5-1990

Forecasting Sunspot Numbers Using Neural Networks

Ming Li

Kishan Mehrotra
Syracuse University, mehrotra@syr.edu

Chilukuri K. Mohan
Syracuse University, ckmohan@syr.edu

Sanjay Ranka
Syracuse University

SU-CIS-90-05

Forecasting Sunspot Numbers Using Neural Networks

Ming Li, Kishan Mehrotra, Chilukuri K. Mohan, and Sanjay Ranka

May 1990

School of Computer and Information Science
Suite 4-116
Center for Science and Technology
Syracuse, New York 13244-4100

(315) 443-2368

Abstract

A recurrent connectionist network has been designed to model sunspot data. Preliminary experimental work shows that the network can produce competitive results as compared to traditional autoregressive models. The method is not problem specific and could be applied to other problems in dynamical system modeling, recognition, prediction, and control fields. It is observed that statistical methods can be used to design an appropriate neural network architecture.

Key Words and Phrases: Sunspots, Neural Network, Autoregressive Models.

1 Introduction

The mysterious cycles in sunspot numbers have intrigued many investigators, including geologists, astronomers, climatologists, economists, historians, dendrochronologists, and statisticians. Sunspots are high-intensity electromagnetic flares of solar radiation of largely unknown and unpredictable causes. They have major effects on various terrestrial phenomena; an example is the long-range weather prediction. Sun-weather relationships in the stratosphere are well understood and sunspot numbers have to be taken into account in telecommunications and interplanetary flight. Moreover, sunspot variations are now proving important in quaternary chronology and paleoclimatology.

Modeling sunspot data is a great challenge to statisticians. In the past half-century many statisticians have tried to provide a functional relationship between the sunspot numbers. As addressed by Izenman (1985),

> "...the sunspot numbers have been shown to contain idiosyncrasies that suggest, quite strongly, that the underlying statistical mechanism by which they are generated is non-linear, non-stationary, and non-Gaussian, and as such they are used primarily as a yardstick to compare and judge new statistical modeling and forecasting methods."

Even though sunspot number prediction via statistical methods has provided promising results, new and traditional techniques are being explored with the hope of further improving the results. Neural networks provide a promising tool for this task and are tested in this paper.

Artificial neural networks are primitive models of the human brain. A neural network can be briefly described as a directed graph with weighted edges. Each node in the network is capable of simple nonlinear processing that typically involves calculating a weighted sum of its inputs and then passing it through a nonlinear function. A series of experiments were performed in which a neural network was trained to follow the curves depicting variations in sunspot numbers (see Figure 1). This trained network was used to predict the sunspot data for "future" periods on which the network had not been trained. These results were compared with four autoregressive models obtained via statistical analysis and our results compete well with traditional statistical methods. The prediction[1] of future observations via neural networks is very good. However, our main result is evidence that traditional statistical methods can be used in obtaining an architecture of the neural network that can provide a better prediction than the statistical model alone because the network does not depend on classical assumptions and is hence robust.

[1]In this paper we use "prediction" to describe the performance of a neural network on the data on which it was trained, and "forecasting" to denote performance on new test data not used in the training phase.

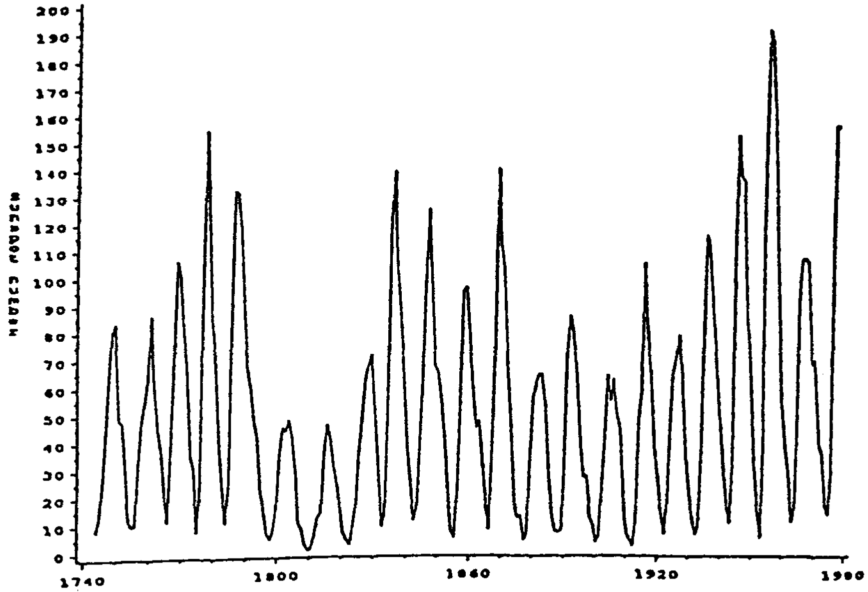

Figure 1. Sunspot cycles 0-21

The remainder of the paper is organized as follows. In Section 2 we describe the network architectures that were designed and used in this study. In Section 3 the sunspot data and statistical models are described. The results are presented in section 4. Finally, we summarize our main findings in Section 5.

2 Network Architecture

In recent years many network models have been proposed for temporal sequence processing tasks. Traditional feedforward neural networks can not learn to trace temporal sequences since a given input will always produce the same output to track temporally changing patterns. Some kind of *recurrent* connections are needed. By adding recurrent connections, feedforward networks can be made to learn complex behavior (see discussions in Rumelhart(1986), Gallant(1988), Elman(1988), Jordan(1986), Williams(1988)). In the present work, we choose a rather general scheme: a three-layer feedforward network with a fully-connected hidden layer. (See figure 2 for a three-layer three units in the hidden layer fully connected network.)

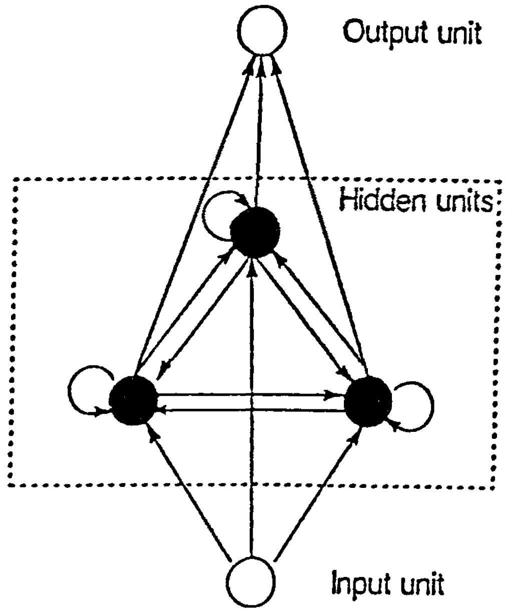

Figure 2. A feedforward neural network with
a fully-connected 3-unit hidden layer

The proposed network has a single unit input layer, a single unit output layer, and a fully-connected hidden layer. All units map real-value inputs to real-value outputs. For network training and experimentation the sunspot data are normalized into real numbers within the range [0, 1].

We experimented with a variable number of hidden layer units ranging between two to eighteen units. The best performance, measured in terms of mean square error (MSE), is offered by a network with eleven hidden larger units. This is an interesting observation because the sunspot data have also shown peaks that occur cyclically with a frequency of almost eleven years. All experimental results reported in Section 4 are based on the network of 11-unit in the hidden layer.

3 Network Training

The network training phase can be described as follows. In each training step an input value, i.e., a normalized sunspot number of some year, is presented to the network. The network is asked to predict the next value in the sunspot number sequence. The error between the predicted value and the value actually observed is then measured and propagated backwards along both feedforward connections and recurrent connections. The weights between any two nodes are modified in the same way as that in the back propagation algorithm. The interconnections between the nodes in the hidden layer do not cause any major difficulty due to the *unfolding* technique [11] which treats a recurrent network as an unfolded feedforward network. In the back-propagation phase, at each time step the hidden units will receive error values, not only from the output node, but also from other hidden units which are directly connected to them.

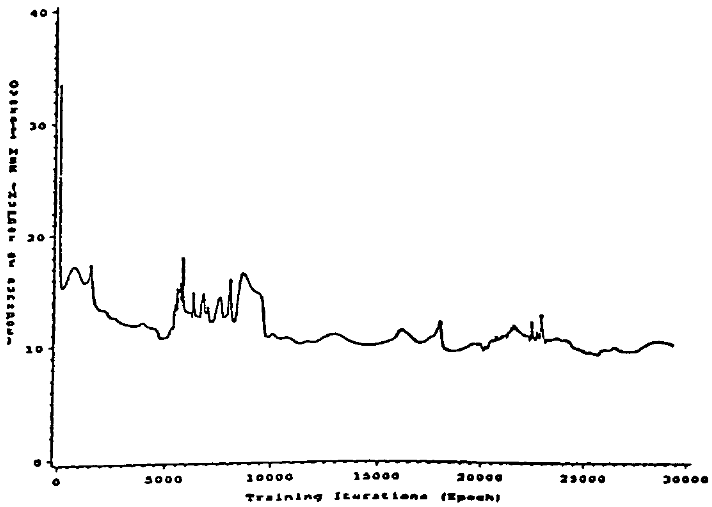

Figure 3. Data set 1: Overall error during training

In training the network, a small learning rate gave better performance, thus indicating that the error surface of the problem has many local minima. In all training cases we chose a learning rate of 0.3, and an associated momentum term of 0.6. The overall mean square error was calculated repeatedly to monitor the training performance. Behavior of MSE plotted against time is shown in Figure 3 for one of the four experiments described below. The number of iterations required to train the network varied for each experiment and ranged from 5000 to 30000.

4 Experiments

We have used sunspot data taken from Anderson (1971) and Schove (1983). In our experiments we have processed only the annual average sunspot numbers, although half yearly data are also available and have been studied in statistical literature. Two sources of data quoted above offer slightly different data for some years. In such cases we have used the data in Anderson (1971).

Our results are based upon the comparisons between the predictive and forecasting abilities of our network model and four well-known *autoregressive* (AR) statistical models. The autoregressive models are briefly described below.

[NOTATION: "x_{t-i}" denotes the sunspot numbers observed in the i-th year preceding the year t currently being considered.]

Model 1: $x_t = 1.239x_{t-1} - 0.55x_{t-2} + 0.128x_{t-9}$

Schaerf (1964) proposed to use this model based on annual sunspot data for the period 1749-1924. In this model x_t is expressed as a function of x_{t-1}, x_{t-2}, and x_{t-9}. This implies that, according to this model, the sunspot number in the year t can be explained in terms of the sunspot numbers observed at years $t-1$, $t-2$, and $t-9$.

Model 2: $x_t = 14.9 + 1.32x_{t-1} - 0.63x_{t-2}$

Box and Jenkins (1970) proposed this model based on annual data for the period 1770-1869. Note that in this model, as in the previous model, x_t is expressed in terms of x_{t-1} and x_{t-2}.

Model 3: $x_t = 5.055 + 1.250x_{t-1} - 0.538x_{t-2} + 0.189x_{t-9}$

Morris (1977) proposed this model based on annual data for the period 1755-1964.

Model 4: $x_t = 9.8 + 1.19x_{t-1} - 0.53x_{t-2} + 0.24x_{t-9} - 0.10x_{t-18}$

The model proposed by Per Hokstad (1983) is based on annual data for the period 1705-1968 and suggests a period of 18 years.

It should be noted here that there exist many other models more complex than the autoregressive models. For instance, the *outburst* models discussed in Morris (1977) is one such class. We do not consider them here because the main goal of this paper is not to find a better model but to show that a neural network can be used toward the purpose of forecasting sunspot numbers.

For a fair comparison, in each set of four cases the neural network was trained on the same data set that was used in developing the AR model. Suppose that the training set consists of data from year t_f to t_l; performance of the neural network was compared with the corresponding AR model in its ability to:

1. Predict sunspot numbers over the training period, i.e., for years t_f to t_l.
2. Forecast sunspot numbers for each year t given the data for $t-1$, $t-2$, etc., for $t = t_l + i, i = 1, \ldots, 12$.
3. Forecast sunspot numbers for years $t_{l+1}, \ldots, t_{l+12}$ based on data available up to t_l only.
4. Forecast sunspot number for years $t_{l+1}, \ldots, t_{l+55}$ based on data available up to t_l only.

To further emphasize the third and fourth comparison above, the AR and neural network models forecast the future data for years $t_{l+i} : i \geq 0$ based on data from t_f to t_l only. These forecasts are then compared with the actually observed data that is available up to 1980. Since models 3 and 4 use data up to 1964 and 1968, respectively, it is not possible to forecast for the next 55 years and compare it with actually observed data. Hence, in the cases of models 3 and 4 and fourth comparison we have trained the networks based on the data up to year 1925.

5 Experimental Results

In this section we present the comparative results both by plotting graphs of predicted vs. actual data and also by calculating the mean square error (MSE). Both methods of comparison are meaningful because, for example, the MSE may show a large value due to one or two large deviations only but that could be noted in the graphical comparison. As stated above, an AR model is compared with the corresponding neural network in (i) learning, i.e., how well the network predicts the data used in its training, (ii) in forecasting sunspot numbers with a lag time of one year, and (iii) in forecasting the sunspot numbers for the next few years.

5.1 Test 1: Comparison on the Training Data Set

The neural network predictions are tested against the corresponding AR models. Recall that in each case the network is trained on the same data set that was used in the AR modeling. The MSE results are shown in Table 1. It is observed that the neural network has better performance than the corresponding AR model (offers lower mean square error) for all four models.

Case		1	2	3	4
MSE	Models	680.51	229.0	235.51	253.16
	Network	79.73	87.76	93.16	90.60
$\left(\frac{\sqrt{MSE}}{mean}\right)$	Models	0.59	0.33	0.31	0.32
	Network	0.20	0.20	0.20	0.19

Table 1. Testing against the training data set

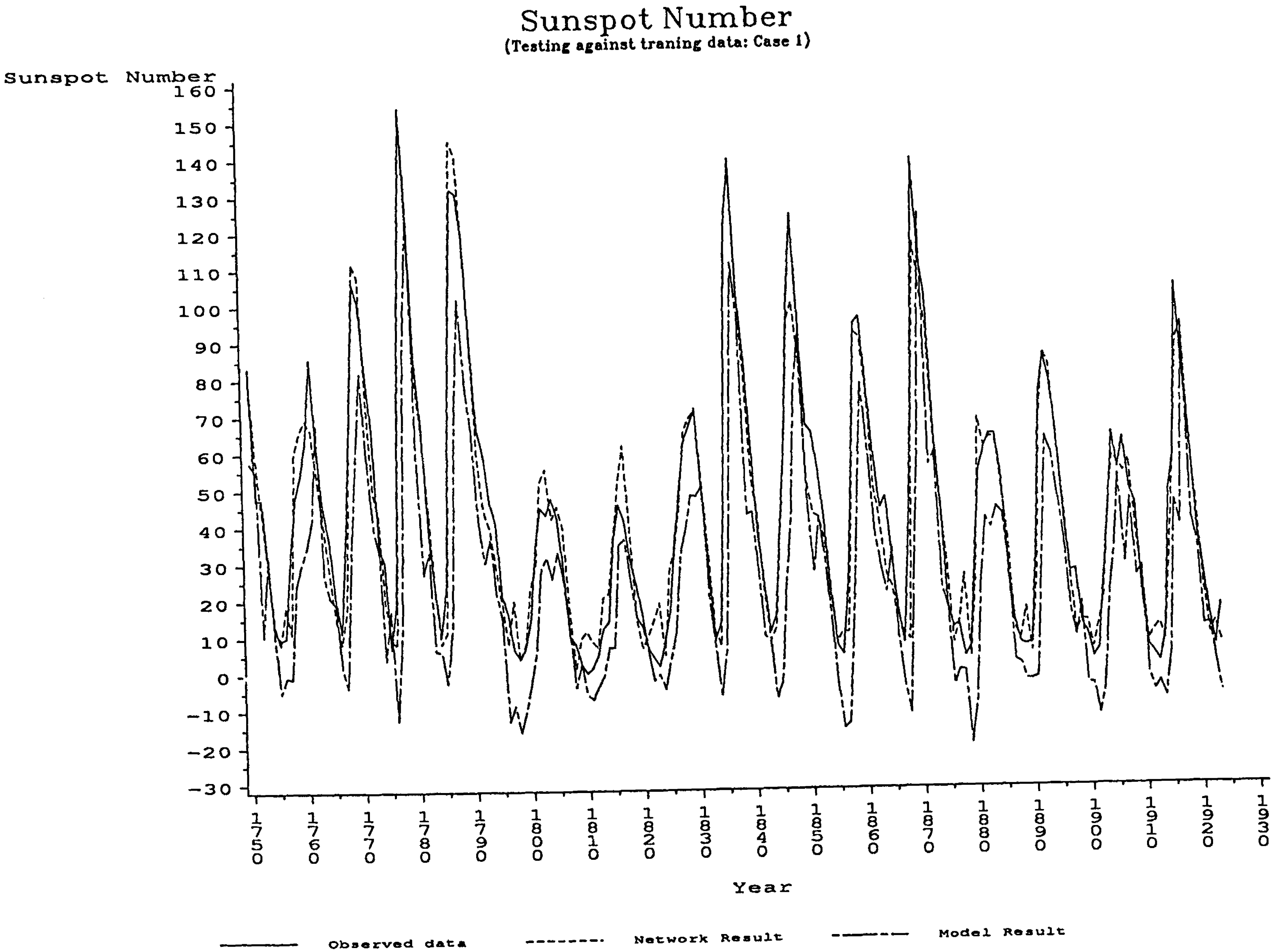

Figure 4. Results on training data in time period 1749-1924, compared to the model of Schaerf (1964)

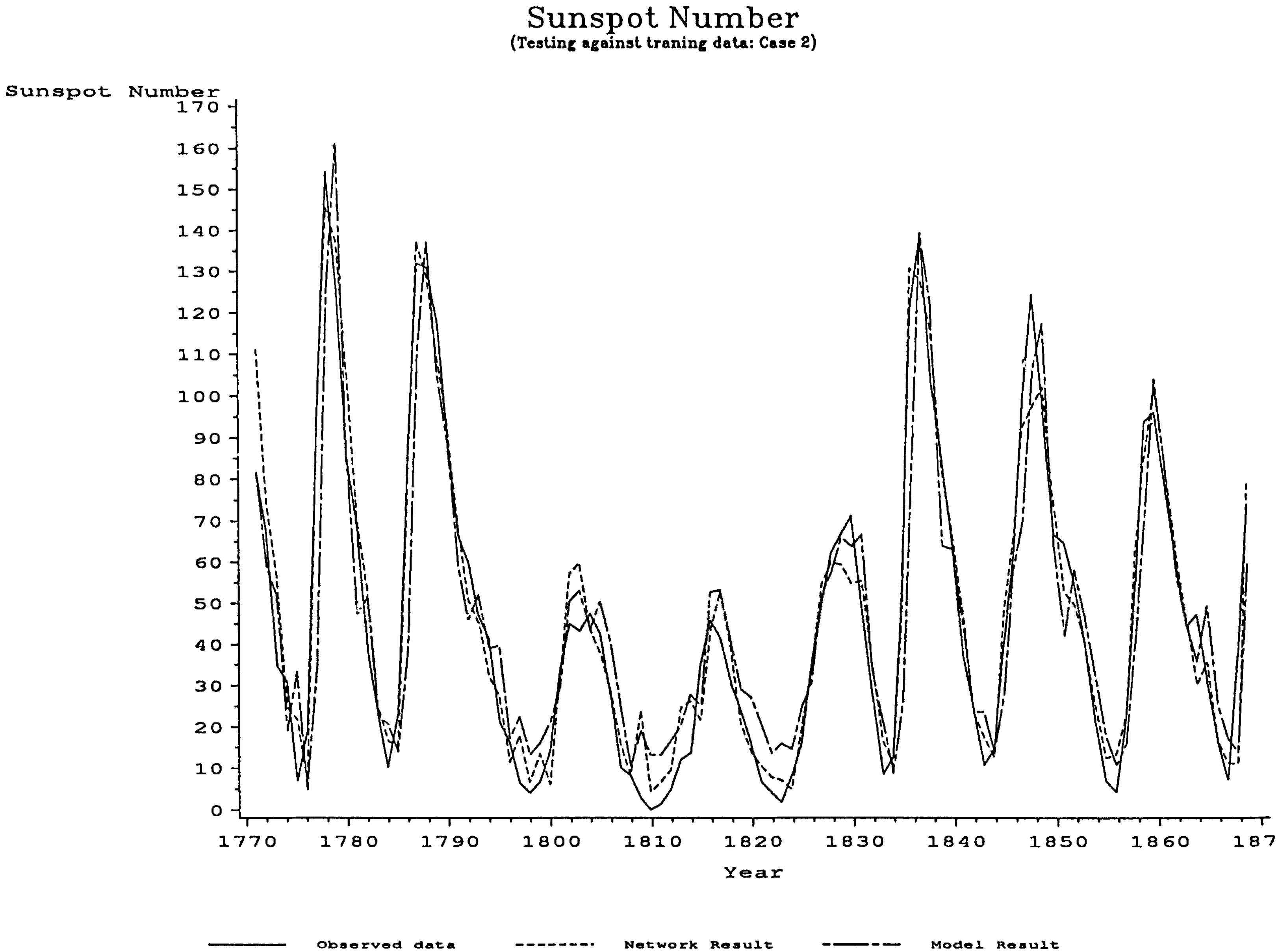

Figure 5. Results on training data in time period 1770-1869, compared to the model of Box and Jenkins(1970)

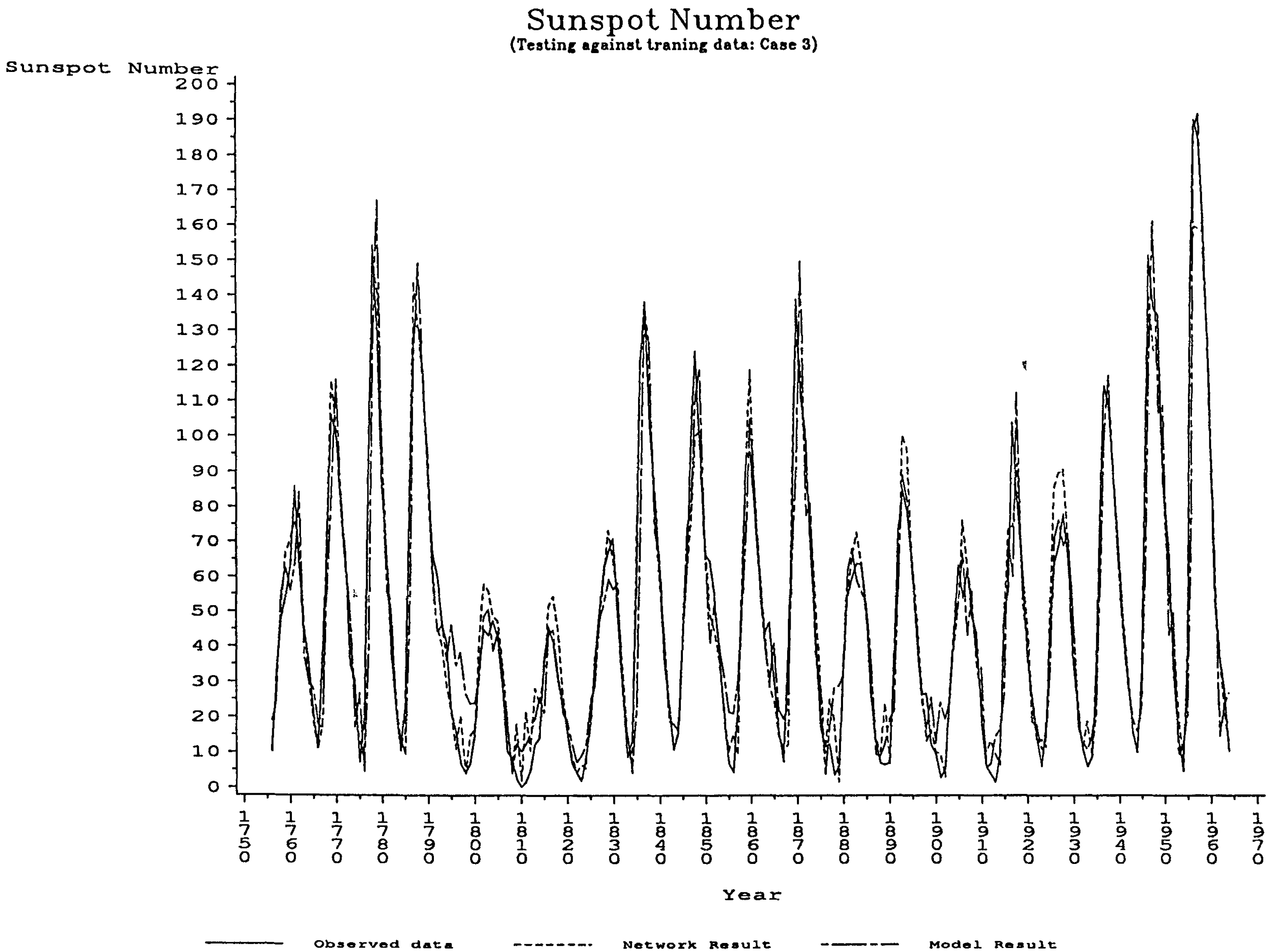

Figure 6. Results on training data in time period 1755-1964,
compared to the model of Morris(1977)

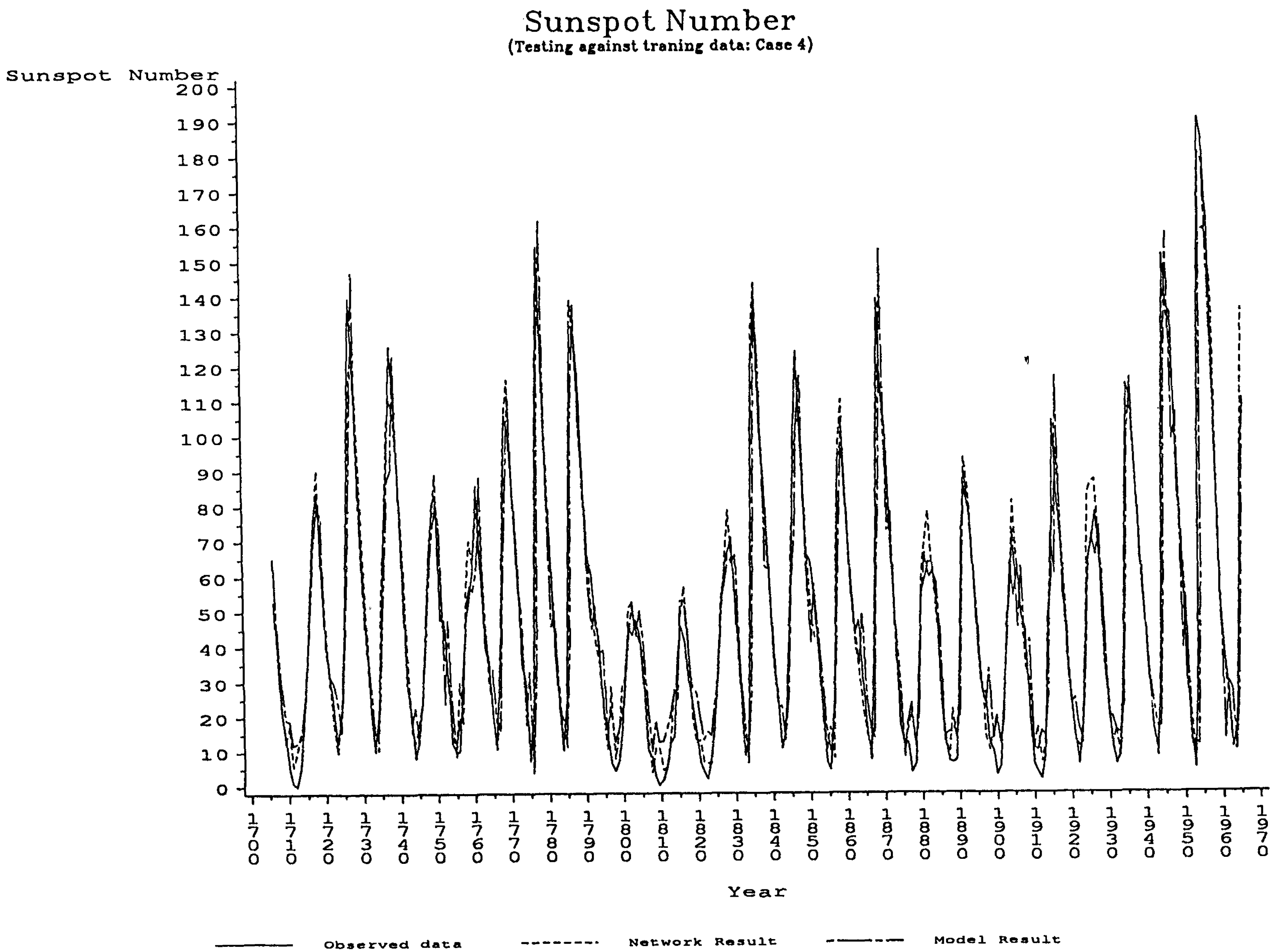

Figure 7. Results on training data in time period 1705-1968, compared to the model of Per Hokstad(1983)

Figures 4 to 7 show how well the network and the AR model predictions match with the actual data. From the figures it is clear that the network prediction follows the actual data more closely than the model prediction.

5.2 Test 2: Lag One Forecasting

In this test, each AR model and our network were asked to forecast the next year's sunspot numbers immediately following the training time period, and such predictions are carried out for a period of 12 years that approximates one sunspot cycle. In other words, this test requires models to look only one year ahead during forecasting, and we always use actual (observed) data at years $t-1$, $t-2$, etc. to forecast the number at year t.

The quantitative result is shown in Table 2. Again we see that the MSE's are higher for all four AR models than for the corresponding neural network. The neural network performs better than the corresponding AR model as verified from the graphical comparison also.

Case		1	2	3	4
MSE	Models	692.07	501.02	292.27	438.02
	Network	150.58	376.37	120.98	277.13
$\frac{\sqrt{MSE}}{mean}$	Models	0.61	0.45	0.29	0.29
	Network	0.28	0.39	0.19	0.23

Table 2. Short-range prediction test

Figures 8 to 11 show the comparisons of the results produced by AR models and our network. (In Figures 8 to 15 —— represents the observed data, ······ represents the network results and, —·— represents the model results.)

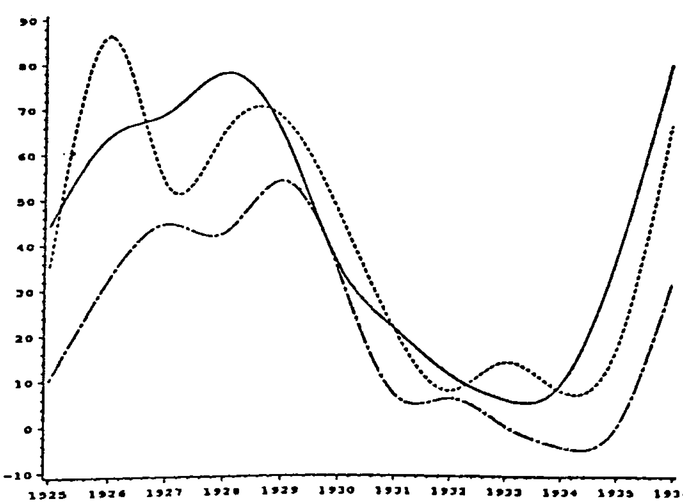

Figure 8. A 12-year prediction for the time period 1749-1924: comparison between the network and Schaerf (1964) model

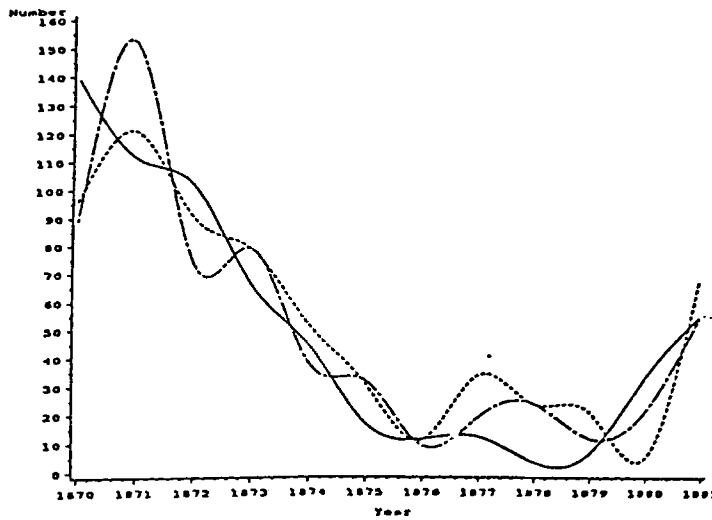

Figure 9. A 12-year prediction for the time period 1770-1869: comparison between the network and Box and Jenkins(1970) model

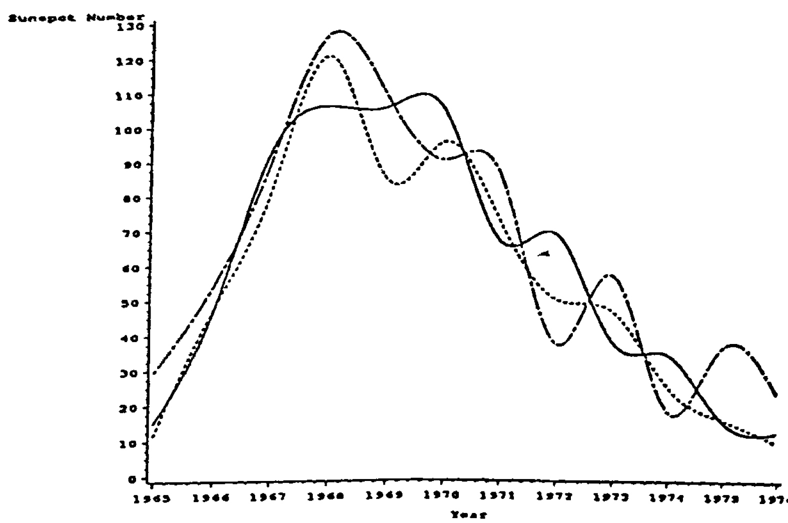

Figure 10. A 12-year prediction for the time period 1755-1964: comparison between the network and Morris (1977) model

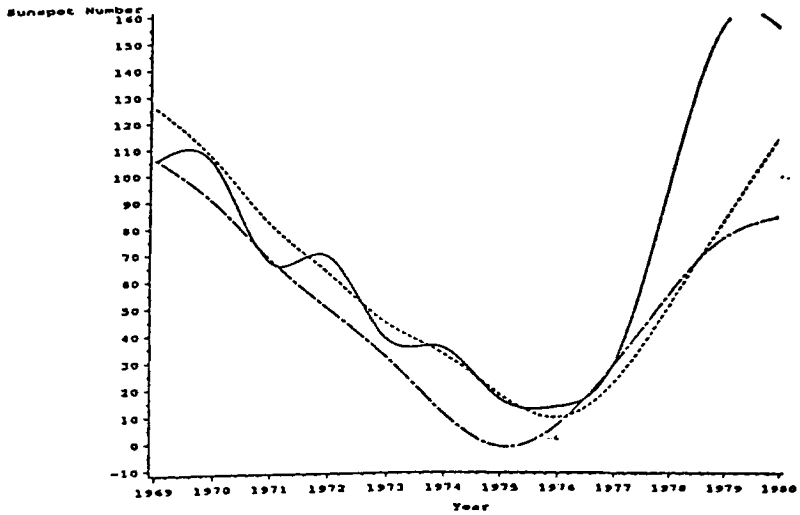

Figure 11. 12-year prediction for the time period 1705-1968: comparison between the network and the model of Per Hokstad (1983)

5.3 Test 3: 12-Year Forecasts

As mentioned before, these forecasts are based on the data predicted by the AR models and the neural network using only the training set. More specifically, the forecasts are made for lag times 1 through 12 using the data in the training set only. The results are shown in Table 3.

Case		1	2	3	4
MSE	Models	4298.64	810.44	514.21	1218.98
	Network	673.55	342.60	854.07	829.60
$\frac{\sqrt{MSE}}{mean}$	Models	1.52	0.57	0.38	0.48
	Network	0.60	0.37	0.50	0.39

Table 3. The result of a 12-year free prediction

It is observed that the network forecasts, measured in terms of MSE, are worse than the forecasts given by the AR model of Morris (1977). Neural network forecasts are superior (smaller MSE) in the remaining three cases. However, a closer look at the graphs indicates that this could be explained by the large difference between forecast value and observed value at only a few points; the network forecasts seem to follow the actually observed data more closely for most years.

5.4 Test 4: 55-Year Forecasts

This is a long-range forecasting task. The task is simple but as a test of a model it is much more difficult than the previous tests; some AR models simply give no prediction at all after one or two cycles. Table 4 shows the quantitative results.

Case		1	2	3	4
MSE	Models	7162.73	840.91	2081.26	2742.06
	Network	1669.14	343.27	1227.08	3435.38
$\frac{\sqrt{MSE}}{mean}$	Models	1.26	0.73	0.66	0.76
	Network	0.61	0.47	0.51	0.85

Table 4. The result of 55-year forecasts

The forecasts produced by Model 1 and 2 decay rapidly. Model 1 is a 9-th order model, and so is Model 3. The only difference between these two models is that the contribution of the lag 9 term (coefficient of x_{t-9}) of Model 1 is smaller than in Model 3.

In these 55-year forecasts, the performance of the neural network is superior to the forecasts of the corresponding AR models for the first three cases and inferior in the fourth case. However, as in the 12-year forecasts above, the mean square error does not explain the behavior of models as well as the comparisons in Figures 12 to 15. Note that the interval between sunspot cycle 18 and 19 (years 1940 to 1950) is only about 9.5 years, inconsistent with the other years. Even though both the neural network and Per Hokstad's model (see Figure 15) missed the peak positions of the cycles, our network does perform better in predicting the maximum number of sunspots in those cycles.

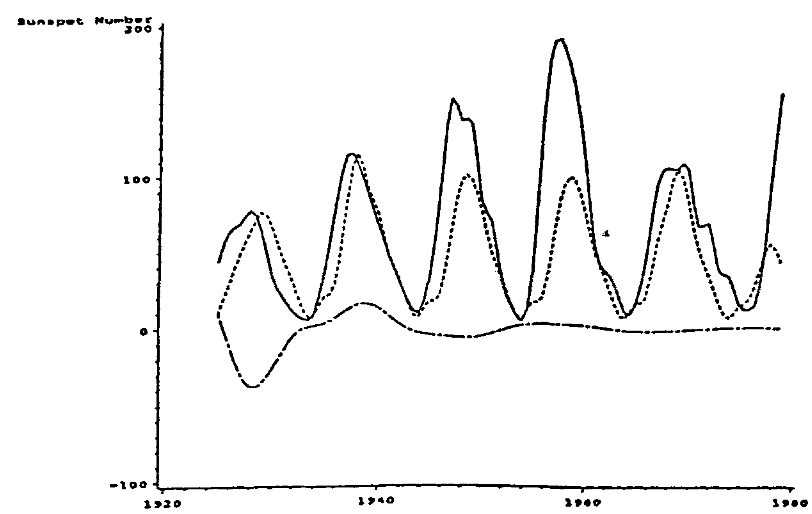

Figure 12. Result of a 55-year free prediction for the time period 1749-1924: comparison between network and the model of Schaerf (1964)

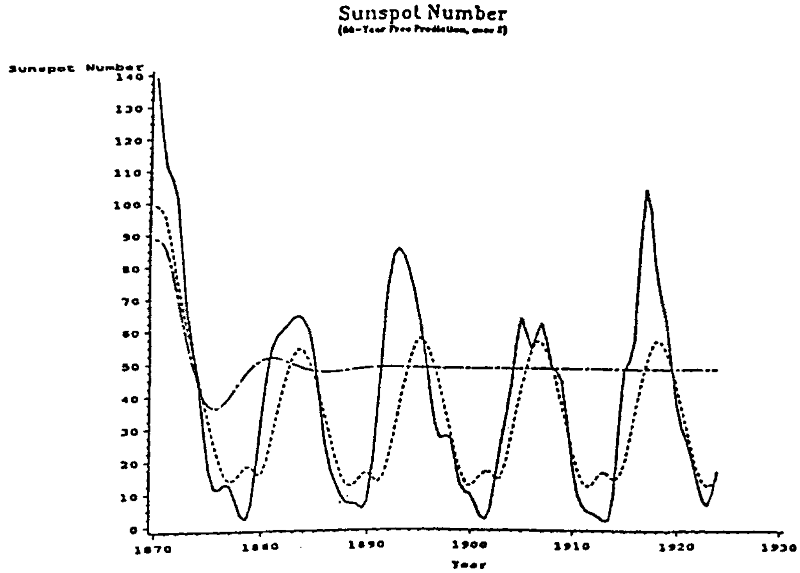

Figure 13. Result of a 55-year free prediction for the time period 1770-1869: comparison between network and the model of Box and Jenkins (1970)

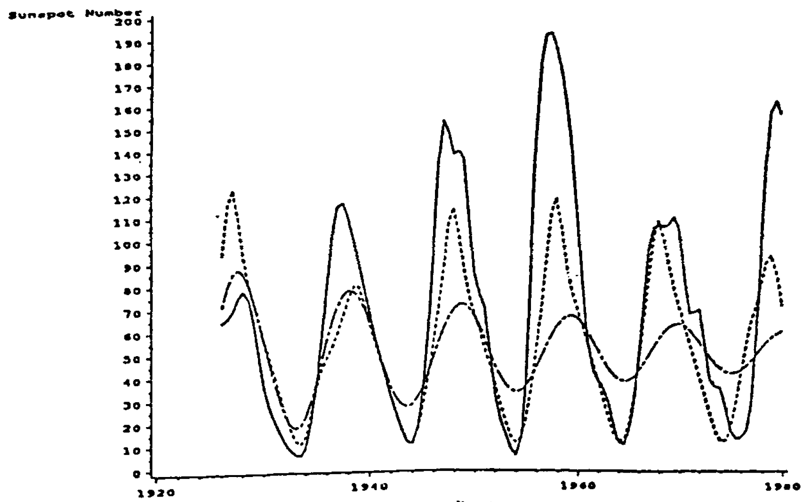

Figure 14. Result of a 55-year free prediction for the time period 1755-1964: the comparison between network and the model of Morris (1977)

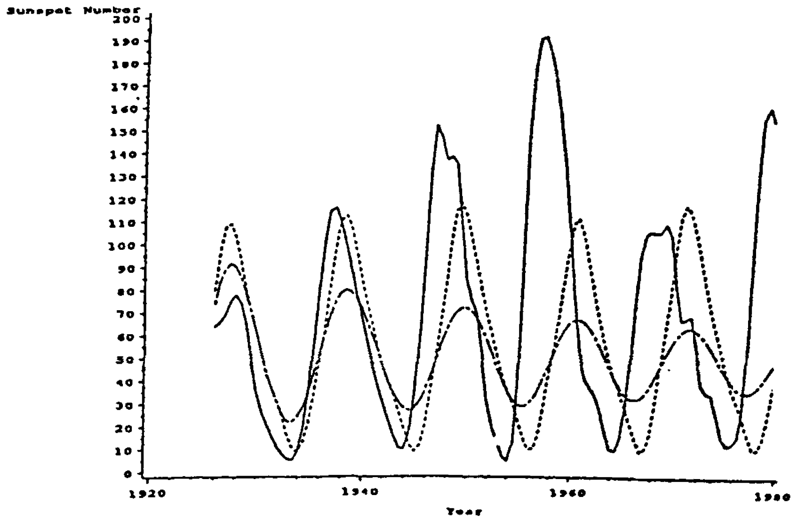

Figure 15. Result of a 55-year free prediction for the time period 1705-1968: the comparison between network and the model of Per Hokstad (1983)

6 Discussion

Our preliminary results indicate that the proposed neural network can produce very good results in modeling sunspot cycles, especially when doing long-range forecasting. Since the network architecture and training scheme used in this work are both general, we have reason to believe that this method could be applied to other temporal sequence processing problems, such as in dynamic system recognition, behavior prediction, and control. More theoretical and empirical work needs to be done to explore the capabilities and limitations of these types of networks (and other recurrent networks as well).

At this stage, we do not have a clear explanation regarding the internal behavior of the network. It is not a coincidence that the best results were produced by the network with the same number of hidden units as the generally agreed upon length of sunspot cycle. One speculation is that the hidden units perform some kind of state transitions, like finite automata, to realize the periodic behavior of sunspot cycles. It follows that the knowledge representation within the hidden layer must be localized. But this conjecture is not supported by analysis of the outputs of hidden units. During a 55-year forecast test, all hidden units actively contribute some potentials to the network output throughout the entire process. All of the hidden units have similar cyclic properties, but no state transition could be found.

We also noticed the phenomenon that when more than 11 hidden units are used in a neural network the majority of those units become inactive; namely, they produce constant outputs throughout training cycles. Although the reason is not clear yet, this phenomenon does hint that we may be able to construct a self-organizing system, using a systematic way to eliminate the redundant hidden units and establish the final network. This method might be more useful in forecasting problems where the processed temporal sequences do not have explicitly identifiable cycles.

We believe that the prediction performance of our network can be improved by further adjusting network configuration and training parameters. It is our belief that the network configuration design can be aided by statistical modeling. This is because models 3 and 4, which contain 9-year and 18-year autoregressive components, do better prediction than the other two models that do not contain these terms; the neural network also contains 11 units in the hidden layer (a number closer to 9 and 10.5; 10.5 is considered to be the cycle length in sunspot numbers).

Acknowledgments

The first author is most grateful to Dr. Curt Burgess in Psychology Department, Syracuse University, for kindly providing computing facilities.

References

[1] Anderson, T. W. *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York (1971).

[2] Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, Holden-Day, Inc., San Francisco (1970).

[3] Elman, J. L., "Finding structure in time," CRL technical report 8801, Center for Research in Language, University of California, San Diego (1988).

[4] Elman, J. L., "Representation and structure in connectionist models," CRL technical report 8903, CRL, University of California, San Diego (1989).

[5] Gallant, S. I., et al., "Experiments with sequential associative memories," College of Computer Science, Northeastern University (1988).

[6] Izenman, A. J., "J. R. Wolf and the Zurich sunspot relative numbers," *The Mathematical Intelligence*, 7, 1, 27–33 (1985).

[7] Jordan, M., "Serial order: A parallel distributed processing approach," ICS technical report 8604, Institute for Cognitive Science, University of California, San Diego (1986).

[8] Moran, P. A. P., "Some experiments on the prediction of sunspot numbers," *J. R. Statist. Soc.*, B16, 112–117 (1954).

[9] Morris, M. J., "Forecasting the sunspot cycle," *JRSS,* A 140, 4, 437–468 (1977).

[10] Per Hokstad, "A method for diagnostic checking of time series models," *Journal of Time Series Analysis*, 4, 3, 177–183 (1983).

[11] Rumelhart, D. E., et al., "Learning internal representations by error propagation," *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol. 1, D. E. Rumelhart et al. (Eds.), 318–362 (1986).

[12] Schaerf, M. C., "Estimation of the covariance and autoregressive structure of a stationary time series," technical report, Dept. of Statistics, Stanford University, Stanford, California (1964).

[13] Schove, D. J. (Ed.), *Sunspot Cycles*, Hutchinson Ross Publishing Company, Stroudsburg, Pennsylvania (1983).

[14] Williams, R. J., et al., "A learning algorithm for continually running fully recurrent neural networks," ICS report 8805, ICS, University of California at San Diego (1988).

[15] Woodward, W. A., et al., "New ARMA models for Wolfer's sunspot data," *Comm. Statist. -Simula Computa.*, B7, 1, 97–115 (1978).

[16] Yule, G. U., "On a method of investigating periodicities in distributed time series with special reference to Wolfer's sunspot numbers," *Phil. Trans.*, A226, 267-298 (1927).