

Auroral electrojet predictions with dynamic neural networks

Hans Gleisner

Lund Observatory, Lund, Sweden

Henrik Lundstedt

Swedish Institute of Space Physics, Solar-Terrestrial Physics Division, Lund, Sweden

Abstract. Neural networks with internal feedback from the hidden nodes to the input [Elman, 1990] are developed for prediction of the auroral electrojet index AE from solar wind data. Unlike linear and nonlinear autoregressive moving-average (ARMA) models, such networks are free to develop their own internal representation of the recurrent state variables. Further, they do not incorporate an explicit memory for past states; the memory is implicitly given by the feedback structure of the networks. It is shown that an Elman recurrent network can predict around 70% of the observed AE variance using a single sample of solar wind density, velocity, and magnetic field as input. A neural network with identical solar wind input, but without a feedback mechanism, only predicts around 45% of the AE variance. It is also shown that four recurrent state variables are optimal: the use of more than four hidden nodes does not improve the predictions, but with less than that the prediction accuracy drops. This provides an indication that the global-scale auroral electrojet dynamics can be characterized by a small number of degrees of freedom.

1. Introduction

Auroral electrojet activity, as measured by the geomagnetic indices AU , AL , and AE , depends on both the external solar wind forcing and the internal dynamics of the magnetosphere. The important roles played by the solar wind density, velocity, and magnetic field have been extensively studied during the last 30 years, first using statistical, correlative methods [e.g., Arnoldy, 1971; Murayama and Hakamada, 1975], and later using linear filters [e.g., Iyemori et al., 1979; Clauer et al., 1981]. The studies based on linear filters have demonstrated the fundamental nonlinearity of the geomagnetic response to the solar wind [Bargatze et al., 1985] and have contributed to a better understanding of the timescales involved.

A consequence of the findings by Bargatze et al. is that accurate predictions of geomagnetic activity from solar wind data require nonlinear methods. Vassiliadis [1993], Price et al. [1994], and Vassiliadis et al. [1995] used a locally linear, but globally nonlinear, filter technique to predict the AE and AL indices. These nonlinear filters were developed from ideas put forward by Casdagli [1992] and Hunter [1992]. The range of applications for the nonlinear filters have later been extended

to the ring current index Dst [Klimas et al., 1998] and to spatial patterns of geomagnetic disturbances at high latitudes [Valdivia et al., 1999].

Neural networks is a related technique that relatively recently has found a wide range of applications. Different type of neural networks have been developed for prediction of geomagnetic indices. Time delay networks (TDNs), taking a sequence of time-lagged solar wind data as input, were used by Lundstedt and Wintoft [1994] and Gleisner et al. [1996] to predict Dst 1 hour ahead. Later studies by Wu and Lundstedt [1996, 1997] have shown that Elman recurrent networks (ERNs) can predict the Dst index from a single sample of solar wind data, i.e. with no explicit reference to time-lagged solar wind inputs.

Hernandez et al. [1993] published a study on AL predictions where they used two types of neural networks: a nonlinear moving-average (MA) filter and a nonlinear autoregressive moving-average (ARMA) filter, the former being identical to a TDN. This work was extended by Weigel et al. [1999] to address certain problems due to clipping of high-amplitude variations. Gleisner and Lundstedt [1997] presented a study on AE predictions using TDNs. It was shown that up to 100 min of solar wind data is required and that the use of raw solar wind parameters gives more accurate AE predictions than any of the most common coupling functions (i.e., a function that combines several solar wind parameters into a single quantity). Takalo and Timonen [1997] also

Copyright 2001 by the American Geophysical Union.

Paper number 2001JA900046.
0148-0227/01/2001JA900046\$09.00

studied *AE* predictions using nonlinear ARMA filters fed with a time-lagged sequence of solar wind data and *AE* itself.

The two types of recurrent neural networks that have been used to predict geomagnetic indices are thus nonlinear ARMA filters and Elman recurrent networks. Unlike nonlinear ARMA filters, ERNs are free to develop their own representation of the states that are fed back to the input. Further, they do not incorporate an explicit memory for past states, other than the most recent. The memory is implicitly given by the feedback structure of the network. If fed with a single sample of solar wind data, any dynamic behavior of an ERN must be the result solely of the network's feedback structure. The studies by *Wu and Lundstedt* [1996, 1997] show that ERNs have the ability to approximate at least a part of the underlying dynamics of the solar wind-*Dst* relation. The aim of the present study is to evaluate the abilities of ERNs to model the dynamics of the solar wind-*AE* relation.

2. Neural Networks

2.1. Nonlinear Dynamic Mappings

At the most general level the geomagnetic activity, O , can be described as a function of a vector, \mathbf{I} , of time-lagged solar wind inputs

$$O_t = F(\mathbf{I}_{t-1}), \quad (1)$$

where

$$\mathbf{I}_{t-1} \equiv \{I_{t-1}, I_{t-2}, \dots, I_{t-T_I}\} \quad (2)$$

and T_I is the temporal length of \mathbf{I} . A more powerful assumption is that the geomagnetic activity depends on both the solar wind input and prior geomagnetic activity

$$O_t = F(\mathbf{I}_{t-1}, \mathbf{O}_{t-1}), \quad (3)$$

where

$$\mathbf{O}_{t-1} \equiv \{O_{t-1}, O_{t-2}, \dots, O_{t-T_O}\} \quad (4)$$

and T_O is the temporal length of \mathbf{O} . Alternatively, it is a set of internal states, \mathbf{Y} , that is fed back to the input

$$O_t = F(\mathbf{I}_{t-1}, \mathbf{Y}_{t-1}), \quad (5)$$

where \mathbf{Y}_{t-1} is a vector containing a number of internal state variables at one instance of time, namely $t-1$. An interesting difference between the feedback mechanisms of (3) and (5) is that in the former equation the time dimension is explicitly represented by a memory for past outputs. In (5), there is no explicit memory for past states (other than the most recent) and time is only represented implicitly by its effects on processing.

Equations (1) to (5) represent low-dimensional magnetospheric dynamics at the most general level. The simplest possible realizations are the linear MA and ARMA filters. For a magnetosphere with linear response properties these filters would be perfectly sufficient. However, as the study by *Bargatze et al.* [1985] pointed out, the response properties vary with the level

of geomagnetic activity. The magnetospheric response is in fact nonlinear.

One way of introducing nonlinearities is to approximate F locally by linear MA or ARMA filters whose response properties vary between different regions of the input-state space. Such locally linear but globally nonlinear filters have been used by *Vassiliadis et al.* [1993, 1995] to study the auroral electrojet response to the solar wind, and the technique has also been described by *Klimas et al.* [1996] in a broader context of nonlinear magnetospheric dynamics.

Neural networks have much in common with the nonlinear filters, but they also differ from them in important respects. For a neural network the function F is both locally and globally nonlinear. The time delay network (Figure 1, and further described in section 2.2), with a sequence of time-lagged solar wind data as input, is essentially a realization of (1). An Elman recurrent network (Figure 2, and further described in section 2.3), with feedback of internal network activations to a set of context nodes is a realization of (5).

2.2. Time Delay Networks

Standard feed-forward neural networks are described thoroughly in many text books [e.g., *Haykin*, 1999]. Time delay networks are simply feed-forward networks that are fed with a temporal sequence of time-lagged external inputs [*Lapedes and Farber*, 1987; *Waibel*, 1989]. It is the organization of the input data that gives the TDN a dynamic behaviour; the mapping itself is static. The processing performed on the input data is given by

$$O^\mu = g_o \left[\sum_{j=1}^{N_{hid}} W_j g_h \left(\sum_{k=0}^{N_{in}} w_{jk} I_k^\mu \right) + B I_0 \right], \quad (6)$$

where the input-output samples $\{I_k^\mu, O^\mu\}$ are labeled by superscript μ . Index j refers to a hidden node and index k refers to an input node. The bias input I_0 is assigned a fixed value and is connected to all hidden and output nodes in the network through a set of bias weights. In the present study the activation functions for nodes in the hidden and output layers are

$$g_h(x) = \tanh(x); \quad g_o(x) = x. \quad (7)$$

The incoming signals at a hidden node are thus processed by a nonlinear, differentiable, saturating function, and the processing performed at the output node is simply a linear, weighted summation.

The number of input and output nodes is determined by the set of data to which the network is applied, while the number of hidden nodes, N_{hid} , essentially is a free parameter. The factors that should be considered when determining N_{hid} are briefly discussed in section 4.1.

2.3. Elman Recurrent Networks

Partially recurrent networks often incorporate the essential features of time delay networks, but they also

include a limited set of fixed feedback connections [Jordan, 1989; Cleeremans et al., 1989; Elman, 1990]. The recurrent networks devised by Elman [1990] feed back the hidden-node activations to a set of context nodes at the input. Unlike the TDN, the dynamical properties of an ERN is a result of the mapping itself being dynamic. The processing performed by an ERN is given by

$$O_t^\mu = g_o \left[\sum_{j=1}^{N_{\text{hid}}} W_j g_h \left(\sum_{k=0}^{N_{\text{in}}} w_{jk} I_k^\mu + \sum_{c=1}^{N_{\text{con}}} w_{jc} Y_{t-1}^c \right) + B I_0 \right], \quad (8)$$

where index c denotes the context units and t is a time variable. The other symbols are identical to those of (6). Similar to a TDN, the external input data to the ERN are organized as a temporal sequence of time-lagged data. This sequence is normally relatively short. It often consists of only the current input data, i.e., the ERN is not fed with any time-lagged input data at all. The activation functions are the same as for the TDN described above, nonlinear at the hidden nodes and linear at the output node.

The actual feedback of an ERN is performed by copying the hidden node activations onto the context nodes. This means that at time t the context input consists of the hidden node activations at time $t-1$ (see (8) and Figure 2). No weights are associated with the feedback connections, which are kept fixed during the training process. This is an important property of a recurrent network. As the modifiable connections are purely feed-forward, we can use the same training algorithm as for a TDN.

2.4. Network Training

A network's ability to produce a "correct" output is quantified by

$$C(\mathbf{w}) \equiv \frac{1}{2} \sum_{\mu=1}^{Q_{\text{trn}}} (O^\mu - T^\mu)^2, \quad (9)$$

where \mathbf{w} is the set of weights, O^μ is the actual output of the network, T^μ is the "correct" output (or target), and Q_{trn} is the number of input-output samples in a set of training data. Network training is the process of optimizing the cost function, $C(\mathbf{w})$, under certain restrictions. In the present study we have used the error back-propagation algorithm [Rumelhart et al., 1986]. The weights are iteratively updated according to the rule

$$\Delta w_i \leftarrow -\eta \left(\frac{\partial C}{\partial w} \right)_i + \alpha \Delta w_{i-1} \quad (10)$$

$$w_{i+1} \leftarrow w_i + \Delta w_i \quad (11)$$

where w is a single weight and subscript i denote the iteration. Normally, it is only a subset of the Q_{trn} training samples that is used in each iteration, and the actual update is in an approximate gradient direction. The size, Q_{bat} , of this subset is a parameter that along with

η and α , controls the training process. For TDNs the Q_{bat} samples in the training batch are selected wholly randomly from the training set. For ERNs, intervals of data, rather than individual samples, are randomly selected.

In the present study the parameters that control the training process have been assigned the values

$$Q_{\text{bat}} = \begin{cases} 800 & \text{TDNs} \\ 3000 - 10000 & \text{ERNs} \end{cases}, \quad (12)$$

$$\eta = \frac{0.012}{Q_{\text{bat}}}, \quad (13)$$

$$\alpha = 0.90. \quad (14)$$

The varying batch size Q_{bat} for the ERNs is a consequence of the fact that the batch consists of a fixed number of data intervals, rather than a fixed number of samples. Further, the training process is more unstable for ERNs than for TDNs, requiring either a smaller learning rate η or a larger batch size Q_{bat} . We have chosen the latter.

Much of the practical use of neural networks relies on their ability to make sensible generalizations. This ability can be defined as the average performance on a randomly chosen data sample. However, the cost function $C(\mathbf{w})$ measures a network's ability to memorize the training data, rather than the ability to generalize to new data. In order to optimize the generalization ability, the training procedure need to be constrained. This is done by excluding a small part of the training set from the actual training, and use these data to determine when to stop the iteration. In this way the problem of overfitting is avoided, or at least lessened.

3. Data Set

The data were obtained from the Bargatze data set, originally compiled for linear MA filter studies [Bargatze et al., 1985]. These data, or subsets of them, have been used in several studies of magnetospheric dynamics [e.g., Hernandez et al., 1993; Vassiliadis et al., 1995; Takalo and Timonen, 1997; Weigel et al., 1999]. A similar data set, largely overlapping in time but with a lower time resolution (5 min instead of 2.5 min), was used in the study by Gleisner and Lundstedt [1997].

The data set contains solar wind data from the Earth-orbiting spacecraft IMP 8 observed between November 1973 and December 1974: velocity V , proton number density n , and the interplanetary magnetic field components B_x , B_y , and B_z given in Geocentric Solar Magnetospheric (GSM) coordinates. The data set also includes the geomagnetic activity index AE at a time resolution of 2.5 min. The AE data have been time shifted to account for the solar wind travel time from the IMP 8 position to the magnetopause. All parameters have further been linearly scaled to be distributed within the interval $[-1.0, +1.0]$.

The data are divided into 34 intervals covering 42,216 samples. Each interval contains isolated auroral activity

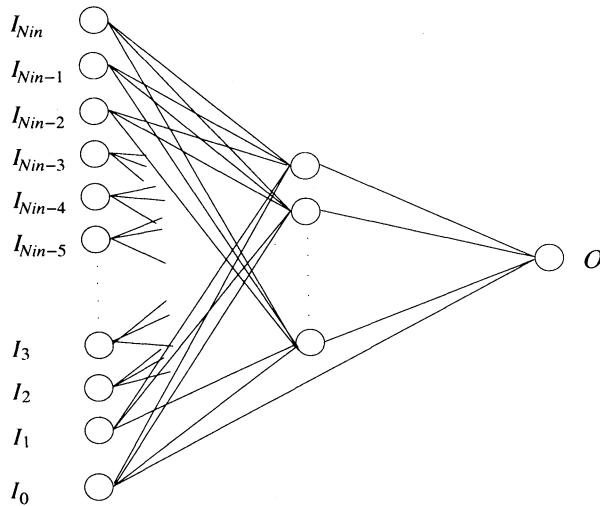


Figure 1. Time delay network (TDN) with a sequence of solar wind data to the left and a single output node to the right. The processing performed at the hidden nodes is nonlinear, while the output node performs a linear, weighted summation.

preceded and followed by relatively quiet periods. The intervals are ordered from low to high activity, such that interval 1 is very quiet and interval 34 very disturbed. Every third interval is used as test data to ensure that the networks are tested on a variety of activity levels. The test data are not used during training; their only role is to evaluate the network performance.

4. Studies

4.1. Network Setup: Input Parameters and Network Size

Neural networks are completely data based in the sense that the model parameters are determined solely from the data available during the training process. Physics enter the neural network model through the choice of physical quantities to use as input and output data. The input data should contain a maximum amount of information on the solar wind but should also exclude all data that are of no relevance for the geomagnetic activity.

Another consideration is the risk for overfitting, which was briefly addressed by *Gleisner and Lundstedt* [1997]. The overfitting problem increases with the number of free parameters in the network, i.e., both with the number of input data and with the number of hidden nodes. The negative effects of overfitting can be reduced by a proper choice of training procedure, but they cannot be completely compensated for. An excessive number of weights tends to decrease the performance of any neural network.

The number of weights in the network, and thus the number of input data, should be as small as possible, while still be large enough to represent the full complexity of the problem. This should favour the use of solar wind coupling functions. However, previous stud-

ies have shown that it is better to use the raw solar wind parameters n , V , and B_z , than to use any of the most common coupling functions constructed from these parameters [*Gleisner and Lundstedt*, 1997], presumably because a loss of relevant information when combining separate solar wind parameters into a single quantity.

In the present studies we have used the solar wind parameters n , V , and B_z as input to the networks. The networks have 10 hidden nodes and a single output node. For an ERN the number of context nodes, N_{con} , is identical to the number of hidden nodes, N_{hid} . The basic ERN, which only use a single solar-wind sample as input, have 3 input nodes in addition to the 10 context nodes, one each for n , V , and B_z

$$\mathbf{I}_t = \{n_t, V_t, B_{z,t}\}.$$

The basic TDN, which take a 100-min sequence of solar wind data as input, have a total number of 120 input nodes, 40 each for n , V , and B_z

$$\mathbf{I}_t = \{n_t, \dots, n_{t-39}, V_t, \dots, V_{t-39}, B_{z,t}, \dots, B_{z,t-39}\},$$

In addition to the solar wind inputs, each network also have a bias input connected to all hidden and output nodes as shown in Figure 1 and Figure 2.

4.2. Length of the Solar Wind Input Sequence

It is the organization of the input data that gives the TDN a dynamic behavior; the mapping itself is static. On the basis of a time-lagged vector of past solar wind inputs the TDN can display such properties as delayed response and modulation of the response properties by prior inputs. For a TDN it is crucial that the temporal size T_I of the time-delay line is large enough to accommodate all relevant dependences on prior solar wind

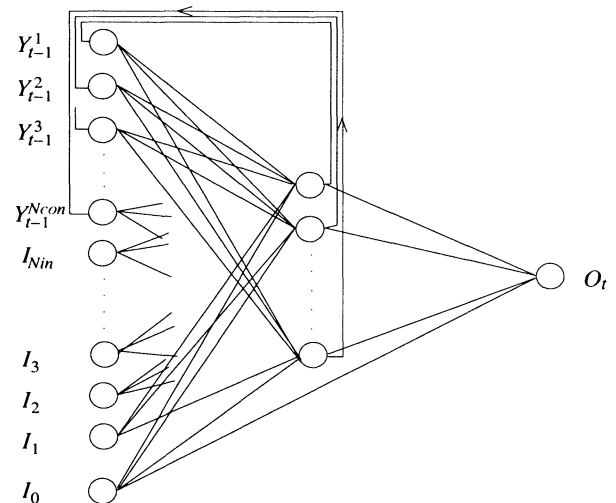


Figure 2. Elman recurrent network (ERN) with feedback connections from the hidden nodes to a set of context nodes at the input. The processing performed at the hidden nodes is nonlinear, while the output node performs a linear, weighted summation.

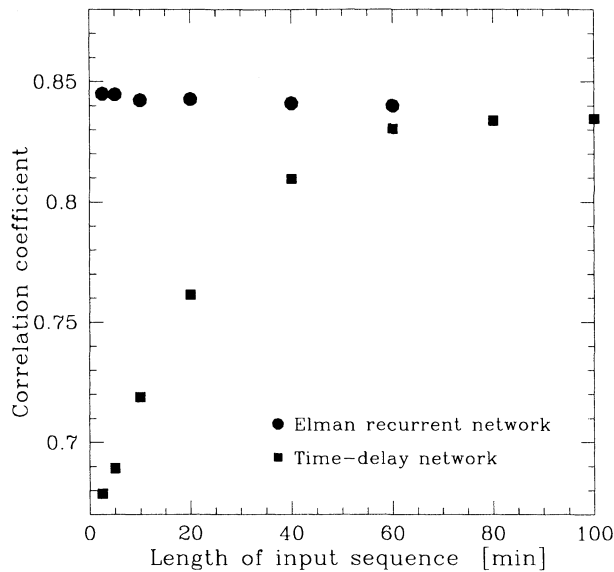


Figure 3. Correlation between observed and predicted test data as a function of T_I , the temporal length of the solar wind input sequence. The predictions are based on solar wind parameters n , V , and B_z . The performance of the Elman recurrent network (circles) is nearly independent of T_I , whereas the performance of the time delay network (squares) depends strongly on T_I .

inputs. The performance of a TDN is systematically improved with a larger T_I , up to roughly 100 min.

Contrary to the TDN, the dynamical properties of the ERN is a result of the mapping itself being dynamic. If the network input consists of only the instantaneous solar wind data, i.e., if no time-lagged solar wind data are presented to the ERN, any dynamic behavior must be the result solely of the network's feedback structure. If, in fact, important aspects of the solar wind-*AE* dynamics are encoded in the feedback structure, we can expect the performance of the network to be relatively independent of T_I . This property of the ERN differs markedly from the TDN and can be used as an indicator of the role played by the feedback.

To study this aspect of recurrent networks, a sequence of ERNs were trained with different temporal lengths of the solar wind input sequence, from 2.5 to 60 min. A corresponding sequence of TDNs were trained on identical input data. All networks used the solar wind parameters n , V , and B_z as input. The results are shown in Figure 3, where the correlation between predicted and observed values over the test set is plotted as a function of T_I .

As expected from previous studies, the performance of time delay network depend strongly on T_I . A TDN with an input-sequence length around 80 to 100 min predicts 70% of the observed *AE* variance. With $T_I = 2.5$ min, i.e., with only the instantaneous solar wind data presented to the network, the TDN only predicts 45% of the observed variance.

Unlike the time delay network, the performance of the recurrent network is nearly independent of the input-

sequence length T_I . Using only the instantaneous solar wind data, i.e., with $T_I = 2.5$ min, the ERN predicts 71% of the *AE* variance, which is marginally better than the best performing TDN.

We thus conclude that adding time-lagged solar wind data to the ERN input does not result in an improved prediction accuracy as measured by the correlation between observed and predicted values. The dynamical behavior of an ERN appears to be entirely due to the feedback structure of the network, and this feedback structure can effectively replace a sequence of time-lagged solar wind data.

4.3. Number of Recurrent State Variables

For an ERN that is fed with only the instantaneous solar wind parameters, i.e., that does not receive any time-lagged external inputs, the number of recurrent state variables becomes an important design parameter of the network. This number provides an upper limit to the dimensionality of the dynamics that can be described by the network. To clarify the role played by the number of state variables, we trained a sequence of ERNs with a varying number of context nodes, and thus a varying number of recurrent state variables. The external solar wind input consists of a single sample of solar wind n , V , and B_z . The results are summarized in Figure 4.

With four context nodes the ERNs predict 71% of the *AE* variance. The prediction accuracy does not improve by adding more context nodes, but with less than four context nodes the network performance suddenly

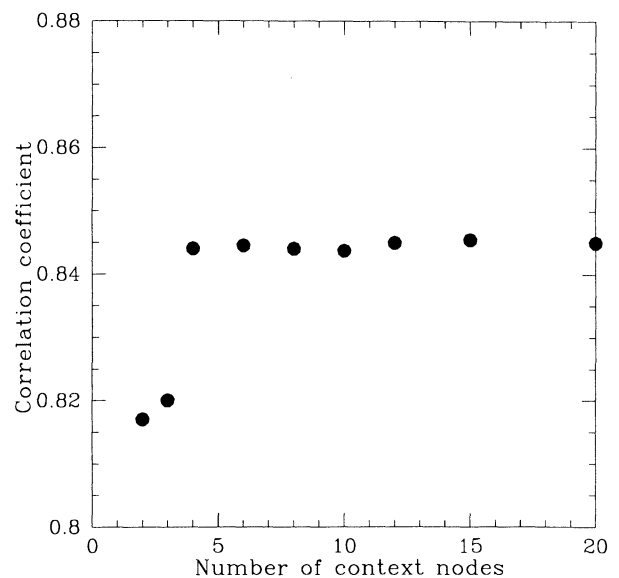


Figure 4. Correlation between observed and predicted test data as a function of the number of context nodes for an Elman recurrent network. The predictions are based on the solar wind parameters n , V , and B_z . Four recurrent state variables are optimal: the predictions are not improved by adding more than four context nodes, but with less than that the prediction accuracy drops.

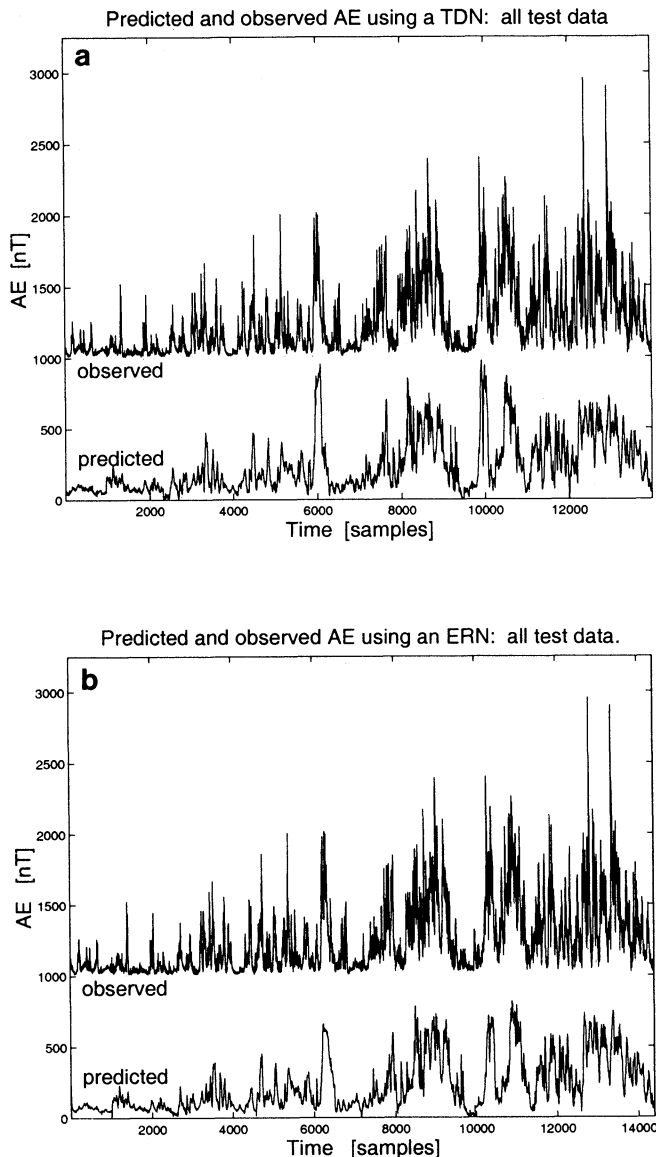


Figure 5. Observed and predicted AE for (a) the basic time delay network using 100 min of n , V , and B_z as input, and (b) the basic Elman recurrent network using a single sample of n , V , and B_z as input. All test data are shown; 11 separate intervals have been concatenated into a single sequence, and the observed values have been vertically shifted.

drops. We conclude that four recurrent state variables are sufficient to give the network a dynamic behaviour that accounts for 71% of the observed AE variance. A larger number of state variables does not contribute significantly to improved predictions.

The drop in network performance from four to three context nodes is rather abrupt (Figure 4). The performance for two or three context nodes is, however, relatively high compared to TDNs with short sequences of solar wind data. Around 66% of the observed variance is still predicted with only two context nodes, corresponding to a TDN using around 45 min of solar wind input data.

4.4. Qualitative Results

All test data are shown in Figure 5: the TDN results are presented in Figure 5a and the ERN results in Figure 5b. The 11 test intervals have been concatenated into a single sequence and observed values are vertically shifted. Observe that the number of test data are not exactly the same for TDNs and ERNs. This is due to the different temporal lengths T_I of the solar-wind input sequences. For the TDN, T_I corresponds to 40 samples, whereas for the ERN it corresponds to a single sample. The first 40 samples are lost from each data interval when using a TDN. When using an ERN, only a single sample is lost from each interval.

Figure 5 gives a rough overview of the correspondence between observation and prediction. The dominating features of the observed AE record are reproduced by the predictions. The amplitudes are, however, underestimated and most narrow, high-frequency features are broadened and some of them are even completely missed. An interesting observation is that even though the statistical performance of the ERN is slightly better than the performance of the TDN, this is not immediately obvious from a visual inspection of Figures 5a and 5b. In fact, it rather tends to give the opposite impression. Several of the dominating features are less underestimated and less broadened by the TDN. The ERN appears to produce smoother predictions than the TDN. This is also confirmed by the variability of the predicted AE : the ERN predictions have a standard deviation of 184 nT, compared to 194 nT for the TDN predictions. As a comparison, the standard deviation of the observed test data is 235 nT.

Despite these obvious differences, the TDN and the ERN predictions are relatively similar, even down to small-scale details. Compare, for example, the interval from $t = 2000$ up to the large peak near $t = 6000$. There is a one-to-one correspondence between the peaks of the two predictions. The same features of the AE records are predicted by both the TDN and the ERN. Where observed features fail to be predicted, as the two narrow peaks around $t = 13,000$, they are simultaneously missed by both the TDN and the ERN.

Figure 6 compares, in much more detail, the predicted and observed values for the TDN (left column) and the ERN (right column). Four of the 11 test data intervals are presented: Bargatze intervals 12, 15, 18, and 27. Most comments on the qualitative aspects of Figure 5 are confirmed by the higher resolution of Figure 6. An interesting case where the TDN and the ERN predictions differ is shown in test interval 9 (Bargatze interval 27). In the TDN plot (bottom left of Figure 6) the activity around $t = 1600$ shows three peaks reaching ~ 400 nT and a fourth peak, barely discernible just after the second peak, that reach 250 nT. The corresponding part of the ERN plot (bottom right of Figure 6) shows only two peaks. The first and the second peaks are not resolved and instead form a broad feature in the ERN plot. The third, smaller peak is not found at all in the

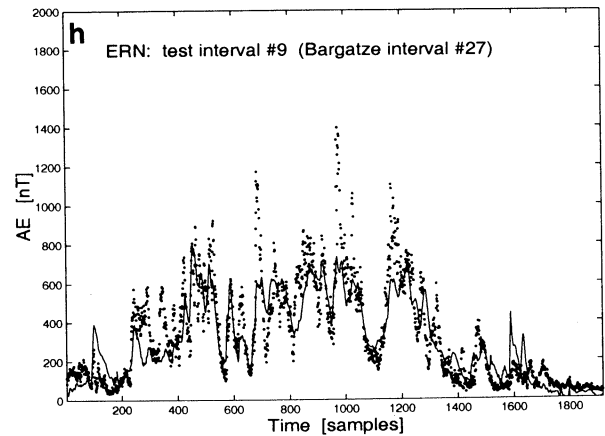
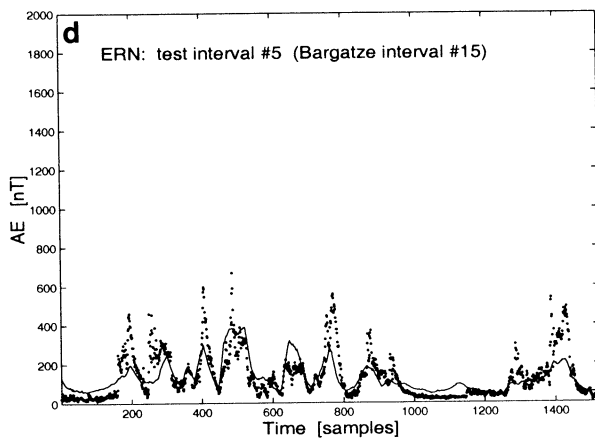
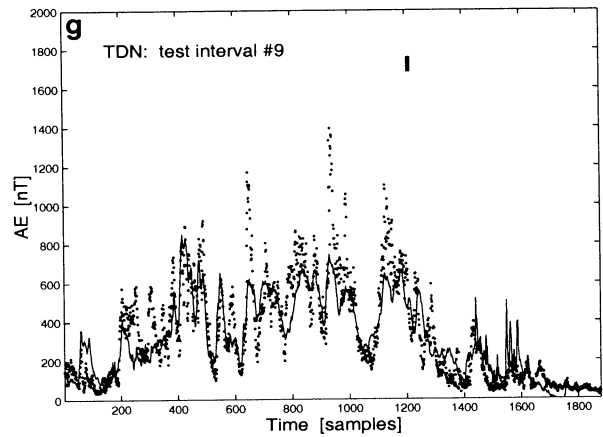
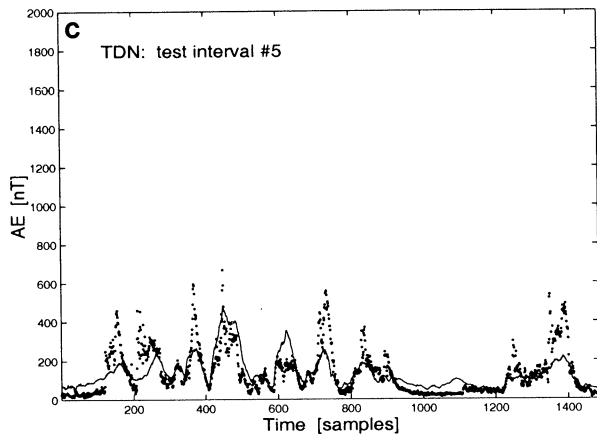
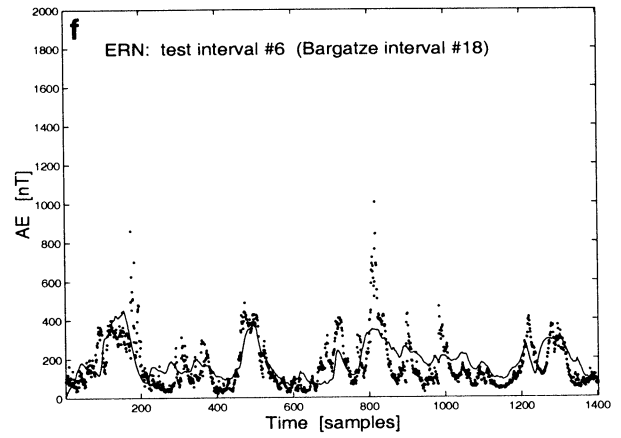
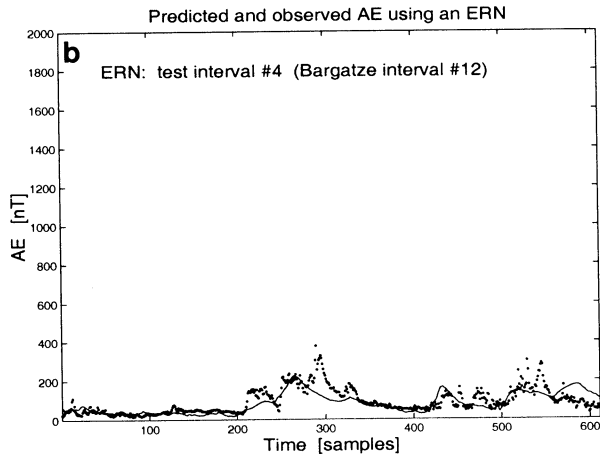
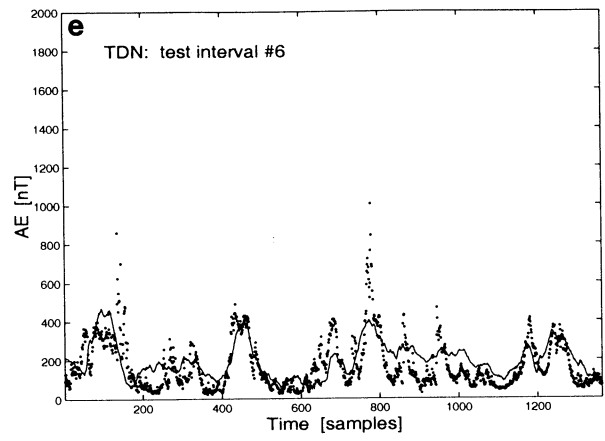
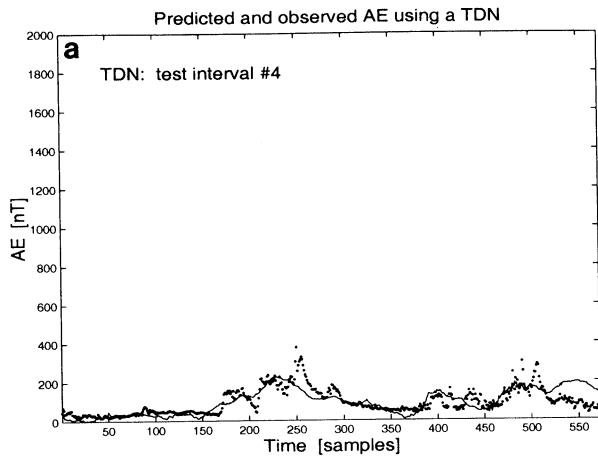


Figure 6. Predicted and observed *AE* for four of the 11 intervals in the test set. The dots show the observed values, and the solid lines show the predictions.

ERN plot. It probably contributes to a broadening of the base of the first two unresolved peaks. The last of the four peaks is shown in the ERN plot, but it is much broader than in the TDN plot. However, for the most part the TDN and the ERN produce predictions that are very similar. The broad activity around $t = 1200$ in the same two plots, is an example of nearly identical predictions, and many more nearly identical features can be found in the other plots of Figure 6.

In general, the ERN tends to smooth high-frequency *AE* variations more than the TDN. The tendency of the TDN to produce predictions with narrow features that not exactly correspond to observations, is probably a part of the explanation to why the ERN is somewhat better than the TDN from a statistical point of view, while a visual inspection of the two predictions gives the opposite impression.

5. Conclusions

The aim of this study is to evaluate the abilities of ERNs to approximate the dynamics of the solar wind-*AE* relation. This was done by a comparison with TDNs, whose predictive capabilities are better known. The results show that Elman recurrent networks can predict approximately the same fraction of the observed *AE* variance as the time delay networks. The predictions are, however, qualitatively somewhat different. The TDN tends to produce predictions with a more irregular appearance, while the ERN tends to smooth out high-frequency irregularities. Observed high-amplitude disturbances tend to be more broadened and underestimated by the ERN than by the TDN. Still, the statistical performance of the ERN is somewhat better than that of the TDN.

The preferred network configurations are vastly different for the two types of network. While a TDN may require up to 100 min of solar wind data, the ERN only requires a single sample of solar wind parameters. The number of weights, and thus the number of free parameters, is much smaller for an optimal ERN compared to a TDN. An ERN can be very simple and still produce relatively accurate predictions: a small number of input nodes, four context nodes, four hidden nodes, and a single output node. While providing an efficient representation of more than 70% of the observed *AE* variance, the simplicity of the model and the low number of internal state variables also indicates that the magnetospheric system governing the global-scale auroral electrojet dynamics can in fact be characterized by a small number of degrees of freedom.

Acknowledgments. Michel Blanc thanks Alex Klimas and Jouni Takalo for their assistance in evaluating this paper.

References

- Arnoldy, R.L., Signature in the interplanetary medium for substorms, *J. Geophys. Res.*, *76*, 5189, 1971.
- Bargatze, L.F., D.N. Baker, R.L. McPherron, and E.W. Hones, Magnetospheric impulse response for many levels of geomagnetic activity, *J. Geophys. Res.*, *90*, 6387, 1985.
- Casdagli, M., A dynamical systems approach to modeling input-output systems, in *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank, pp. 265-281, Addison-Wesley-Longman, Reading, Mass., 1992.
- Clauer, C.R., R.L. McPherron, C. Searls, and M.G. Kivelson, Solar-wind control of auroral zone geomagnetic activity, *Geophys. Res. Lett.*, *8*, 915, 1981.
- Cleeremans, A., D. Servan-Schreiber, and J.L. McClelland, Finite state automata and simple recurrent networks, *Neural Comput.*, *1*, 372, 1989.
- Elman, J.L., Finding structure in time, *Cognitive Sci.*, *14*, 179, 1990.
- Gleisner, H., and H. Lundstedt, Response of the auroral electrojets to the solar wind modeled with neural networks, *J. Geophys. Res.*, *102*, 14,269, 1997.
- Gleisner, H., H. Lundstedt, and P. Wintoft, Predicting geomagnetic storms from solar-wind data using time-delay neural networks, *Ann. Geophys.*, *14*, 679, 1996.
- Haykin, S., *Neural Networks - A Comprehensive Foundation*, 2nd ed., Prentice Hall, Old Tappan, N. J., 1999.
- Hernandez, J.V., T. Tajima, and W. Horton, Neural net forecasting for geomagnetic activity, *Geophys. Res. Lett.*, *98*, 7673, 1993.
- Hunter, N.F., Applications of nonlinear time series models to driven systems, in *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank, pp. 467-491, Addison-Wesley-Longman, Reading, Mass., 1992.
- Iyemori, T., H. Maeda, and T. Kamei, Impulse response of geomagnetic indices to interplanetary magnetic fields, *J. Geomagn. Geoelectr.*, *31*, 1, 1979.
- Jordan, M.I., Serial order: A parallel, distributed processing approach, in *Advances in Connectionist Theory: Speech*, edited by J.L. Elman and D.E. Rumelhart, Erlbaum, Hillsdale, N.J., 1989.
- Klimas, A.J., D. Vassiliadis, D.N. Baker, and D.A. Roberts, The organized nonlinear dynamics of the magnetosphere, *J. Geophys. Res.*, *101*, 13,089, 1996.
- Klimas, A.J., D. Vassiliadis, and D.N. Baker, *Dst* index prediction using data-derived analogues of the magnetospheric dynamics, *J. Geophys. Res.*, *103*, 20,435, 1998.
- Lapedes, A., and R. Farber, Nonlinear signal processing using neural networks: prediction and system modelling, *Tech. Rep. LA-UR-87-2662*, Los Alamos Nat. Lab., Los Alamos, N. M., 1987.
- Lundstedt, H., and P. Wintoft, Prediction of geomagnetic storms from solar-wind data with the use of a neural network, *Ann. Geophys.*, *12*, 19, 1994.
- Murayama, T., and K. Hakamada, Effects of solar wind parameters on the development of magnetospheric substorms, *Planet. Space Sci.*, *23*, 75, 1975.
- Price, C.P., D. Prichard, and J.E. Bischoff, Nonlinear input/output analysis of the auroral electrojet index, *J. Geophys. Res.*, *99*, 13,227, 1994.
- Rumelhart, D.E., G. Hinton, and R. Williams, Learning representations by back-propagating errors, *Nature*, *323*, 533, 1986.
- Takalo, J., and J. Timonen, Neural network prediction of *AE* data, *Geophys. Res. Lett.*, *24*, 2403, 1997.

- Valdivia, J.A., D. Vassiliadis, A.J. Klimas, and A.S. Sharma, Modelling the spatial structure of the high latitude perturbations and the related current systems, *Phys. Plasmas*, *6*, 4185, 1999.
- Vassiliadis, D., The input-state space approach to the prediction of auroral geomagnetic activity from solar wind variables, in *Proceedings of the International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics*, edited by J. Joselyn, H. Lundstedt, and J. Trolinger, pp. 145-151, NOAA-SEL, Boulder, 1993.
- Vassiliadis, D., A.J. Klimas, D.N. Baker, and D.A. Roberts, A description of the solar wind-magnetosphere coupling based on nonlinear filters, *J. Geophys. Res.*, *100*, 3495, 1995.
- Waibel, A., T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, Phoneme recognition using time-delay neural networks, *IEEE Trans. Acoustics Speech Signal Proc.*, *37*, 328, 1989.
- Weigel, R.S., W. Horton, T. Tajima, and T. Detman, Forecasting auroral electrojet activity from solar wind input with neural networks, *Geophys. Res. Lett.*, *26*, 1353, 1999.
- Wu, J.-G., and H. Lundstedt, Prediction of geomagnetic storms from solar wind data using Elman recurrent neural networks, *Geophys. Res. Lett.*, *23*, 319, 1996.
- Wu, J.-G., and H. Lundstedt, Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks, *J. Geophys. Res.*, *102*, 14,255, 1997.
-
- H. Gleisner, Lund Observatory, Box 43, SE-22100 Lund, Sweden (Hans.Gleisner@astro.lu.se).
- H. Lundstedt, Swedish Institute of Space Physics, Solar-Terrestrial Physics Division, Scheelev. 17, SE-22370 Lund, Sweden (henrik@irfl.lu.se).

(Received October 26, 2000; revised March 1, 2001; accepted March 3, 2001.)