

Time series predictions with neural nets: application to airborne pollen forecasting

C.M. Arizmendi, J.R. Sanchez, N.E. Ramos, and G.I. Ramos

Departamento de Física, Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Av. J.B. Justo 4302, 7600 Mar del Plata, Argentina

Received July 13, 1992; revised February 26, 1993; accepted March 17, 1993

Abstract. Pollen allergy is a common disease causing rhinoconjunctivitis (hay fever) in 5–10% of the population. Medical studies have indicated that pollen related diseases could be highly reduced if future pollen contents in the air could be predicted. In this work we have developed a new forecasting method that applies the ability of neural nets to predict the future behaviour of chaotic systems in order to make accurate predictions of the airborne pollen concentration. The method requires that the neural net be fed with non-zero values, which restricts the method predictions to the period following the start of pollen flight. The operational method outlined here constitutes a different point of view with respect to the more generally used forecasts of time series analysis, which require input of many meteorological parameters. Excellent forecasts were obtained training a neural net by using only the time series pollen concentration values.

Key words: Air pollen forecasting – Back propagation – Chaos prediction – Neural networks – Argentina

Introduction

The most abundant biological particles in the atmosphere, with concentrations that can reach several thousands of particles/m³ are pollen grains and spores (Comtois and Sherknies 1987). For various reasons, both biological and medical, it is interesting to determine the number of pollen grains of different plants that are to be found in the air. Pollen allergy is a common disease causing hay fever in 5–10% of the population; although not a life threatening disease, the symptoms can be very troublesome, and furthermore, the costs to the social sector due to pollen related diseases are high. Self-protection of hay fever patients is possible through the prediction of future pollen contents in the air (Leuschner 1991).

Models to forecast pollen concentration in the air are principally based on pollen and atmospheric weather interactions. Several statistical techniques (O'Rourke 1986, 1989; Goldberg et al. 1988; Comtois and Sherknies 1987, 1989; Moseholm et al. 1987) have been used to predict future atmospheric pollen concentrations from daily weather conditions of the current and of preceding days. In spite of these attempts, it has not been possible to predict the pollen concentrations with great accuracy, and about 25% of the daily pollen forecasts have resulted in failures (Goldberg et al. 1988).

A new analysis of the dynamic characteristics of time series of atmospheric pollen was developed recently by Bianchi et al. (1992). The analysis is carried out by searching for a strange attractor of the pollen time series through the study of the correlation dimension. The value of the dimension found was a non-integer and less than 1, indicating that the system can be described by a one-dimensional chaotic map. This chaotic behaviour places a fundamental limit on long-term prediction while suggesting possibilities for short-term prediction (Farmer and Sidorowich 1987). A short introduction to chaos is given in Bianchi et al. (1992) and references therein. The existence of a one-dimensional function connecting each two neighbour members of the pollen concentration time series indicates that it is possible to apply the predicting abilities of artificial neural nets. Artificial neural networks have been used recently to predict future values of chaotic time series by extracting knowledge from the past (Lapedes and Farber 1987). The problem of prediction, usually framed as extrapolation in time, is then re-framed as interpolation in space. This essential idea is captured by approximating the past relationships between subsequent values by smoothing the interpolation in space.

In this work we have developed an operational forecasting model based on the techniques of neural networks. By training the network using present and past values of the pollen time series accurate future predictions are generated. Traditional approaches to pollen forecasting are consistently surpassed although, in prin-

ciple, it is not possible to predict the day of the year when the pollen flight starts. This kind of forecasting requires phenological dating for which environmental parameters are necessary. The method outlined here provides good forecasting once the pollen season has started. We apply the forecasting method to time series of total pollen concentration; however, it may be used for pollen concentration of any particular species of interest whenever the corresponding time series is available.

Materials and methods

The material used in this work was from our chaos study of pollen series (Bianchi et al. 1992). Data of airborne pollen concentrations were obtained with an automatic and volumetric Burkard pollen and spore trap, situated on the roof of the Facultad de Ciencias Exactas y Naturales, 12 m above ground level. The area surrounding the sample is typical of Mar del Plata. The great distance from the sampling site to the emission sources makes the particular emission spectra unimportant.

Ten litres of air/min were sucked through a $14 \times 2 \text{ mm}^2$ orifice, always orientated against the wind flow; the sucking rate was checked weekly. Behind the slit, a drum rotated at a speed of 2 mm/h. The particles were collected on a cellophane tape (Melinex), 19 mm wide, located just below the orifice. The sticky collecting surface comprised 9 parts vaseline: 1 part paraffin in toluene. The exposed tape was removed from the drum, cut into pieces of 48 mm, corresponding to 24-h intervals, then embedded into a solution of polyvinylalcohol (Gelvatol), water and glycerol and covered with a cover glass. Slides were studied as 12 transects per day. The pollen was counted at a magnification of $\times 400$ for the first year cycle (August 1987–1988) and at $\times 200$ for the second (August 1988–1989), and correspond to 13.5 and 27 min of sampling every 2 h respectively. The method of counting pollen followed that of K  pyl   and Penttinen (1981). Hourly counts were stored in a database file for further analysis; statistics of the hourly counts are presented in Tables 1 and 2 of Bianchi et al. (1992). The concentration values correspond to total pollen grains. The main species found were: *Cupressus*, Gramineae, *Eucalyptus*, Pinace, Chenopodaceae, *Plantago*, Cyperaceae, *Betula*, Cruciferae, Compositae, Tueulflorae, *Ambrosia*, *Ulmus*, Umbelliferae, *Plantanus*, *Fraxinus*.

Neural networks

Robert Hecht-Neilson defines a neural network as “a dynamical system with the topology of a directed flow graph that can carry out information processing by means of its state response to continuous or episodic inputs” (Hecht-Neilson 1988). A neural network is fundamentally different from a traditional program in a computer. Neural networks have the following characteristics: they do not sequentially execute instructions; respond in parallel to a set of inputs; are more concerned with transformations than algorithms and procedures; are comprised of a large number of simple devices doing little more than computing weighted sums.

Information in neural networks is not stored in explicit memory locations, but is represented by the interconnection strengths and the current output of the neurons. An “artificial” neuron, referred to hereafter simply as a neuron, models the behaviour of the biological neuron composed of a body, one axon, and a multitude of den-

drites. The body of the j^{th} neuron is generally represented by a weighted linear sum, z_j , of the input signals followed by a linear or a nonlinear function $y_j = f(z_j)$. The function $f(z_j)$ is called an activation function, a function which uses the input value to determine the output activity of the neuron. Different neural networks use different activation functions, but the internal structure of the neuron, i.e. linear sum of the inputs followed by a function $f(z_j)$, is common to most networks. Examples of functions f are linear functions, threshold functions and non-linear Gaussian or sigmoid functions.

The “synapse” or connection between neurons is represented by the interconnection weight or simply weight, W_{ij} , from the i^{th} neuron to the j^{th} neuron. The weights can be positive (excitatory), negative (inhibitory) or zero. The number of neurons and its relative positions is the “architecture” of the neural network. The neural net architecture used in this work is shown in Fig. 1. For a given network architecture the weights and the functions f dictate the operation of the network; usually the functions f will be fixed. Thus varying the weights will allow the network to carry out different computations; the outputs of the neurons determine the current state of the network. Networks with adaptable weights use learning laws to adjust the values of the interconnection strengths.

Learning laws determine how the network is to adjust its weights using an error function or some other criteria. There are two types of learning: supervised and unsupervised. Supervised learning occurs when the network is supplied with both the input values and the correct output values, and the network adjusts its weights based upon the error of the computed output. Unsupervised learning occurs when the network is only provided with the input values, and the network adjusts the weights based solely on the input values and the current network output. Adaptable neural networks may have two phases of operation. The first phase is the training of the network. The user provides the network with a suitable number of input patterns, and output patterns if required, and the network adjusts the interconnection strengths until the output is “close enough” to the correct output. The second phase is the computation phase. The network is presented an input pattern and the network simply computes its output. More details concerning neural networks are given in Rumelhart et al. (1986a).

Neural nets for time series predictions

A typical architecture of a layered artificial neural net can be seen in Fig. 1. The input neurons are fed with the input patterns, while at the output neuron the response of the net is obtained. Placed between the input layer and the output layer, a layer of hidden neurons fully connected with both layers can be found. The number of neurons in the input layer is equal to the number of inputs in each input pattern. The output of the neurons in the output layer are the outputs of the network. The number of neurons in the hidden layer depends on the network designer: too many neurons

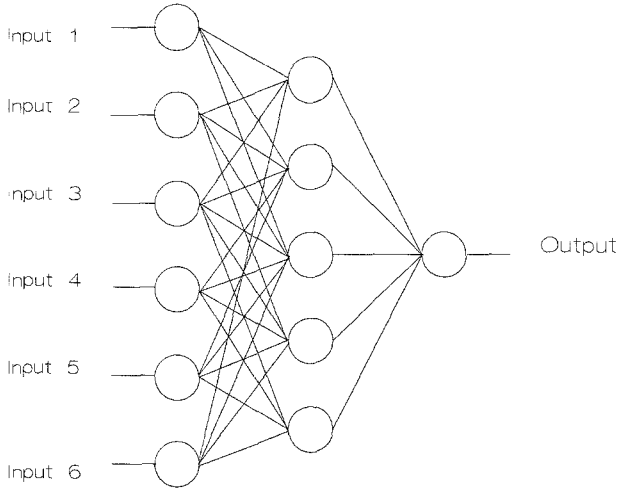


Fig. 1. Architecture of the three-layer neural network used to predict the pollen concentrations. The number of output nodes is set at 1 corresponding to the fact that we are making predictions of the subsequent 2-h concentration value. The number of input nodes is set at 6 and the number of hidden nodes is set at 5. Results from trial runs indicated that adding input and/or hidden nodes did not significantly improve the network prediction capabilities, but rather only slowed the convergence. The values at the input nodes are lagged values of the pollen time series

will make it difficult to deal with new input patterns, e.g. making generalizations. Too few neurons will not allow it to make sufficient internal representations so that accurate approximations for the maps from the input patterns to the output patterns can be made. There is no weight connecting input to output that skip the hidden layer.

The neurons have a nonlinear activation function. The output y_l of a given neuronal element l obey the following equation:

$$y_l = f(z_l), \quad (1)$$

where the argument z_l is either an external stimulus (if l is a neuron in the input layer) or is of the form

$$z_l = \sum_m W_{lm} z_m, \quad (2)$$

z_m being the output of the element m of the preceding layer and W_{lm} the weight connecting the l and m neurons. The transfer function $f(u)$ is chosen to be a typical sigmoid function:

$$f(u) = [1 + \exp(-u)]^{-1}. \quad (3)$$

In this way, the output values z_l of the neuronal elements range over the continuum $[0, 1]$. The output of the network is the z_l value of the output neuron and is determined by the input to the input neurons and the values of the different weights W_{lm} . To “train” the net consists in adjusting the couplings W_{lm} to achieve a prescribed input-output mapping. In a “supervised learning” scheme, the couplings are adjusted in order to minimize a “cost function”. Defining O_i as the net-

work output when the input I_i is present, and T_i as the desired output, the error associated with a possible cost function is:

$$C(W) = \sum_i (O_i - T_i)^2. \quad (4)$$

Currently, the most popular method for training this type of nets is “backpropagation” (BP; Rumelhart et al. 1986b). This process is an iterative one in which, at each step, the connection strengths are improved incrementally by altering them in a way which produces a decrease of $C(W)$ by the gradient descent method. The training process works as follows: error signals are calculated comparing the set of input-output pairs obtained by the net with the set of the desired input-output pairs. The weights are then incrementally altered using the error signals (batch learning). The equation for the increment of the weights is given as

$$W_{ij}^{n+1} = W_{ij}^n - \eta \delta C(W) / \delta W_{ij}, \quad (5)$$

where the sum is overall inputs. The parameter η is the learning rate and is set to a value that ensures convergence. The time spent in this process is an epoch, the basic unit of training time. A complete outline of the BP method can be found in Rumelhart et al. (1986b). If the input-output pairs used to train the net constitute a sequential time series, the essential idea is that the net will be able to approximate the relationships between subsequent values. Thus if new inputs, out of the training set, are presented, the output of the net will be close to corresponding to the “real” output. The net outputs can then be considered as “predictions” of future values of the time series.

Pollen time series: learning and predictions

The goal of this work was to build a neural network that can predict pollen concentrations in the immediate future. This objective was to be achieved by means of the backpropagation learning scheme explained above. The learning of any function [for example $y = \sin(x)$] is made by training the net with a certain quantity of pairs (x, y) or $[x, g(x)]$. The function $y = g(x)$ is called a “map”. The network will then be able to approximate this map, that is, if the network is fed with the input x , the output of the network will be approximately $y = g(x)$. The precision of the approximation will depend on the number of pairs (x, y) used and the final “cost function” obtained in the learning process, which will be proportional to the training time or number of epochs. The pollen time series may be considered in terms of a map: each member of the series, x_1 , is related to the next member, x_2 , by this map, i.e. $x_2 = g(x_1)$. The map or function g is, of course, unknown, but the neural network may be trained with the pairs (x_1, x_2) and after the training process the network will approximate this unknown map.

The map associated with the pollen time series is variable with time. The Lyapunov exponent (Guckenheimer

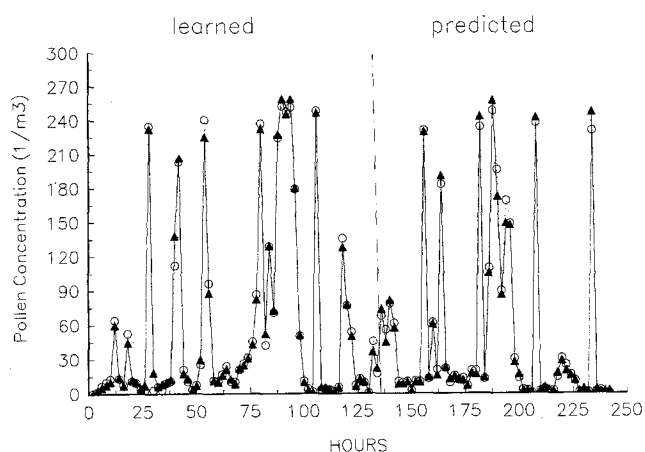


Fig. 2. Learned, predicted and measured values of the 2-h pollen concentration time series for September (beginning of spring in the southern hemisphere), characterised by its high pollen activity. The first 65 values of the 2-h pollen time series of September, used as the training set during 100000 epochs are shown in the *left* half of the figure. The agreement of the network outputs and the measured values is remarkable. The September predicted values of the 2-h time series for the week that follows the days taken for learning are shown at the *right* side. Due to the good agreement between the learned and the measured values achieved in the training of the net, an excellent correspondence between the measured and the predicted values can be obtained. ▲ Measured; ○ learned/predicted

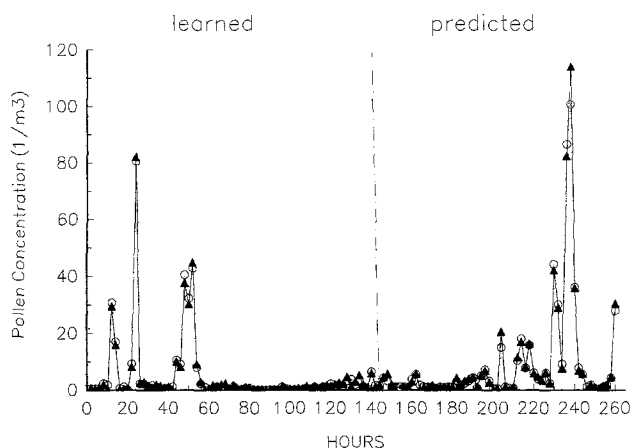


Fig. 3. Learned, predicted and measured values of the 2-h pollen concentration time series for June (winter). The *left side* presents the first 65 values of the 2-h pollen time series of June, used as the training set during 100000 epochs. The agreement of the network outputs and the measured values is remarkable. The June predicted values of the 2-h time series for the week that follows the days taken for learning are shown at the *right* side. Due to the good agreement between the learned and the measured values achieved in the training of the net, an excellent correspondence between the measured and the predicted values is obtained with a low pollen concentration as well as at the peak. ▲ Measured; ○ learned/predicted

and Holmes 1983) is characteristic of chaotic maps; it measures the exponential separation in time of the trajectories of two infinitesimally close initial points. We have estimated the Lyapunov exponent and observed that it changes from month to month. These studies of the pollen time series dynamics (to be published elsewhere) indicate that the dynamic map that produces this time series

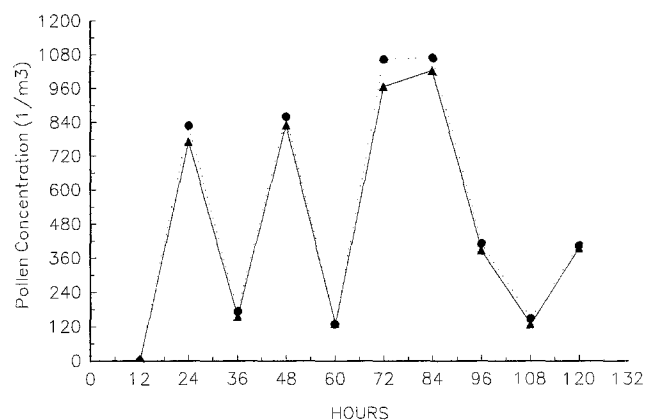


Fig. 4. Half day prediction: "Multistep" predicted values of pollen concentration over 5 days in September. Real measured values are also indicated for comparison. To obtain multistep prediction, one-step prediction is successively feedback to the neural network in such a way that a prediction is obtained half a day in advance. No additional learning is required. ▲ Predicted; ○ measured

has a characteristic time of around 1 month. Due to this fact the learned map with pollen values taken from the first days of a month will be a good approximation for that month but not for the next one. The studies performed also indicate that the Lyapunov exponent or degree of unpredictability is greater for September in accordance with the fact that this month is the beginning of spring in the southern hemisphere.

There are no strict rules for the choice of the neural net architecture, i.e. the number of neurons in each layer. The choice being a matter of trial and error. We found the best results with a neural net (Fig. 1) composed of 6 input neurons, 5 neurons in the hidden layer and 1 output neuron. The single output represents some future value of the series we wish to predict. The inputs are the components of a reconstructed n -dimensional space consisting of successive time-delayed values of the series. The method is similar to that used by Weigend et al. (1990) to predict annual sunspot numbers. For example, if we represent the series as $x(t_i)$ where $i=1, 2, \dots, L$, then with $\tau=1$, and using a 6-dimensional phase of space beginning with the first value of the part of the pollen time series we wish to predict, the first set of inputs is $x(1), x(2), \dots, x(6)$ and the output we are trying to predict is $x(7)$. Similarly, the second set of inputs is $x(2), x(3), \dots, x(7)$ and the output we are trying to predict is $x(8)$. Training continues over all training pairs (set of inputs and output) for several thousand iterations.

The first 65 values of the 2-h pollen time series of each month, scaled linearly to the $[0, 1]$ interval were used as the training set during 100000 epochs. The results obtained have been analysed and are shown in Figs. 2–7. The plots correspond to two quite different months: September (with the highest pollen activity) and June which can be considered as a typical winter month. The dynamic behaviour of the pollen time series obtained was similar during any month of the year except for September. In Fig. 2 the learned, predicted and measured values for September are shown. The September

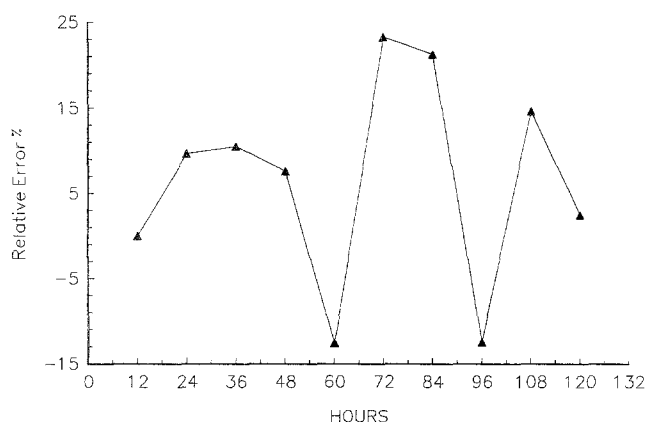


Fig. 5. Relative error for half-day prediction values over 5 days in September

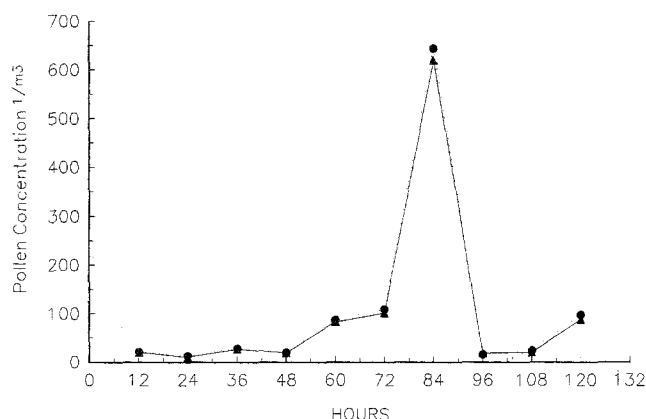


Fig. 6. Half-day prediction: Multistep predicted values of pollen concentration over 5 days in June. Real measured values are also indicated for comparison. Δ Predicted; $\bullet$ measured

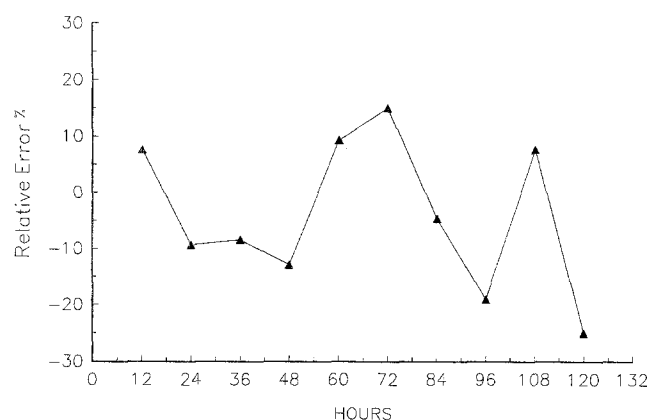


Fig. 7. Relative error for half-day prediction values over 5-days in June

predicted values of the 2-h time series for the week that follows the days taken for learning are shown in the right side of the same figure. Due to the good agreement between the learned and the measured values achieved in the training of the net, an excellent correspondence was obtained between the measured and the predicted values. The trained net was able to predict accurately the following 2-h value if the 6 measured values up to

the "present" moment were used as the net input. The correspondence between learned, predicted and measured values for the typical month June is shown in Fig. 3.

Although these are accurate predictions, a time interval 2 h in advance is not sufficient to prevent from high pollen concentrations. A reasonable period for advance prediction is half a day. A 12 h advance prediction may be obtained including in the set of inputs those predicted as if they were real inputs: The first input set is $x(1), \dots, x(6)$ where $x(i)$ are all inputs read from the pollen series. The output is $x'(7)$. The next input set will be $x(2), \dots, x(6), x'(7)$. Repeating several times, the sixth input set will be $x'(7), \dots, x'(12)$ where $x'(i)$ are predictions of the net. This kind of prediction with feedback of previous times of prediction is usually known as "multistep" prediction. Prediction over 5 days in September using the described method is shown in Fig. 4 with the corresponding measured values. The relative error for this kind of prediction is shown in Fig. 5. The errors are considered as the difference from the measured data. To obtain the error from the real concentration values, the measurement error must be added. The results for multistep prediction over 5 days in June are shown in Fig. 6 with the relative error shown in Fig. 7. The relative error of the 12-h predictions, although greater than that the 2-h predictions, is small and therefore the predictions obtained are a good forecast.

Conclusions

The ability of neural nets to predict the future behaviour of chaotic systems is used in order to make accurate predictions of the airborne pollen concentration, whose chaotic character was recently reported (Bianchi et al. 1992). We have shown that a neural net, trained with measured values of pollen concentration, is able to predict near future values. Since no environmental parameters are used in our approach it is not possible to predict the day of the year when the pollen flight starts. Thus the method may be used after the first appearance of the pollen flight. Very good forecasts for the subsequent 2 h and 12 h respectively were obtained. The operational method outlined here constitutes a different point of view compared to the traditional forecast using time series analysis, which requires input of many meteorological parameters. The present method does not require data other the pollen concentration values.

The utilization of this neural network method in practice is highly dependent on the possibility of calibration of the network, i.e. the availability of historical data and corrections on output data. Although the method is applied for total pollen concentration, it may be used with no modification for the pollen concentration of a particular species, if at least 1 month of the time series of interest is available.

Acknowledgement. The authors wish to acknowledge M.M. Bianchi for supplying the data of pollen concentrations used in this work. The software for IBM-PC and compatibles, developed for this method, is available from the authors.

References

- Bianchi MM, Arizmendi CM, Sanchez JR (1992) Detection of chaos: New approach to atmospheric pollen series analysis. *Int J Biometeorol* 36:172–175
- Comtois P, Sherknies D (1987) An aerobiological model for pollen forecasting. 18th Conference on Agricultural and Forest Meteorology – 8th Conference on Biometeorology and Aerobiology, W. Lafayette, Indiana. American Meteorological Society, Boston, Mass
- Comtois P, Batchelder G, Sherknies D (1989) Pre-season forecasting. In: Comtois P (ed) *Aerobiology health environment: a symposium*. University of Montreal, Montreal, Canada
- Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. *Phys Rev Lett* 59:845–848
- Goldberg C, Buch H, Moseholm L, Weeke E (1988) Airborne pollen records in Denmark, 1977–1986. *Grana* 27:209–218
- Guckenheimer J, Holmes P (1983) *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*. Springer, Berlin Heidelberg New York
- Hecht-Neilson R (1988) “Neurocomputer applications” in neural computers. In: Eckmiller, Rolf, von-der-Malsburg (eds) *Springer*, Berlin Heidelberg New York
- Lapedes A, Farber R (1987) Technical Report LA-UR-87, Los Alamos National Laboratory
- Käpylä M, Penttinen A (1981) An evaluation of microscopical counting methods of the tape in Hurst-Burkard pollen and spore trap. *Grana* 20:131–141
- Leuschner R (1991) Twenty-one years of airborne pollen determinations. *J Palyno* 27:305–320
- Moseholm L, Weeke E, Petersen B (1987) Forecast of pollen concentration of Poaceae (Grasses) in the air by time series analysis. *Pollen et Spores* 29:305–322
- O'Rourke M (1986) The implications of atmospheric pollen rain for pollen profiles in the arid southwest. PhD Thesis University of Arizona, Tucson, Arizona
- O'Rourke M (1989) Relationship between airborne pollen concentration and weather parameters in an arid environment. In: Comtois P (ed) *Aerobiology health environment: a symposium*. University of Montreal, Montreal, Canada
- Rumelhart DE, Hinton GE, Williams RJ (1986a) *Nature* 323:533
- Rumelhart DE, Hinton GE, Williams RJ (1986b) *Parallel distributed processing*. MIT Press, Cambridge, Mass
- Weigend AS, Huberman BA, Rumelhart DE (1990) Predicting the future: a connectionist approach. *Int J Neural Syst* 1:193