H index. Its prediction capability is examined and compared with Elman network and previous work for Dst prediction. The NARX model introduced in this paper shows much better capability than Elman network for SYM-H prediction, which can partly be attributed to a key feedback to the input layer from the output neuron with a suitable time delay length. An extensive work of training and testing shows that when the length of the output feedback sequence (i.e., the context unit number) L is 24 corresponding to a time length of 120 min, the NARX network gets the best performance.

It shows that although Elman network can make very excellent prediction of Dst, it is not able to predict the variation of SYM-H index with satisfied high accuracy at least in its present frame. The prediction errors are particularly large

Table 5. Comparison of Dst prediction by Elman network with SYM-H by NARX ANN.

Model	IMF/SW data source	Input history (Mins)	Prediction result		
			ρ	RMSE (nT)	ARV
Wu & Lundstedt Dst prediction model	OMNI	60	0.90	15.5	0.19
		240	0.92	14.1	0.16
		480	0.92	13.8	0.15
		720	0.92	14.1	0.15
SYM-H prediction model of NARX	ACE	90	0.95	17.1	0.11
	WIND	90	0.91	14.6	0.18

near the maximum of the main phase (minimum SYM-H) and for the recovery phase. This may be attributed to a lack of the inner status information of the ring current in the network input. The input of the BP network is only the parameters of outer space containing nothing on the information of inner status of the magnetosphere, while the context units of the Elman network fed back from the hidden neurons contain the history information of inner status only 1 step, i.e. 5 min in this study. This length is too short to catch the characteristic time scale of RC decay that is at least tens of minutes.

The development, evolution and decay of the ring current are controlled by both aspects of solar wind driving and magnetospheric internal physical status. The solar wind driving dominates the initial and main phase of the storm, which can be somewhat easily implemented by taking their key parameters with proper history length as the external input of the network. The internal status of the magnetosphere affect all the course of RC growth, evolution and decay, especially dominate the recovery phase of the storm. How to effectively bring the internal status information of the ring current to take part in the prediction is the key point to promote the prediction performance. The global feedback in NARX network developed in this study directs the predicted SYM-H which contains the nearly real ring current information into the external input, thus provides an effective way to get this point. Further, how to shape the feedback to contain necessary and sufficient information for a good prediction becomes an important problem. Here the time scale of the feedback should match the characteristic time of the ring current decay. The time scale of the ring current decay ranges from a few tens of minutes to more than tens of hours (Gonzalez et al., 1989; Daglis, 2001). Too short delay time of the feedback, during which the ring currents do not get change significantly, would contribute little to the prediction. An amount of trials (taking the L from 5 min to 24 h or more) indicates that the output-feedback length seems to be at least larger than about 60 min so as to get fair prediction performance, while it does also using feedback length as long as 1–2 days. The proper history length of the output-feedback in our NARX model is about 120 min. This length may mainly reflect the characteristic time of ring current decay on average and weighted

especially for the early recovery phase, which involves various decay mechanisms with ion lifetimes from tens of minutes for pitch angle scattering by EMIC wave-particle interactions (Jordanova et al., 1997, 2001) to tens of hours or more for charge-exchange and Coulomb collision loss etc. This proper output-feedback delay time length may somewhat change with the collected storm samples, because different type of individual magnetic storm would have different characteristic evolution and decay time.

Different combinations of solar wind and IMF parameters for input and different time length of input parameters would affect the prediction performance of the NARX network, which have not been verified yet in this study. So the present NARX network and its prediction result are not necessarily the most optimum and the best. They will be examined in an accompanying investigation. However, our study has proved the great advantage of the NARX neural network in prediction of storm-time SYM-H index.

The Elman neural network has potential capability to catch more complex processes due to the connections from the hidden layer to the output. It might be true in principle that the Elman neural network could make storm prediction at least as well as NARX even for higher resolution data. But it needs major improvement by using new techniques like pruning method etc. It is worthy of optimizing Elman network and other ANNs in SYM-H prediction with some new techniques in the further study.

Acknowledgements. The authors thank ISTP PIs and teams for providing ACE/WIND/GEOTAIL satellite data of IMF/SW parameters and magnetic indices through internet. This study is supported by Doctoral Fund of Ministry of Education of China (Grant No. 200804860012).

Topical Editor I. A. Daglis thanks J. Wanliss and another anonymous 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, *Planet. Space Sci.*, 29, 1151–1158, 1981.

- Amata, E., Palloccchia, G., Consolini, G., Marcucci, M. F., and Bertello, I.: Comparison between three algorithms for Dst predictions over the 2003–2005 period, *J. Atmos. Solar-Terr. Phys.*, 70, 496–502, 2008.
- Baker, D. N.: Statistical analyses in the study of solar wind/magnetosphere coupling, in: *Solar Wind-Magnetosphere Coupling*, vol. 17, edited by: Kamide, Y. and Slavin, J. A., D. Reidel Publishing Company, Dordrecht, Holland, 1986.
- 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.
- Daglis, I. A.: Ring current dynamics, *Space Sci. Rev.*, 124, 183–202, doi:10.1007/s11214-006-9104-z, 2006.
- Daglis, I. A.: The storm-time ring current, *Space Sci. Rev.*, 98, 343–363, 2001.
- Daglis, I. A.: The role of magnetosphere-ionosphere coupling in magnetic storm dynamics, in: *Magnetic Storms*, edited by: Tsurutani, B. T. and Gonzalez, W. D., *Geophys Monogr, Ser.*, 98, Washington, AGU, 107–116, 1997.
- Elman, J. L.: Finding structure in time, *Cognitive Sci.*, 14, 179–211, 1990.
- Gleisner, H., Lundstedt, H., and Wintoft, P.: Predicting geomagnetic storms from solar-wind data using time-delay neural networks, *Ann. Geophys.*, 14, 679–686, 1996, <http://www.ann-geophys.net/14/679/1996/>.
- Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, 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 Gonzalez, A. L. C.: Solar wind-magnetosphere coupling during intense magnetic storms (1978–1979), *J. Geophys. Res.*, 94(A7), 8835–8851, 1989.
- Gosling, J., McComas, D., Phillips, J., and Bame, S.: Geomagnetic Activity Associated with Earth Passage of Interplanetary Shock Disturbances and Coronal Mass Ejections, *J. Geophys. Res.*, 96(A5), 7831–7839, 1991.
- Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 2nd Edition, Prentice-Hall Inc., USA, 1999.
- Jordanova, V. K., Farrugia, C. J., Thorne, R. M., Khazanov, G. V., Reeves, G. D., and Thomsen, M. F.: Modeling ring current proton precipitation by electromagnetic ion cyclotron waves during May 14–16, 1997 storm, *J. Geophys. Res.*, 106(A1), 7–22, 2001.
- Jordanova, V. K., Kozyra, J. U., and Nagy, A. F.: Kinetic model of the ring current-atmosphere interaction, *J. Geophys. Res.*, 102(A7), 14279–14291, 1997.
- Lundstedt, H., Gleisner, H., and Wintoft, P.: Operational forecasts of the geomagnetic Dst index, *Geophys. Res. Lett.*, 29(24), 2181, doi:10.1029/2002GL016151, 2002.
- Mayaud, P. N.: *Derivation, Meaning, and Use of Geomagnetic Indices*, *Geophysical Monograph 22*, American Geophysical Union, Washington D.C., 1980.
- McPherron, R. L., Bargatze, L. F., Holzer, R. E., Baker, D. N., and Clauer, C. R.: IMF control of geomagnetic activity, *Adv. Space Res.*, 8, 71–86, 1988.
- Palloccchia, G., Amata, E., Consolini, G., Marcucci, M. F., and Bertello, I.: Geomagnetic D_{st} index forecast based on IMF data only, *Ann. Geophys.*, 24, 989–999, 2006, <http://www.ann-geophys.net/24/989/2006/>.
- Rangarajan, G. K.: Indices of geomagnetic activity, in: *Geomagnetism*, edited by: Jacobs, J. A., Academic Press, London, UK, 323, 1989.
- Ridley, A. J. and Liemohn, M. W.: A model-derived storm time asymmetric ring current driven electric field description, *J. Geophys. Res.*, 107(A8), 1151, doi:10.1029/2001JA000051, 2002.
- Rostoker, G. and Fälthammar, C. G.: Relationship between changes in the interplanetary magnetic field and variations in the magnetic field at the earth's surface, *J. Geophys. Res.*, 72, 5835–5863, 1967.
- Sugiura, M. and Poros, D. J.: Hourly values of equatorial Dst for the IGY, *Ann. Int. Geophys. Year*, 35, 9–45, 1964.
- 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.
- Tsurutani, B., Gonzalez, W., Tang, F., Akasofu, S. I., and Smith, E.: Origin of Interplanetary Southward Magnetic Fields Responsible for Major Magnetic Storms Near Solar Maximum (1978–1979), *J. Geophys. Res.*, 93(A8), 8519–8531, 1988.
- 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.
- Wanliss, J. A. and Showalter, K. M.: High-resolution global storm index: Dst versus SYM-H, *J. Geophys. Res.*, 111, A02202, doi:10.1029/2005JA011034, 2006.
- Wu, J. and Lundstedt, H.: Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks, *J. Geophys. Res.*, 102, 14255–14268, 1997.
- Wu, J. and Lundstedt, H.: Prediction of geomagnetic storms from solar wind data using Elman recurrent neural network, *Geophys. Res. Lett.*, 23(4), 319–322, 1996.
- Yermolaev, Yu. I., Yermolaev, M. Yu., 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.
- Zhou, Y. L., Ma, S. Y., Luehr, H., Xiong, C., and Reigber, C.: An empirical relation to correct storm-time thermospheric mass density modeled by NRLMSISE-00 with CHAMP satellite air drag data, *Adv. Space Res.*, 43, 819–828, doi:10.1016/j.asr.2008.06.016, 2009.