Neural network prediction of solar activity

R. A. Calvo, H. A. Ceccatto and R. D. Piacentini

Instituto de Física Rosario, CONICET-UNR, Rosario, Argentina

Abstract

The neural network technique is used to analyze the time series of solar activity, as measured through the relative Wolf number. Firstly, the embedding dimension of the time-series characteristic attractor is obtained. Secondly, after describing the design and training of the net, the performance of the present approach in forecasting yearly mean sunspot numbers is favorably compared to that of conventional statistical methods. Finally, predictions for the remaining part of the 22th and the whole 23th cycle are presented.

1. Introduction

The solar activity is characterized, among other quantities, by means of the relative Wolf sunspot number. In the present work we will concentrate our studies on this index, since it has the largest data set compared to other activity measurements. In particular, we will consider the predictability of its annual mean value. The time series of the Wolf number shows a frequency spectrum with a large maximum at around 11 years (the usually cited short period), and also other periods between 9.5 and 13.7 years. Even longer periodicities from 58 to 200 years have been determined (Charvàtovà and Strestik 1991, and references therein). This complex behavior, and the superimposed noisy structure due to measurement errors, require sophisticated methods in order to reconstruct the solar-activity intrinsic dynamics.

The use of artificial neural networks has been recognized recently as a promising way of making predictions on temporal series with chaotic or irregular behavior (Weigend, Huberman, and Rumelhart 1990; Elsner and Tsonis 1992). This technique has already been applied in the framework of solar-terrestrial physics for the prediction of geomagnetic induced currents and storms (Lundstedt 1992; Lundstedt and Wintoff

1993), and as a way of recognizing a pattern in the onset of a new sunspot cycle (Koons and Gorney 1990). In this work we use neural networks to determine the solar-activity intrinsic dynamics, as measured by the relative Wolf number. First, in section 2 we give a brief description of the method, and in section 3 we determine the embedding dimension of the Wolf number time-series attractor (which is needed to design the network's architecture). In section 4 we describe the training of the network, compare the performance of the method with those of conventional statistical techniques, and forecast the remaining part of the 22th cycle and the complete 23th cycle. Finally, in section 5 we discuss the results and present some conclusions.

2. Neural networks: A capsule introduction

Neural networks are parallel computational structures of highly interconnected simple processors -called neurons-, which simulate to some extent the structure and functioning of the brain (Rumelhart and McClelland 1986). In particular, in this work we will use the so called "feed-forward" neural networks. These architectures have a group of neurons -the input layer-,which are fed by external stimuli **I,** where **I** is a column vector (Fig.1). The input units send these stimuli to hidden neurons (not connected to the environment), grouped in one or more internal layers. These hidden units process the information they receive, and pass their results to the last group of neurons -the output layer-. Neurons in the output layer produce the final response **O** to the external stimuli . The units are connected through information channels, whose strengths ("weights") have to be determined in order to properly relate inputs to desired outputs. This process is known as network training, and the algorithm which performs the weights adjustment is the so called "backpropagation rule" (Rumelhart, Hinton, and Williams 1986). It essentially consists of a gradient-descent algorithm to reduce the error between actual and desired network outputs, which modifies weights going backward from the output-layer to the input-layer connections.

The computations carried out inside each neuron amount to: i) performing the weighted average of its impinging inputs, ii) sending this average through a (biased) sigmoid function, and iii) forwarding the sigmoid-function output to the next layer of neurons (Fig.2). These calculations are generally performed synchronously by all the neurons in a given layer, so that the stimulus-response delay depends on the number of layers. Despite the simplicity of the calculations involved, a feed-forward neural network with enough number of hidden units is capable of performing any arbitrary mapping of n-dimensional inputs to m-dimensional outputs (Cybenko 1988; 1989). Although in general the optimal network architecture able to perform this task is not known, the nature of the problem often gives hints in this sense (see next section).

Typically, solving a problem by using neural networks requires: a) choosing a suitable network architecture (number of layers and neurons in each layer), b) selecting a large, representative set of training patterns (input-output pairs carrying enough information on the data set), c) training the network to relate the inputs to the corresponding outputs, by modifying weights according to the backpropagation method. As mentioned above, this corresponds to minimizing an error function

$$\mathrm{Err} = \sum_{k=1}^{N} (\mathbf{T}_k - \mathbf{O}_k)^2 \quad , (1)$$

where **T**k is the desired output ("target"), **O**k is the actual output produced by the network as a response to input **I**k, and N is the total number of training patterns. In general, every pair (**I**k,**T**k) has to be presented many times to the network in order for it to (approximately) learn the mapping. The length of the training process is usually measured in terms of single presentations of the whole training set ("epochs"). A frequent pitfall in this part of the process is getting trapped in high local minima of the error function given by equation (1).

A successful training experiment generally allows the network to capture the essential relationships among inputs and outputs. In such cases, the trained net shows remarkably generalization capabilities, being able to correctly relate inputs to outputs not included in the training set. However, the architecture used is crucial: smaller-than-needed nets do not learn the examples, while bigger-than-required ones, in most cases overfit the data, learning undesirable (noisy) features which degrade its generalization performance. In the last case, a way to cope with the problem is to keep a small number of input-output examples without being presented to the network. The generalization performance is then monitored on these examples while learning the rest of the training set ("cross-validation"). This performance is usually appraised by following the time evolution of the Average Relative Variance:

$$\mathrm{ARV} = \frac{\mathrm{Err}}{\sigma^2 M} . \quad (2)$$

Here Err is given by equation 1 with k running from 1 to M, where M is the number of patterns in the cross-validation set, and σ2 is the variance of the validation targets. Notice that ARV=1 for a network which only learns to predict the target average. In section 4 we will use the cross-validation method to stop the training process before overfitting the noisy time series of solar activity.

Other useful quantities to characterize the network's performance are the root mean square error (RMSE) and the correlation coefficient (CC) between predicted and actual targets, and the explained variance (EV), defined as the ratio of variances of predicted and desired outputs.

3. The embedding dimension of the solar dynamics attractor

In this section we estimate the minimum number of variables required to reconstruct the solar dynamics by determining the embedding dimension of its time-series attractor (Packard, Farmer and Shaw 1980; Grassberger and Procaccia 1983). The method is well known, and has been recently applied to meteorological data (Fraedrich 1986; Henderson and Wells 1988). It essentially amounts to:

a) embedding the attractor in a μ-dimensional pseudophase space spanned by the time series and its lagged values, so that a point in this space is described by:

$$\mathbf{x}(t) = (x(t), x(t-1), \ldots, x(t-\mu+1)) \quad (3)$$

where x(t) is the time series and the lagged time constant is taken to be 1 year in this case,

b) calculating the cumulative correlation function:

$$C(r) = \frac{1}{N^2}\sum_{ij=1}^{N} H(r - r_{ij})$$

where H is the Heaviside function (H(x)=0 if x<=0, H(x)=1 otherwise), rij is the Euclidian distance between points i and j, and r is a variable distance, and

c) estimating the slope d(μ) of the linear part of the C(r) vs. ln(r) curve, by fitting with a least-square line.

By repeating these steps for increasing values of μ (μ=1,2,3..), one gets succesive estimates of the attractor dimension. If the slopes converge to a limiting value d_∞ for a large value μ=M, then d_∞=d(M) is the true correlation dimension of the attractor, and M is a measure of the number of variables sufficient to model the dynamics. A non integer value of d_∞ indicates the presence of a strange attractor, signifying deterministic but chaotic dynamics. Notice that for a purely random time series there will be no saturation of the slopes. In our case, following steps a) to c) leads to the results shown in Figures 3 and 4. In particular, from Figure 4 we see that there is a saturation of slopes for M aproximately 12. We note that this number is close to the most important periodicity of the sunspot cycle.

We will assume that the solar dynamics is given by x(t)=F(**x**(t-1)), with F some unknown function. The above calculation suggests the use of a network with 12 units in its input layer and a single unit in the output one. Thus, the output of the network - sunspot number at year t- becomes a function of the previous sunspot values at (t-1),...,(t-12). Training of the network then determines the dynamical function F. Notice, however, that the number of hidden units has not been specified yet. Following the thumb's rule that the number of training patterns should be 5 to 10 times the number of network parameters, we have used 3 hidden units. The resulting 12:3:1 network has 43 adjustable parameters -39 weights plus 4 sigmoid function biases-, while the complete record 1700-1991 of annual mean sunspot numbers comprises 292 values.

4. Neural network forecasts of the Wolf number

After describing the neural network approach in previous sections, we will present here the results obtained using this method to predict the annual mean sunspot number. First, in order to establish the reliability of our approach, we will compare the neural network performance with those of more standard forecasting techniques. We will do it with the methods considered to be the best for this application (Cerrito 1992).

Using the data for the years 1770-1869, Box and Jenkins (1976) showed that, whithin their technique of time series analysis, the optimal model for this data is a third order autoregressive model (AR(3)). However, they also stated that the model is not a particularly good fit of the data, which lead Wei (1989) to suggest that the sunspots are periodic of size 11. Cerrito (1990) proposed a new technique using the nonparametric kernel density estimator. The use of the density estimator requires the assumption that the data be strictly stationary, completely excluding the possibility of periodicity. Forecasts for the years 1870-1889 obtained with these two methods are shown in Figures 5a and 5b.

To be able to compare with these results, we used the sunspot values for the years 1770-1869 as the training set (data set in the statistical methods), and the interval 1870-1889 for cross-validation (prediction set). Since for this smaller training set the network's architecture is too large, we avoided overtraining by monitoring the generalization performance during learning, and terminated the process when there was no longer improvement. After 20 experiments with different initialization of weights and biases, Figure 5c shows the scattered predictions of the best 19 experiments. We also show in Figure 5d the best results obtained in a single trainning process, which will be taken as the neural network predictions in order to compare with other methods.

The performances of the different techniques can be appraised from table 1. As can be seen from these values, the neural network predictions are much better than the corresponding to other methods. Furthermore, Figure 5c indicates that the network's ability to capture the intrinsic dynamics of solar activity does not rely crucially on a fortunate training process. Moreover, the prediction power seems not to deteriorate sensibly for long-term forecasting (a whole cycle in this case).

After validating the neural network capabilities, in the rest of this section we present our predictions for the rest of the 22th cycle and the complete 23th cycle. As above, we have first split the whole record (1700-1991) in a training set (1700-1975,1987-1991), and a cross-validation interval (1976-1986), corresponding approximately to the last complete cycle. We performed 10 training experiments, following the network's performance in a multiple-step prediction over the cross-validation interval. Figures 6a,b show respectively the results for the 9 best experiments and for the training process having the smallest error on the independent data set. In the training section of the record we found:

while in the cross-validation interval the performance was:

RMSE=9.91 ARV=0.032 CC=0.99 EV=0.82.

The predicted sunspot numbers for the years 1992-2008 are given in Table 2. The error to be expected in these forecasts can be estimated from Figures 5d and 6b, and from the above RMSE values.

5. Conclusions

We have used neural networks to study the solar dynamics, as measured by the annual mean value of the Wolf number. First, we determined the embedding dimension of the time-series attractor, and used this number to define the network's architecture. Secondly, we compared the neural network approach with conventional techniques in order to establish the reliability of our results. Finally, we presented the predictions for the remaining part of the 22th and the whole 23th cycle.

From this study we conclude, in agreement with previous works (Weigend, Huberman and Rumelhart 1990, Elsner and Tsonis 1992), that neural networks are a reliable tool for time series analysis. In particular, they seem to be able to capture the intrinsic dynamics of solar activity, producing good long-term forecastings for periods of at least a complete cycle. Moreover, Figures 5d) and 6b) show no obvious deterioration in the forecasting performance in a multiple-step prediction. This seems to indicate

(Elsner and Tsonis 1992) that the dynamics is essentially non-chaotic in this time scale, but perturbed by a fairly large amount of noise.

Acknowledgments

The authors thank Dra. Marta Rovira and Lic. Horacio Ghielmeti of the Instituto de Astronomía y Física del Espacio, Buenos Aires, Argentina, and Dr. Bernardo Huberman of Xerox Palo Alto Research Centre, USA; for fruitful discussions on different parts of the present work.

Table 1

	RMSE	ARV	CC	EV
AR(3)	31.2	0.68	0.92	0.12
Kernel Density	26.5	0.49	0.87	0.31
Neural Network	11.2	0.09	0.98	0.59

Table 2

1992	120.6
1993	93.9
1994	55.5
1995	23.1
1996	7.8
1997	8.7
1998	39.9
1999	111.3
2000	166.5
2001	161.4
2002	142.5
2003	103.5
2004	75.0
2005	42.0
2006	17.7
2007	5.4
2008	8.7

References

Box G. and Jenkins G. 1976, Time Series Analysis, Revised Edition (Oakland: Holden)

Cerrito P. B. 1990, *Astron. J.* **100**, 579

Cerrito P. B. 1992, *The Astrophysical Journal*, **393**, 795

Charvàtovà, I. and Strestik J. 1991, *Bull. Astron. Inst. Czechosl.* **42**, 90

Cybenko G. 1988, Technical Report Department of Computer Science, Tufts University, Medford, MA

Cybenko G. 1989, *Math Control Signals Syst.*, **2**, 303

Elsner J. B. and Tsonis A. A. 1992, *Bulletin American Meteorological Society*, **73**, 49

Fraedrich K. 1986, *J. Atmos. Sci.*, **432**, 419

Grassberger P. and Procaccia I. 1983, *Phys. Rev. Lett.*, **50**, 346

Henderson H. W. y Wells R. 1988, *Adv. Geophys.* **30**, 205

Koons H. C. and Gorney D. J. 1990*, EOS, Trans. AGU*, **71**, 677

Lundstedt H. 1992, *Planet. Space Sci.*, **40**, 457

Lundstedt H. and Wintoff P. 1993, private communication.

Packard N. H., Farmer J. D. and Shaw R. S. 1980, *Phys. Rev. Lett.*, **45**, 712

Rumelhart D. E. and McClelland J. L 1986, Parallel Distributed Processing (Cambridge: MIT Press)

Rumelhart D. E., Hinton G. E., and Williams R. J. 1986, *Nature,* **323**, 533

Wei, W. 1989, Time Series Analisis (California: Addison Wesley)

Weigend A. S. , Huberman B. A. and Rumelhart D. 1990, *Intl. J. Neural Systems*, **1**, 193

Figure Captions

Fig.1: Schematic diagram of a feed-forward neural network with a single hidden layer.

Fig.2: The jth neuron in the hidden layer. The net input is the weighted average $net_j = \Sigma_i W_{ji} I(i)$. The corresponding output is given by $O_j=F(net_j+\theta_j)$ where F is a sigmoid function and θ_j is the bias of the unit j.

Fig.3: Cumulative correlation function for the sunspot time series. Curves from left to right correspond to embedding dimensions m=2,4,...14 respectively.

Fig.4: Estimated dimension of the sunspot time-series attractor as a function of the embedding dimension.

Fig.5: Comparison between the 1870-1888 observed sunspot numbers (solid line) and the predictions of different models: a) AR(3) (closed circles). The dashed lines give the 95% confidence interval. (Cerrito 1992). b) Idem as a) but for the nonparametric kernel density estimator using the conditional median with the assumption of periodicity. c) Neural network's best 19 out of 20 different trainning experiments. The dashed lines are only to guide the eyes. d) Best trained neural network (closed circles).

Fig.6: Comparison between the 1976-1986 observed values (solid line) and the neural network predictions. a) Predictions of the best 9 out of 10 training experiments for the cross-validation interval. The dashed lines are only to guide the eyes. b) Best trained neural network (closed circles).

Table Captions

Table 1: Forecasting performances of different techniques for the period 1870-1889.

Table 2: Neural Network predictions of sunspot numbers for the years 1992-2008.

R. A. Calvo:

Instituto de Física Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, and Universidad Nacional de Rosario, Boulevard 27 de Febrero 210 bis, 2000 Rosario, Argentina.

Instituto Rosario en Ciencias de la Educación, Consejo Nacional de Investigaciones Científicas y Técnicas, and Universidad Nacional de Rosario, Boulevard 27 de Febrero 210 bis, 2000 Rosario, Argentina.

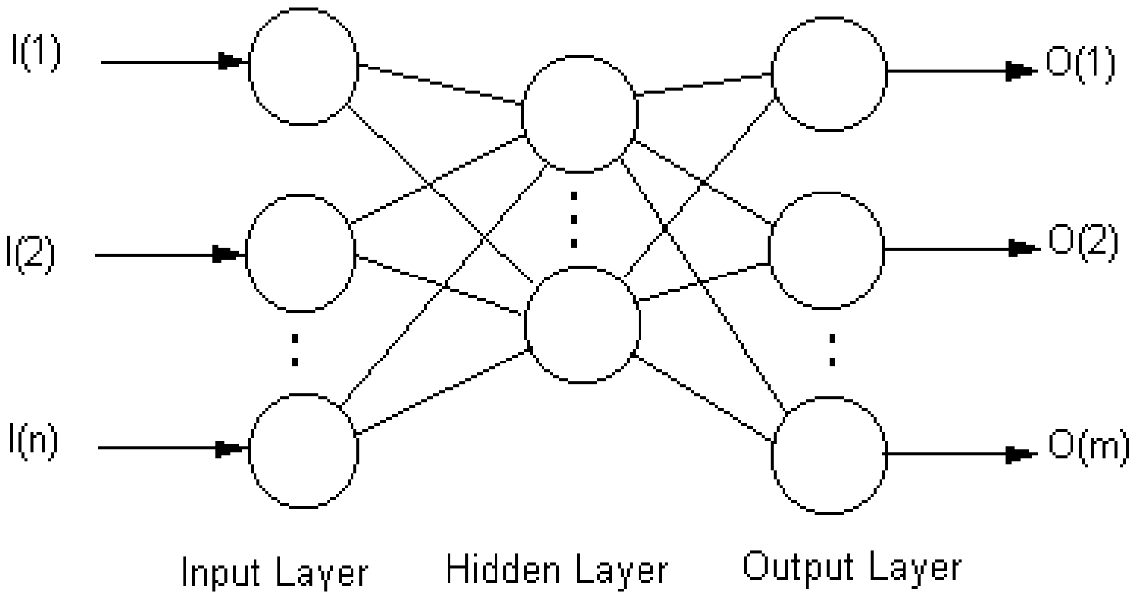
H.A. Ceccatto:

Instituto de Física Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, and Universidad Nacional de Rosario, Boulevard 27 de Febrero 210 bis, 2000 Rosario, Argentina.

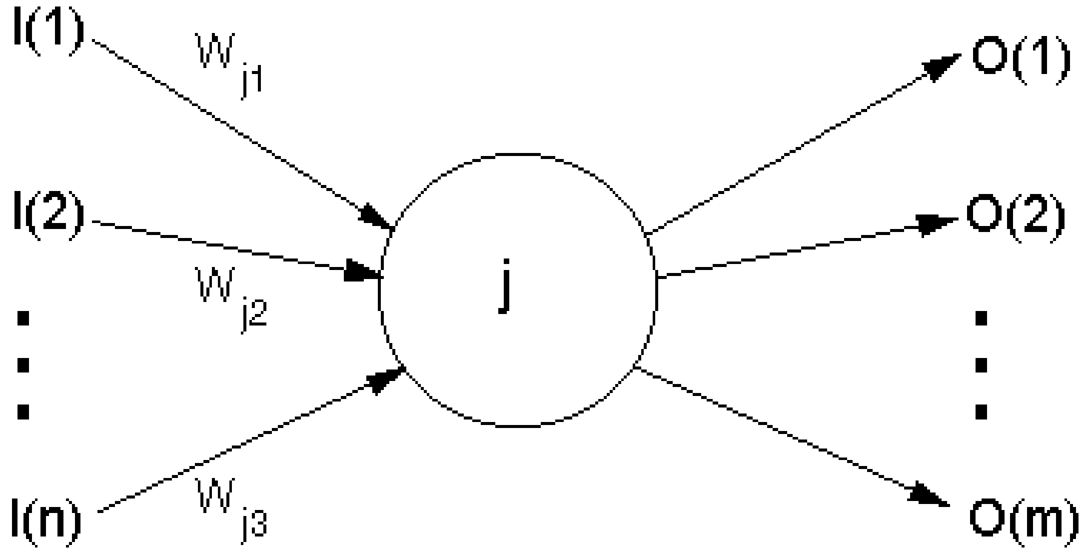
R.D. Piacentini:

Instituto de Física Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, and Universidad Nacional de Rosario, Boulevard 27 de Febrero 210 bis, 2000 Rosario, Argentina.

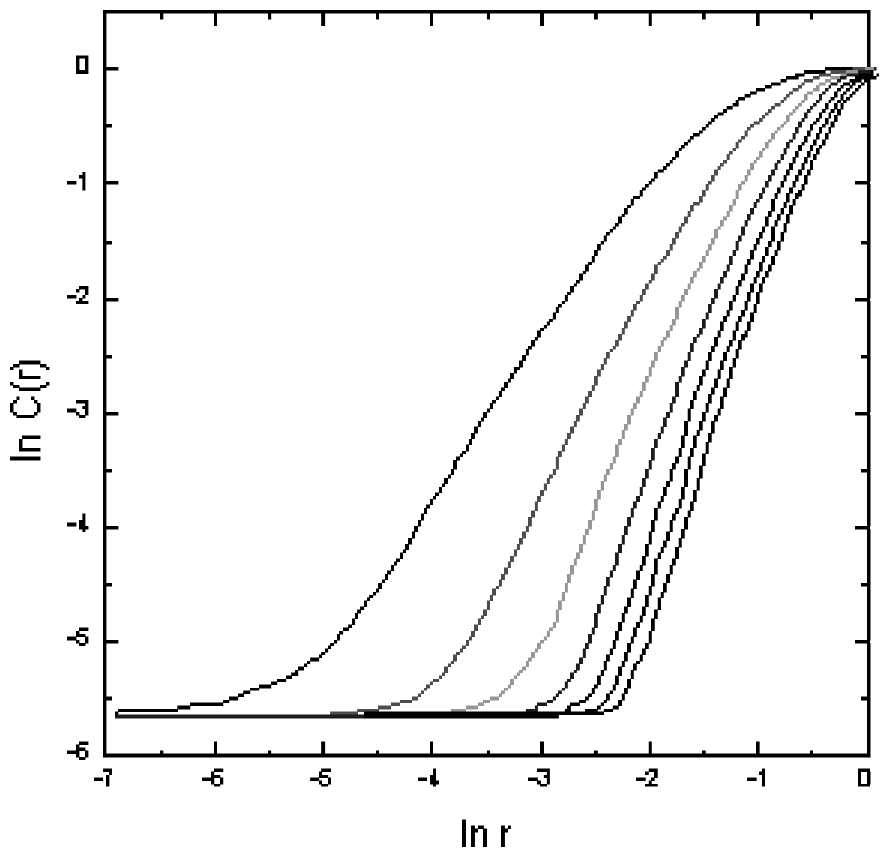
Observatorio Astronómico de Rosario, CC 606, 2000 Rosario, Argentina



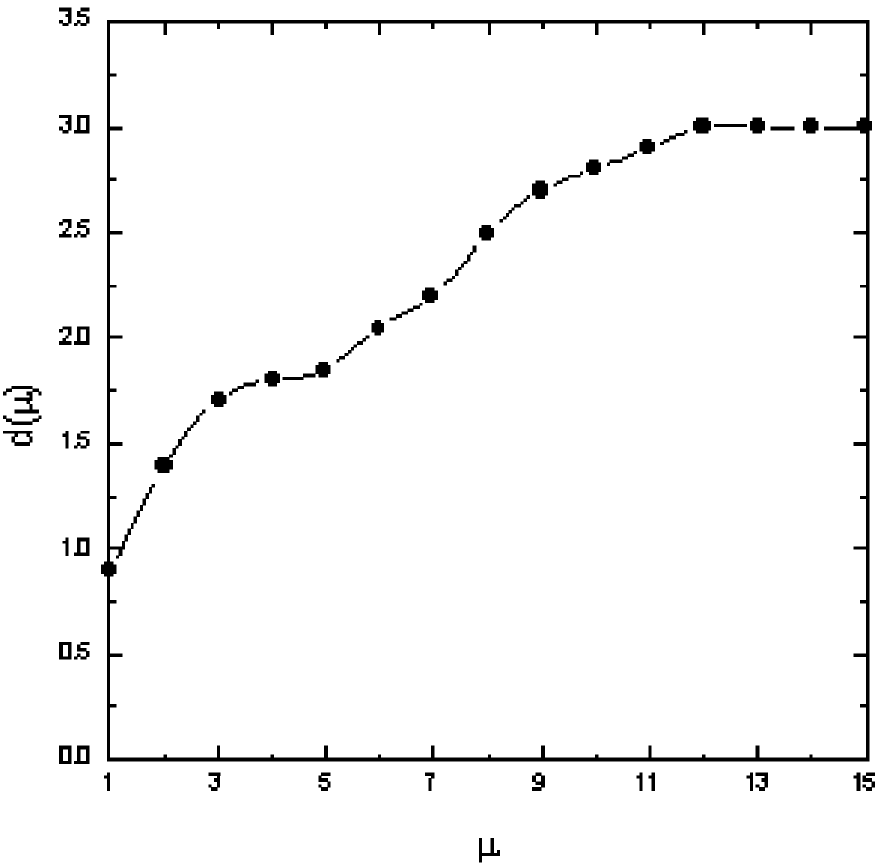
I(1)
I(2)
I(n)
O(1)
O(2)
O(m)
Input Layer
Hidden Layer
Output Layer



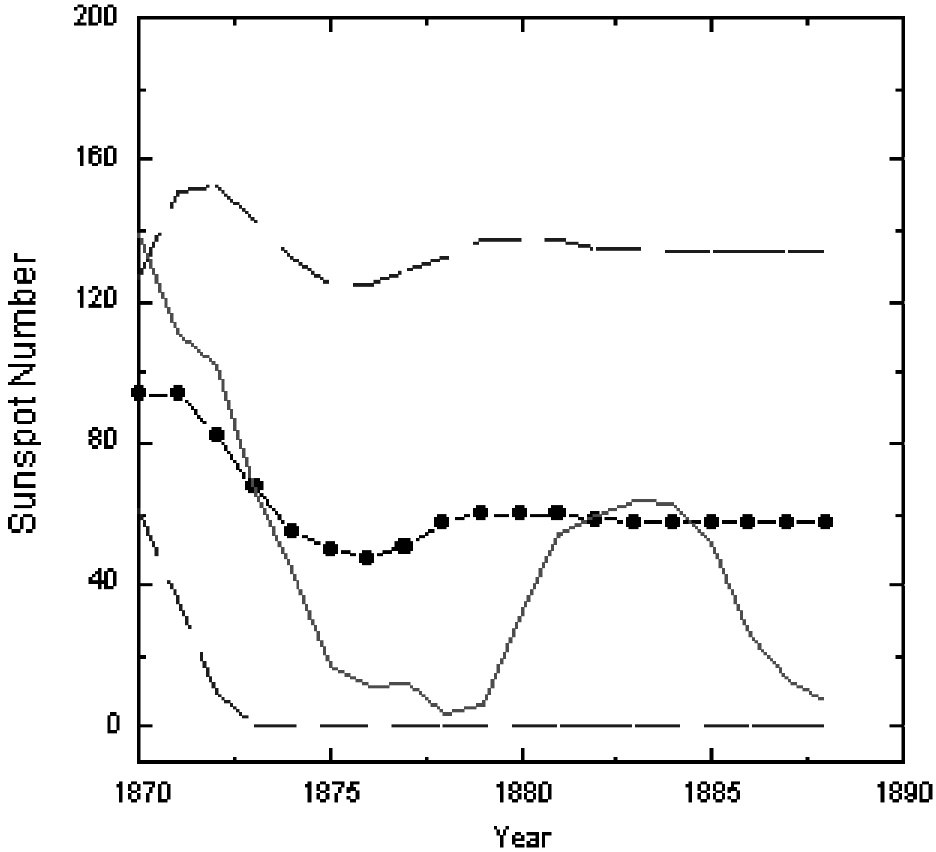
I(1)
I(2)
I(n)
W j1
W j2
W j3
j
O(1)
O(2)
O(m)

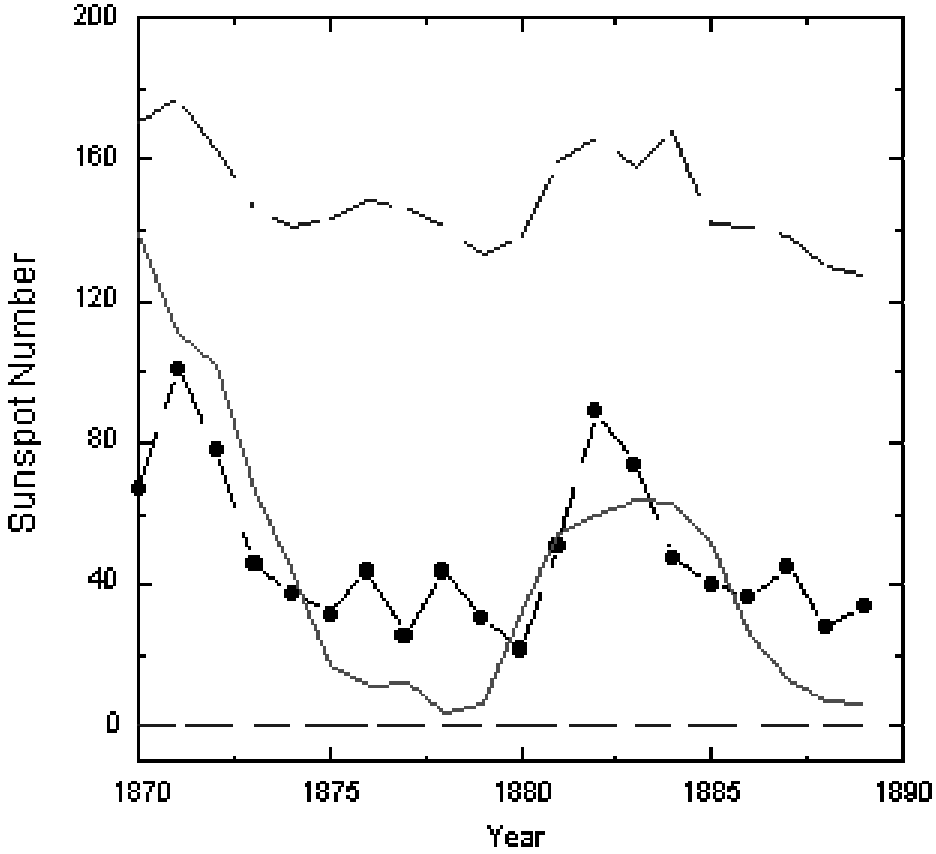


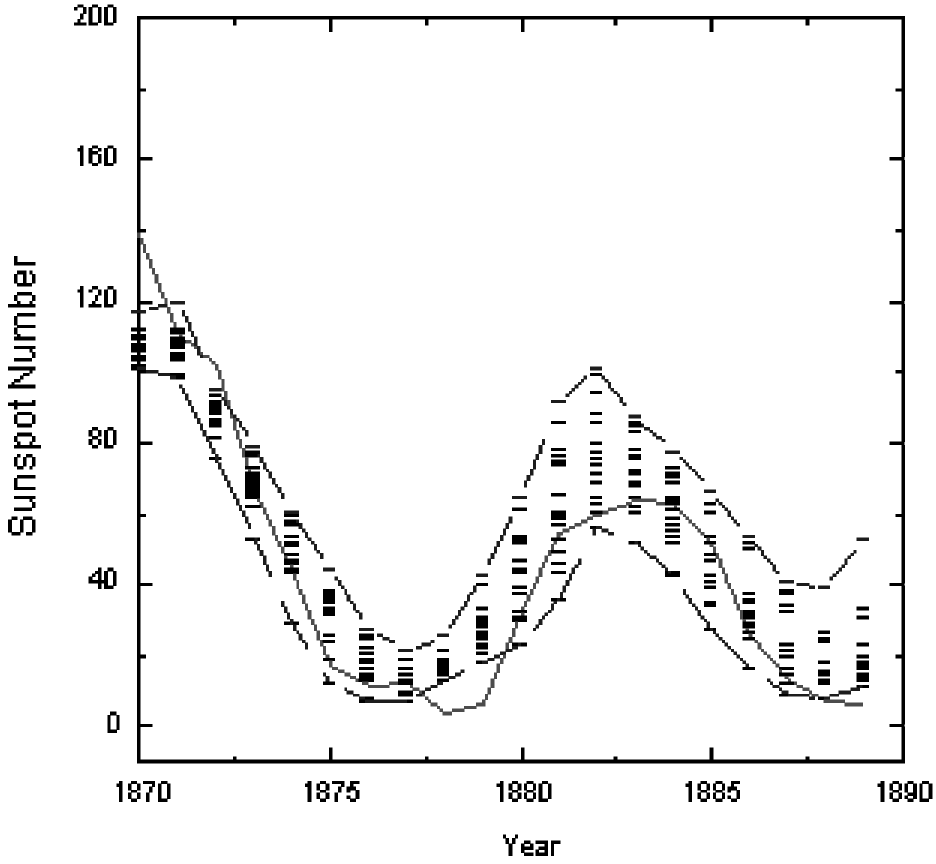
0
-1
-2
-3
-4
-5
-6
ln C(r)
-7
-6
-5
-4
-3
-2
-1
0
ln r

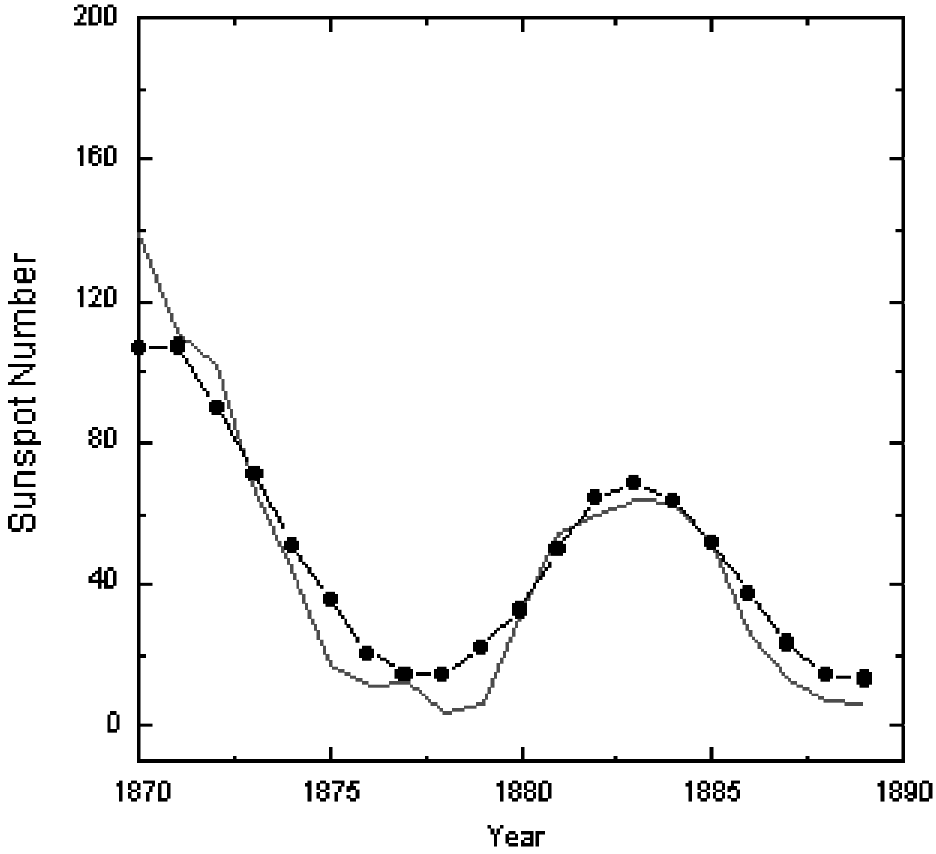


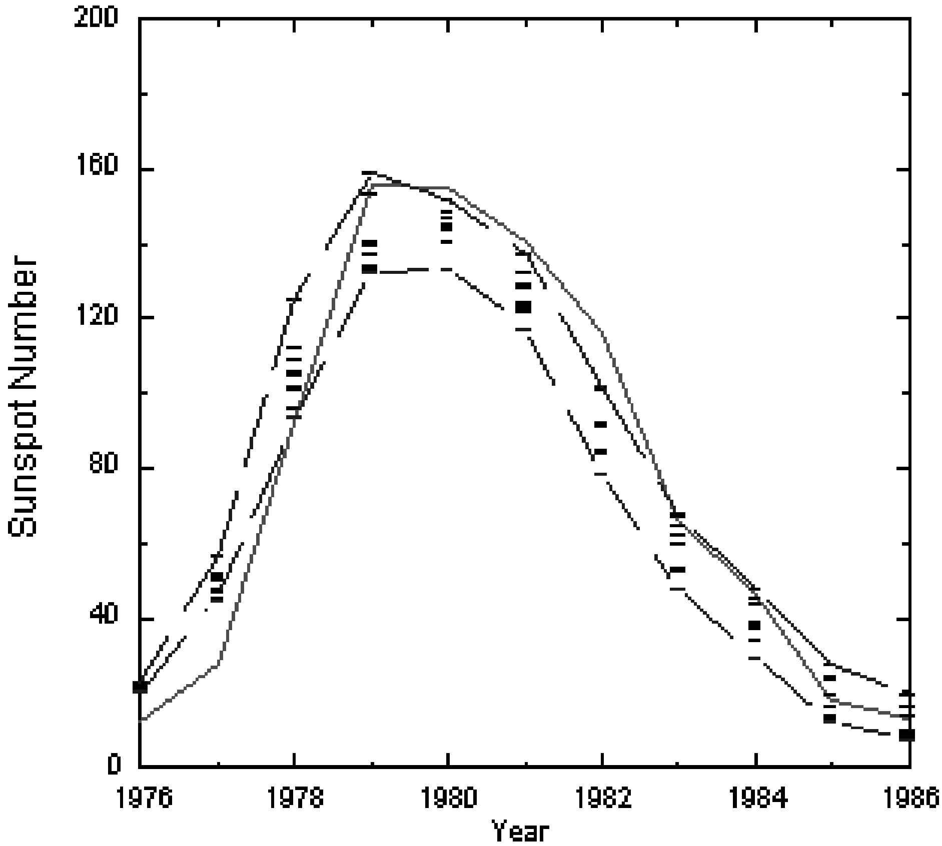
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
d(μ)
1
3
5
7
9
11
13
15
μ

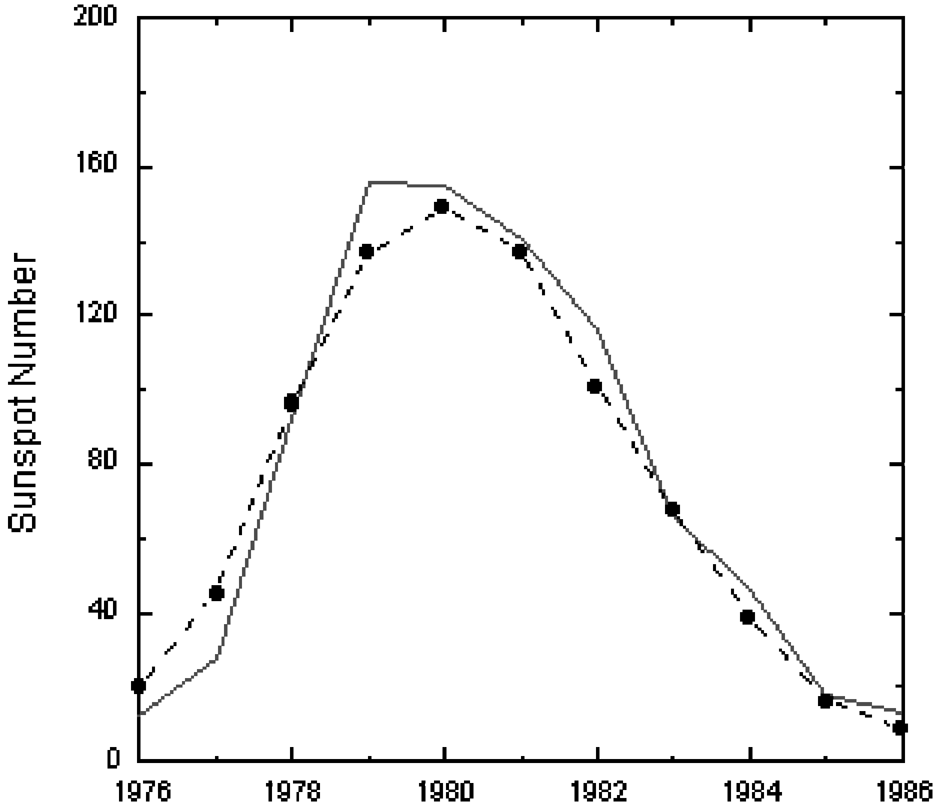

200
160
120
80
40
0
Sunspot Number
1870
1875
1880
1885
1890
Year


200
160
120
80
40
0
Sunspot Number
1870
1875
1880
1885
1890
Year


200
160
120
80
40
0
Sunspot Number
1870
1875
1880
1885
1890
Year


200
160
120
80
40
0
Sunspot Number
1870
1875
1880
1885
1890
Year


200
160
120
80
40
0
Sunspot Number
1976
1978
1980
1982
1984
1986
Year



Sunspot Number
200
160
120
80
40
0
1976
1978
1980
1982
1984
1986