GEOPHYSICAL RESEARCH LETTERS, VOL. 20, NO. 23, PAGES 2707-2710, DECEMBER 14, 1993

NEURAL NET FORECASTING FOR GEOMAGNETIC ACTIVITY

J.V. Hernandez, T. Tajima, and W. Horton
Institute for Fusion Studies, University of Texas at Austin

Abstract. We use neural nets to construct nonlinear models to forecast the AL index given solar wind and interplanetary magnetic field (IMF) data. We follow two approaches: 1) the state space reconstruction approach, which is a nonlinear generalization of autoregressive-moving average models (ARMA) and 2) the nonlinear filter approach, which reduces to a moving average model (MA) in the linear limit. The database used here is that of Bargatze et al. [1985].

Introduction

Linear filter studies [Iyemori et al., 1979; Clauer et al., 1981; Bargatze et al., 1985; Baker et al., 1986; McPherron et al., 1988; Vassiliadis et al., 1993] suggest that magnetospheric response to energy transfer from the solar wind contains a nonlinear component. On the other hand, dissipation in the magnetospheric system has the potential for a significant reduction in the number of relevant degrees of freedom. Though this is a controversial issue, since the magnetosphere is randomly driven by the solar wind [Prichard and Price, 1992], the success of nonlinear analogue models [Baker et al., 1990; Roberts et al., 1991; Klimas et al., 1992; Goertz et al., 1993] in reproducing several features of geomagnetic activity indicates that some global aspects of magnetospheric dynamics may be modeled by low-dimensional systems of equations. State space reconstruction studies [Vassiliadis et al., 1990; Roberts et al., 1991; Shan et al., 1991; Klimas et al.,1992; Sharma et al., 1993] using techniques valid only for autonomous systems indicate that the relevant dimension d is in the range $2 \lesssim d \lesssim 4.6$.

We view the magnetosphere as a driven system and construct models of geomagnetic activity using neural nets. The input to the system is $I(t) = V(t)B_s(t)$, where $V(t)$ is the solar wind velocity, and $B_s(t)$ is the southward component of the IMF, and the output is $O(t) = AL(t)$. The choice of $I(t) = V(t)B_s(t)$ is consistent with the results of Akasofu [1979]. The database used here is that of Bargatze et al. [1985]. This database consists of 34 uncorrelated data intervals of solar wind plasma, IMF and AL index measurements at a time resolution of $\tau = 2.5$ min.

Neural Networks and Nonlinear Models

Neural nets have been applied to autonomous systems by Lapedes and Farber [1987] for forecasting purposes. An introduction to neural computation is given by Hertz et al. [1991].

Paper number 93GL02848
0094-8534/93/93GL-02848$03.00

We use feed forward networks, which are universal approximants to mappings of the form $f : R^q \to R^s$. Feed forward nets consist of one input layer with q inputs, one or several hidden layers, with one or several units per layer, and the output layer with s outputs. Let $V_i^{(k)}$ represent the output of unit i in the $k-th$ layer. Then $V_i^{(k)}$ is a function $g_i^{(k)}$ (the activation function) of the weighted sum of the outputs of all the units in the previous layer,

$$V_i^{(k)} = g_i^{(k)} \left(\sum_{j=1}^{N} W_{ij}^{(k-1)} V_j^{(k-1)} \right) , \tag{1}$$

where $W_{ij}^{(k-1)}$ is the weight connecting unit j to unit i and the sum runs over all the N units in the $k-1$ layer. The network is 'trained' to approximate f by choosing the set of all the weights $\mathbf{W}$ according to a 'learning rule'. The activation function $g(x)$ is a nonlinear, saturating function which we take to be $\tanh(x)$. In the limit where the activation function is linear, $g(x) = x$, the system reduces to a linear filter.

Let O denote the actual output of f, and let $\hat{O}$ denote the neural net output corresponding to the input $\mathbf{I} \epsilon R^q$. For a given set of M inputs, we define the mean square error by

$$\mathcal{E}(\mathbf{W}) \equiv \frac{1}{M} \sum_{j=1}^{M} (O_j - \hat{O}_j)^2 . \tag{2}$$

A widely used learning rule is the back propagation algorithm [Rumelhart et al., 1986], which consists of minimizing $\mathcal{E}(\mathbf{W})$, for a given (measured) set of input-output pairs (the training set), by adjusting the weights according to a gradient descent recipe,

$$W_{ij}^{\text{new}} = W_{ij}^{\text{old}} + \Delta W_{ij}; \quad \Delta W_{ij} = -\eta \frac{\partial \mathcal{E}}{\partial W_{ij}} + \alpha \Delta W_{ij} , \tag{3}$$

with η, the learning rate, a small positive constant, $0 < \eta \ll 1$ and with α, the inertial term, a number between 0 and 1.

The goodness of the models is characterized by the average relative variance [Casdagli, 1989; Weigend et al., 1992],

$$ARV = \frac{\sum_{j=1}^{M} (O_j - \hat{O}_j)^2}{\sum_{j=1}^{M} (O_j - \langle O \rangle)^2} , \tag{4}$$

where $\langle O \rangle$ is the mean value of the set of M actual outputs, and by the correlation coefficient ρ,

$$\rho = \frac{\frac{1}{M} \sum_{j=1}^{M} (O_j - \langle O \rangle) \left(\hat{O}_j - \langle \hat{O} \rangle \right)}{\sigma_O \sigma_{\hat{O}}} , \tag{5}$$

where σ_O and $\sigma_{\hat{O}}$ are the standard deviations of the actual and the model outputs, respectively.

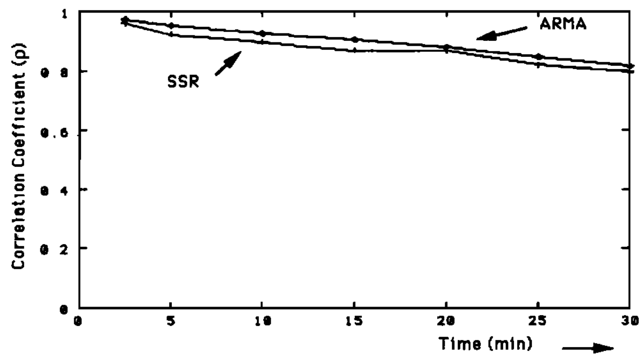

Fig. 1. Correlation coefficient as a function τ, for single-step predictions using the SSR net (crosses) and the ARMA net (dots).

Forecasting Geomagnetic Activity

We attempt single-step predictions of magnetospheric response to the variable solar wind conditions. The predictions are computed with the known records of Bargatze et al. [1985]. We describe two kinds of models to perform single-step predictions.

State Space Reconstruction Approach

In autonomous systems dissipation may lead their evolution to become trapped onto a low-dimensional attractor. Let $O(t)$ be some variable characterizing the system. Then [Packard et al., 1980] knowledge of m previous measurements of $O(t)$, $O_{t-\tau}, O_{t-2\tau}, \ldots, O_{t-m\tau}$, where τ is the time interval between successive measurements, should provide enough information to predict the evolution of $O(t)$, as long as $m > d$, where d is the dimension of the attractor. Takens [1981] has proven that in order to predict the evolution of the system the dimension m of the delay vector is at most $m = 2d + 1$. Sauer et al. [1992] have extended Takens proof to the case of fractional d.

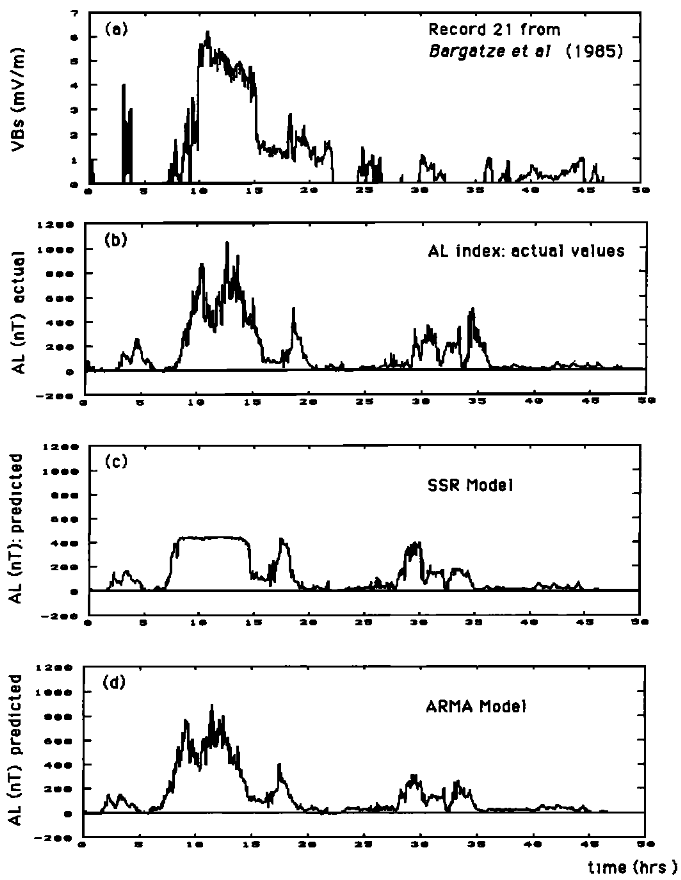

Fig. 2. Single-step predictions of the SSR and ARMA nets: (a) solar wind input, (b) actual AL index, (c) SSR net with $\tau = 15$ min., (d) ARMA net with $\tau = 15$ min.

Recently [Casdagli, 1992; Hunter, 1992] interest has shifted to driven dissipative systems. In such systems the relevant information is contained in the delay vector which includes input and output measurements,

$$I_{t-\tau}, I_{t-2\tau}, \ldots, I_{t-l\tau}, O_{t-\tau}, O_{t-2\tau}, \ldots, O_{t-m\tau} . \quad (6)$$

This approach has also been taken by Price and Prichard [1993] to forecast the AE index using a local linear type of nonlinear input-output method [Casdagli, 1992].

We construct (6) using $I(t) = V(t)B_s(t)$ and $O(t) = AL(t)$. As for the time lags, we use $m = l = 7$. We use τ in the range from 2.5 to 30 min. For each τ we train, keeping the architecture fixed, a corresponding neural net. In previous studies [Vassiliadis et al., 1990; Roberts et al., 1991; Shan et al., 1991; Sharma et al., 1993], time steps between 15 and 175 min. have been used.

In order to get robust predictions we need to monitor the generalization ability of the net, and thus we use three different data sets: training, validation and prediction sets. The validation set is used to determine the number of training epochs and the number of hidden units for which the neural net yields best generalizations, and the prediction set is used to check the out-of-sample performance of the net. We use the intervals 1 and 31 with 1704 points as training set, the interval 15 with 1524 points as validation set, and the interval 21 with 1176 points as prediction set.

The network consists of the input layer with $n = m + l = 14$ inputs, two hidden layers, the first with 7 and the second with 2 units, and one output for the predicted AL. The activation functions for the hidden layers are $g(x) = \tanh(x)$, and the activation function for the output is linear. If all the activation functions are linear, then the SSR net model reduces to a linear ARMA(m, l) model of the form

$$\hat{O}(t) = \sum_{j=1}^{l} \alpha_j I_{t-j\tau} + \sum_{k=1}^{m} \beta_k O_{t-k\tau} . \quad (7)$$

Because the solar wind input and the AL index vary over several orders of magnitude, we use, for training purposes, the natural log of their values shifted by suitable constants to avoid logarithmic singularities. The learning rate is $\eta = 0.0005$, and the momentum term is $\alpha = 0.9$.

Figure 1 displays, for both the SSR net and its corresponding ARMA model, the single-step correlation coefficient (5) for different time steps τ for the prediction set. Similar results hold for the prediction error (4). For the intervals considered, the results obtained from the ARMA model are better than those obtained from the SSR net. In previous studies [Bargatze et al., 1985; McPherron et al., 1988] typical correlations are of the order of 50 – 80%. Higher correlation ($\sim$ 90%) [McPherron et al., 1985; Goertz et al., 1993] have been obtained by pre-setting the parameters of the models for the event under consideration.

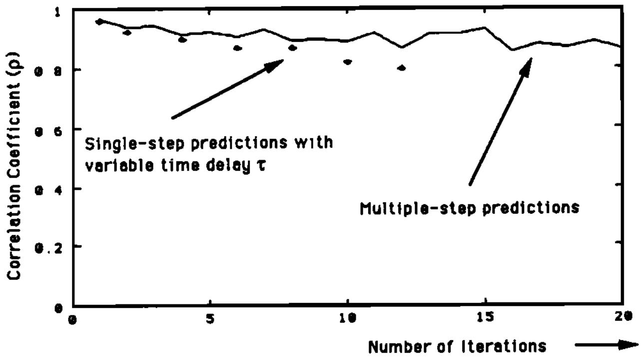

Fig. 3. Correlation coefficient for multiple-step (solid lines) predictions of the SSR net as a function of the number of iterations. The results of single-step predictions (dots) with variable τ are shown just for reference.

The single-step prediction performance of the SSR net is shown in Figure 2 for the prediction set. The upper panel displays the solar wind input, the second panel is the actual AL index, the third and fourth panels correspond to single-step predictions with $\tau = 15$ min. for both the SSR net and the corresponding ARMA model, respectively. For the third (fourth) panel the number of training epochs is 360 (880), $ARV = 0.25$ (0.28), and $\rho = 0.87$ (0.91). The SSR net clips large variations in the AL index. On the other hand, the ARMA model follows the general trend of the AL index.

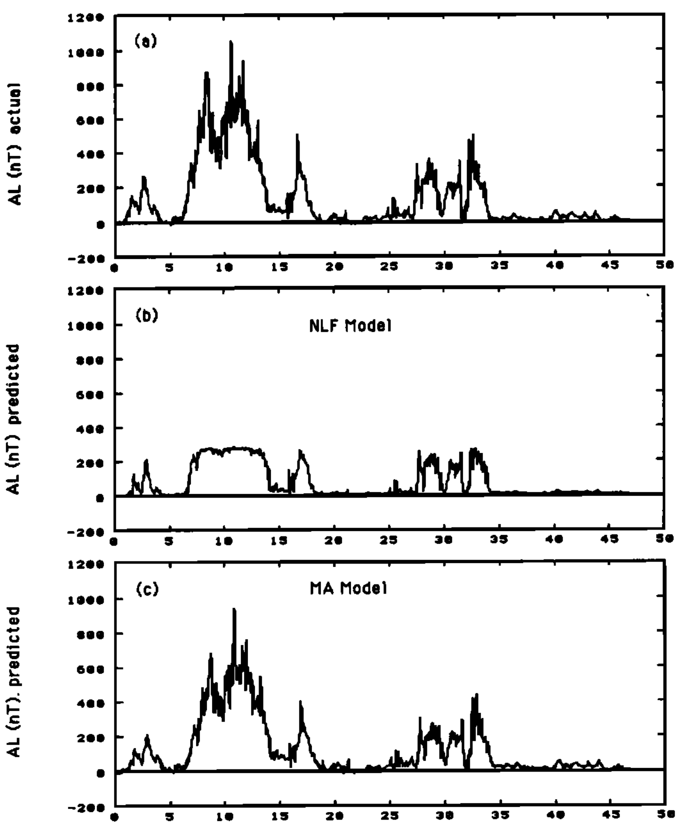

Fig. 4. Single-step prediction performance of the NLF and MA nets: (a) actual AL index, (b) NLF net results, (c) MA net results.

Predictions further into the future can be achieved through multiple-step predictions, that is by feeding back the predicted values to the input for several iterations. Figure 3 is a plot of the correlation coefficient (5) for multiple-step predictions (solid line) as a function of the number of iterations using the SSR net with $\tau = 2.5$ min. The multiple step prediction is shown to be better than a single prediction with an increasing τ.

Nonlinear Prediction Filter

In the nonlinear filter approach the input is given by a set of q past values of solar wind data, $I(t) = V(t)B_s(t)$, and the output is $O(t) = AL(t)$. For linear activation functions the NLF net reduces to a MA(q) model of the form

$$\hat{O}(t) = \sum_{j=1}^{q} \alpha_j I_{t-j\tau} \,. \tag{8}$$

For a time step $\tau = 15$ min., the NLF net consists of nine inputs, spaning 2.25 hrs., two hidden layers with three units in the first and one unit in the second, and one output for the predicted AL(t). Figure 4 shows (a) the actual AL index for data interval 21, (b) the single-step NLF net predictions, and (c) the single-step MA model predictions. For the NLF (MA) model the number of training epochs is 27 (440), $ARV = 0.46$ (0.2), $\rho = 0.87$ (0.9), $\eta = 0.001$ (0.0005), and $\alpha = 0.9$ (0.9). Large amplitude variations in the AL index are clipped by the NLF. In general, the results obtained by the SSR model are better than those obtained by the NLF.

Conclusions

The purpose of this work is to explore the problems associated with forecasting geomagnetic activity using neural networks. We use two modeling approaches: the state space reconstruction approach, which contains information on both the driven and the loading-unloading components, and the nonlinear filter approach, which models the driven component of magnetospheric activity. The forecasting results are tested in data sets different from those used for training. Our main findings are:

- The single-step prediction performance of the SSR net is better than the performance of the NLF net. The reason for this may be that the SSR net contains information on both the driven and the loading-unloading responses. For the data intervals considered the performance of the linear models is slightly better than the performance of their nonlinear counterparts. On the other hand, the multiple-step prediction performance of the SSR net is better than that achieved through single-step predictions with an increasing time step τ.

- Given the success of the linear stochastic models, it may seem that the magnetospheric response is not deterministic. However, it may well be that the neural net models are not fully capturing the underlying deterministic dynamics and further exploration is necessary.

Acknowledgments. We would like to thank S. Sharma, D. Vassiliadis, K. Wang, J.M. Rodriguez, and R. McPherron for their help, comments and interest in the present work. This work was supported by the NSF Grant ATM-8506646 and NASA Grant NAGW-2590.

References

Akasofu, S.-I., Interplanetary energy flux associated with magnetospheric substorms, *Planet. Space Sci., 27*425, 1979.

Baker, D.N., L.F. Bargatze and R.D. Zwickl, Magnetospheric response to the IMF: substorms, *J. Geomag. Geoelec., 38*, 1047, 1986.

Bargatze, L.F., D.N. Baker, R.L. Mc Pherron and E.W. Hones, Magnetospheric impulse response for many levels of geomagnetic activity, *J. Geophys. Res., 90*, 6387, 1985.

Casdagli, M., Nonlinear prediction of chaotic time series, *Physica D, 35*, 335, 1989.

Casdagli, M., A dynamical systems approach to modeling input-output systems, In *Nonlinear Modeling and Forecasting,* M. Casdagli and S. Eubanks, eds., Addison-Wesley, 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.

Goertz, C.K., L.H. Shan and R.A. Smith, Prediciton of Geomagnetic Activity, *J. Geophys. Res.,*submitted, 1991.

Hertz, J., A. Krogh and R.G. Palmer, *Introduction to the Theory of Neural Computation,* ch. 6, Addison-Wesley, 1991.

Hunter, N.F., Jr., Applications of nonlinear time-series models to driven systems, In *Nonlinear Modeling and Forecasting,* M. Casdagli and S. Eubanks, eds., Addision-Wesley, 1992.

Iyemori, T., H. Maeda and T. Kamei, Impulse response of geomagnetic indices to interplanetary magnetic field, *J. Geomagn. Geoelectr., 31*, 1, 1979.

Klimas, A.J., D.N. Baker, D.A. Roberts, D.J. Fairfield and J. Büchner, A nonlinear dynamical analogue model of geomagnetic activity, *J. Geophys. Res., 97,* 12253, 1992.

Lapedes, A.S. and R.M. Farber, Nonlinear signal processing using neural networks: prediction and system modeling. Technical Report LAUR-87-2662, Los Alamos National Laboratory, 1987.

McPherron, R.L., D.N. Baker, L.F. Bargatze, C.R. Clauer, and R.E. Holzer, IMF control of geomagnetic activity, *Adv. Space Res., 8*, 71, 1988.

Packard, N.H., J.P. Crutchfield, J.D. Farmer, and R.S. Shaw, Geometry from a time series, *Phys. Rev. Lett., 45,* 712, 1980.

Price, C.P. and D. Prichard, The non-linear response of the magnetosphere: October 30, 1978, *Geophys. Res. Lett., 20,* 771, 1993.

Prichard, D. and C.P. Price, Spurious dimension estimates from time series of geomagnetic indices, *Geophys. Res. Lett., 19,* 1623, 1992.

Roberts, D.A., D.N. Baker, A.J. Klimas, and L.F. Bargatze, Indications of low dimensionality in magnetospheric activity, *Geophys. Res. lett., 18,* 151, 1991.

Rumelhart, D.E. G.E. Hinton and R.J. Williams, Learning representations by back-propagating errors, *Nature, 323,* 533, 1986.

Sauer, T., J.A. Yorke, and M. Casdagli, Embedology, *J. Stat. Phys., 65,* 579, 1991.

Shan, L.-H., C.K. Goertz, and R.A. Smith, On the embedding-dimension analysis of AE and AL time series, *Geophys. Res. lett., 18,* 1641, 1991.

Sharma, A.S., D. Vassiliadis and K. Papadopoulos, Reconstruction of low dimensional magnetospheric dynamics by singular spectrum analysis, *Geophys. Res. Lett., 20,* 335, 1993.

Takens, F., Detecting strange attractors in turbulence, In *Dynamical Systems and Turbulence,* Springer, Berlin, 1981.

Vassiliadis, D., A.S. Sharma, T.E. Eastman, and K. Papadopoulos, Low-dimensional chaos in magnetospheric activity from AE time series, *Geophys. Res. Lett., 17,* 1841, 1990.

Vassiliadis, D., A.S. Sharma and K. Papadopoulos, An empirical model relating the auroral geomagnetic activity to the interplanetary magnetic field, *Geophys. Res. lett., 20,* 1731, 1993.

Weigend, A.S., B.A. Hubermand, and D.E. Rumelhart, Predicting sunspots and exchange rates with connectionist networks, in *Nonlinear Modeling and Forecasting,* M. Casdagli and S. Eubanks, Eds. (Addison-Wesley, 1992).

J.V. Hernandez, T. Tajima, and W. Horton, Department of Physics and Institute for Fusion Studies, University of Texas, Austin, TX 78712.

(Received April 30, 1993;
revised September 3, 1993
accepted September 28, 1993)