

Juan Antonio Sánchez Mesa · Carmen Galán ·  
César Hervás

## The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate

Received: 17 November 2004 / Revised: 9 February 2005 / Accepted: 22 February 2005 / Published online: 24 March 2005  
© ISB 2005

**Abstract** Biological particles in the air such as pollen grains can cause environmental problems in the allergic population. Medical studies report that a prior knowledge of pollen season severity can be useful in the management of pollen-related diseases. The aim of this work was to forecast the severity of the Poaceae pollen season by using weather parameters prior to the pollen season. To carry out the study a historical database of 21 years of pollen and meteorological data was used. First, the years were grouped into classes by using cluster analysis. As a result of the grouping, the 21 years were divided into 3 classes according to their potential allergenic load. Pre-season meteorological variables were used, as well as a series of characteristics related to the pollen season. When considering pre-season meteorological variables, winter variables were separated from early spring variables due to the nature of the Mediterranean climate. Second, a neural network model as well as a discriminant linear analysis were built to forecast Poaceae pollen season severity, according to the three classes previously defined. The neural network yielded better results than linear models. In conclusion, neural network models could have a high applicability in the area of prevention, as the allergenic potential of a year can be determined with a high degree of reliability, based on a series of meteorological values accumulated prior to the pollen season.

**Keywords** Aerobiology · Allergy · Neural network · Poaceae pollen

### Introduction

Biological particles originating from natural sources, particularly pollen grains, pose an environmental problem today in that they are a major source of adverse reactions among the population; for that reason, research into their precise role in allergies takes on considerable importance. Of all airborne pollen grains, those of the Poaceae family are considered one of the main causes of pollen allergy in Europe (Spiekma et al. 1989; Frenguelli et al. 1989; D'Amato et al. 1998; Norris-Hill and Emberlin 1991). In Córdoba, 73% of pollen-allergic patients are sensitive to this pollen type (Sánchez-Mesa et al. 2005).

Following production in the anthers, pollen grains undergo an aerobiological process involving their emission, dispersion and deposition. All of these mechanisms are greatly influenced by meteorological variables. The relationship between concentrations of grass pollen in the air and environmental factors has been analyzed by several authors [Galán et al. (1989), Emberlin (1994)]. A number of studies carried out over the last 20 years have demonstrated the relationship between weather parameters prior to the pollen season and the characteristics defining the main pollination period, such as the start of the pollen season and the peak day (Frenguelli et al. 1989; Laaidi 2001).

The Mediterranean climate is characterized by great variability. Normally, wet periods alternate with more frequent dry spells. Significant differences can even exist within a single year, giving rise, for example, to a very rainy winter and a very dry spring, or vice versa. The development of grasses and consequently the amount of pollen emitted into the atmosphere depends directly on rainfall levels. The second meteorological variable that most influences herbaceous plants is temperature.

The primary aim of this study was to group years according to the allergenic potential of airborne Poaceae pollen,

J. A. Sánchez Mesa (✉) · C. Galán  
Department of Plant Biology, University of Córdoba,  
Edificio Celestino Mutis, Campus de Rabanales,  
E-14071 Córdoba, Spain  
e-mail: bv2samej@uco.es  
Tel.: +34-957-218719  
Fax: +34-957-218598

C. Hervás  
Departments of Computing and Numerical Analysis, University  
of Córdoba,  
Edificio Celestino Mutis, Campus de Rabanales,  
E-14071 Córdoba, Spain

using weather and pollen data. This preliminary research makes it easier to forecast the severity of the Poaceae pollen season before it starts, by using pre-seasonal meteorological parameters. This could be of particular value for prevention purposes, since the allergenic potential of a given year could be determined with a high degree of reliability, based on a series of weather data accumulated prior to the pollen season. Early awareness of the severity of the pollen season could enable the use of a more appropriate model for predicting daily airborne grass pollen concentrations (Sánchez-Mesa et al. 2002). Discriminant linear analysis and neural network analysis were used to construct models.

Discriminant analysis is a very useful tool for detecting the variables that allow the researcher to discriminate between different (naturally occurring) groups, and for classifying cases into different groups with a better than chance accuracy.

A neural network consists of a mathematical model that performs a computational simulation of the behavior of neurons in the human brain by replicating, on a small scale, the brain's patterns, in order to form results from the events perceived, i.e. it is a model based on learning a set of training data. It is about being able to analyze and reproduce the learning mechanisms and recognition of events possessed by the more highly evolved animal species. In recent years, neural network use has spread to practically all sciences, particularly in pattern classification and non-linear models of prediction.

## Materials and methods

The study was carried out in the city of Cordoba (4° 45' W, 37° 50' N; height above sea level 123 m), in the south-west of the Iberian Peninsula. The climate is influenced by the Mediterranean Sea, the annual average temperature being 17.5°C and total annual rainfall 536 mm (according to the records of the National Institute of Meteorology for the last 30 years).

This study used a historical database of 21 years of pollen counts and weather data, from 1982 to 2002. Airborne pollen concentrations were measured using a Hirst-type volumetric sampler, located at a height of 15 m above ground level. Daily data obtained from the sampler were read following the method recommended by the Spanish Aerobiology Network (Domínguez et al. 1992). The result obtained was a daily average, expressed as pollen grains/m<sup>3</sup>. The beginning of the pollen season was considered when a mean value of at least 1 pollen grain/m<sup>3</sup>/day was detected, and at least 1 grain/m<sup>3</sup> on the following days, with no more than one consecutive day of 0 grains/m<sup>3</sup>. This technique eliminates the long tails at the beginning that lead to serious errors in statistical analyses. Meteorological data (daily rainfall and daily minimum temperature) were supplied by the National Institute of Meteorology, located 4 km from the sampling point.

## Grouping years

The primary aim was to group the 21 study years into classes according to their potential allergenic load, by analyzing conglomerates of mean values. *K*-means clustering assumes *n* individuals or objects and *p* measurements. We denote by  $X(I, j)$  the value of the *I*th individual on the *j*th variable;  $I=1,2, \dots, n, j=1,2, \dots, p$ . We will assume throughout the discussion that the measurements collected have properties that allow a Euclidean distance between individuals to be considered. Let  $P(n, K)$  be the partition that results in each of the *n* individuals being allocated to one of clusters 1, 2, . . . *K*. The mean of the *j*th variable in the *l*th cluster will be denoted by  $\bar{X}(l, j)$  and the number of individuals belonging to the *l*th cluster by  $n(l)$ . In this notation we may express the distance between the *l*th individual and the *l*th cluster as

$$D(i, l) = \left( \sum_{j=1}^p [X(i, j) - \bar{X}(l, j)]^2 \right)^{1/2}$$

We can define the error component of the partition by:

$$E[P(n, K)] = \sum_{i=1}^n D[i, l(i)]^2,$$

where  $l(i)$  is the cluster that contains the *i*th individual, and  $D[i, l(i)]$  is the Euclidean distance between individual *i* and the cluster mean of the cluster containing the individual. The procedure for clustering is as follows: search for a partition with a small error component *E* by moving individuals from one cluster to another, until no transfer of an individual results in a reduction in *E*.

Pre-season weather variables were used, together with a series of seasonal characteristics, following the methodology proposed in previous studies (Sánchez-Mesa et al. 2002). When considering pre-season meteorological variables, winter variables were separated from early spring variables, thus reflecting the nature of the Mediterranean climate, as explained in the introduction section.

The following pre-season variables were used:

- (1) RainDJ = 1 December to 31 January accumulated rainfall (mm).
- (2) DaysrainDJ = number of rainy days from 1 December to 31 January.
- (3) MintDJF = December, January and February accumulated minimum temperature (°C).
- (4) Winter = winter factor obtained from the following equation:

$$\text{Winter} = (\text{RainDJ} \times \text{DaysrainDJ}) / \text{MintDJF}$$

This factor was used to quantify the degree to which winter conditions were favorable for grass development;

a directly proportional relationship was considered between rainfall and grass development. Water, an integral part of living systems, is ecologically important because it is a major force in shaping climatic patterns and biochemically important because it is a necessary component in physiological processes (Brown 1995). Minimum temperature was considered to be inversely proportional, as grasses require cold during the winter in order to avoid premature growth. Over-development of grasses in winter could lead to plant death due to late frosts in winter. Frosts in Cordoba city usually occurs from December to February. For this reason, accumulated minimum temperature in February has been included in the MintDJF variable.

- (5) RainFMA = 1 February to 15 April accumulated rainfall (mm).
- (6) DaysrainFMA = number of rainy days from 1 February to 15 April.
- (7) MintMA = 1 March to 15 April accumulated minimum temperature ( $^{\circ}\text{C}$ ).
- (8) Spring = spring factor obtained from the following equation:

$$\text{Spring} = (\text{RainFMA} \times \text{DaysrainFMA} \text{MintMA}) / 100,000$$

This factor was used to quantify the degree to which early spring conditions were favorable for grass development; a directly proportional relationship was considered between rainfall and grass development. Grass development in Cordoba usually starts in March, but accumulated rainfall in February influences this phenomenon. For this reason, precipitation in February has been included in the RainFMA variable. Minimum temperature was considered to be directly proportional, as grasses require heat during this part of the year for normal vegetative development. The total result was divided by 100,000 in order to correct the scale.

The following seasonal variables were used:

- (1) TotPollen = whole-season total pollen ( $\text{grains}/\text{m}^3$ ).
- (2) Maximum = pollen concentration of the highest peak day ( $\text{grains}/\text{m}^3/\text{day}$ ).
- (3) Daysbefpeak = number of days from the start of the season to the peak day.

## Classification

The second aim was to classify years according to the classes determined before, using only the pre-seasonal variables defined previously. By this means, it was possible to categorize a year even before the start of the pollen season, using only weather-related parameters. Both discriminant linear analysis and neural network models were used for this purpose. Analysis covered the first 15 years (from 1982 to 1996), while the last 6 years (from 1997 to 2002) were used for validation.

## Discriminant linear analysis

Discriminant analysis allows us to build a predictive model of group membership based on observed characteristics of each case. The variables tested in this study were those previously described in the “grouping years” section. The procedure generates a discriminant function (or, for more than two groups, a set of discriminant functions) based on linear combinations of the predictor variables that provide the best discrimination between the groups. The functions are generated from a sample of cases for which group membership is known; the functions can then be applied to new cases with measurements for the predictor variables but unknown group membership. Determining the number of statistically significant discriminant functions is particularly important. The number of discriminant functions that provide statistically significant among-groups variation essentially defines the dimensionality of the discriminant space. In this study, a Wilks’ Lambda test was used to determine the significance of each discriminant function. The Wilks’ Lambda test also measures the differences between groups. Thus, multiple discriminant analysis can be viewed as a reduction technique, since it can, by uncovering a small number of discriminant functions, provide a condensed version of the factors that contribute to the among-group differences.

## Neural network models

The main characteristic of neural networks is their capacity for learning by example. This means that by using a neural network there is no need to program how the output is obtained, given certain input; but rather examples are shown of the relationship between input and output, and the neural network will learn the existing relationship between the two by means of a learning algorithm. This learning will materialize in the network’s topology and in the value of its connections. Once the neural network has “learnt” to carry out the desired function, it can be used, i.e. input values for which the output is unknown can be entered, and the neural network will calculate the output.

The ANN model was inspired by what is known about the physical structure and mechanism of the nervous system and the biological cognition and learning process, on an oversimplified scale. Although based on functionality of the nervous system, the cornerstone of ANN is in fact its structure. The basic structure of the nervous system is the neuron. Basically, a neuron consists of three major functional elements: dendrites, cell body, and axon (Fig. 1).

Dendrites receive signals from other neurons and send them to the cell body. The axon receives signals from the cell body and carries them away through the synapse to the dendrites of neighboring neurons. The dendrites of the second neuron generate a new electric signal depending upon the intensity of the signal received, the synaptic strengths and the threshold of the receiving neuron. Since a neuron has a large number of dendrites it can receive many signals simultaneously, and all neurons thus form a network (Fig. 2).

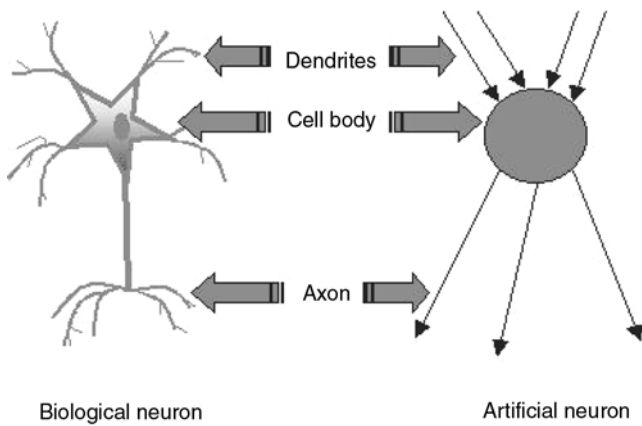


Fig. 1 Analogy between artificial neuron and biological neuron

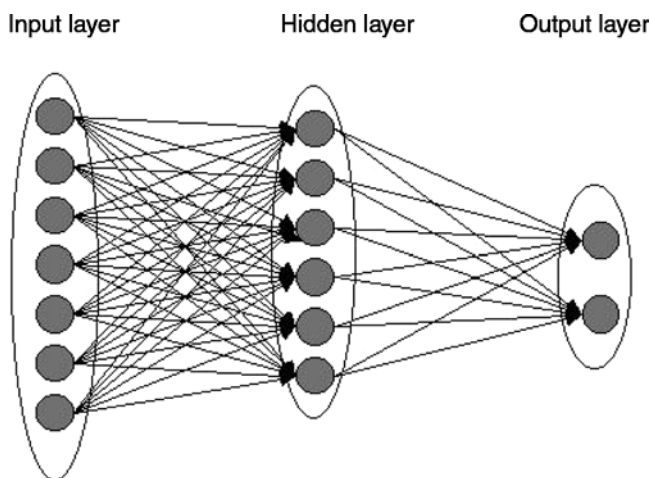


Fig. 2 Artificial Neural Network diagram showing the connections between neurons

The analogy between the artificial neuron and the biological neuron is that the connections between nodes represent the axon and dendrites, the connection weights represent the synapses, and the threshold approximates the activity in the soma.

Classification often uses a major class of artificial neural networks, mainly multilayer feedforward networks. Usually, the network consists of a set of sensory units (source nodes), each node being associated with a variable or characteristic of the problem (in our case RainDJ, DaysrainDJ, MintDJF, etc.), that constitutes the *input layer*, one or more *hidden layers* of computation nodes, and an *output layer* of computation nodes, each of which is associated with a class identification; in our case, class 1 (1, 0, 0), class 2 (0, 1, 0) or class 3 (0, 0, 1). The input signal propagates through the network in a forward direction, on a layer-by-layer basis.

The nomenclature of the network architecture is as follows:

(no. input neurons):(no. hidden neurons) s:(no. output neurons) s, (output)  
in our case 8: 2s:3s, (output)

where:

- $s$ , the transfer function for the neurons in the hidden and the output layer is a sigmoid;
- “output” in the example is the class identification estimated;
- for classification, the transfer function in the output layer is a sigmoid, but for prediction the transfer function is linear.
- The design of the layers includes an intercept term called “bias”.

Among the different algorithms that have been used to train the network is the traditional back propagation algorithm (BP) first proposed by Werbos (1974) and later used by Rumelhart and McLelland (1986). This algorithm is a generalization of the Delta rule (Widrow et al. 1988) and, like it, undergoes a slow but sure learning process. Another algorithm employed is the Extended-Delta-Bar Delta (EDBD) algorithm, later used by Williams and Minai (1990), which is at the same time a modification of the Delta-Bar-Delta algorithm proposed by Jacobs (1978) to improve the learning speed.

One of the main problems involved in the application of artificial neural networks is the selection of the most appropriate net architecture to be used. That is, the correct number of nodes in the hidden layers of the net must be determined as well as the amount of connections between the nodes in the different layers making up the net. Generally, the size of the net affects its complexity and the time necessary for training but, even more importantly, it affects its capacity to generalize (that is, its capacity to produce reliable and satisfactory results for data different to those used during training). In fact, the same learning error can be obtained in nets with different structures, although the error of generalization would probably be different. In practice, it seems that a bigger net size contributes to a lower error level in the training group, although at the same time it could increase the error committed in the generalization group. Because of this, there ought to be an analysis of the degree of complexity that a neuronal net should have to resolve each problem in order for the learning and generalization errors to be considered acceptable.

To find the number of nodes to be included in this layer, as well as the number of connections between the nodes in the different layers that make up the net, the evolutive algorithm proposed by Bebis et al. (1997) was used, coupled with Williams’ pruning algorithm (Williams 1994). Evolutive algorithms in the field of neuronal networks have proved in many studies to be useful in the optimization of the architecture of the net and its weights (Whitley and Schaffer 1992).

Evolutive algorithms are stochastic search algorithms that execute a global search in weight space, avoiding the fall to local optimum often produced by overtraining of the net (Angeline et al. 1994; Yao 1999; García et al. 2003). Here, we used evolutive algorithms coupled with pruning algorithms (Hervás et al. 2000) to search for the neuronal net architecture that will allow us to verify, with a minimum of information, the classification of years.

**Table 1** Mean variable values for each class considered in this study

Variable	Class 1	Class 2	Class 3
RainDJ(mm)	91.06	194.52	262.13
DaysrainDJ*	12	20	26
MintDJF(°C)*	310.65	471.58	533.37
Winter	3.70	9.04	13.53
RainFMA(mm)	94.74	125.26	185.25
DaysrainFMA*	12	16	24
MintMA(°C)	333.09	337.33	395.50
Spring	4.36	7.93	17.99
TotPollen (grains/m <sup>3</sup> )*	1871	4033	7681
Maximum (grains/m <sup>3</sup> /day)*	159	347	811
Daysbefpeak*	39	63	73

\*Significance level of 0.05 according to ANOVA test

## Results

### Grouping years

As a result of the classification, the 21 study years were divided into three classes. All the pre-season and season variables described above were introduced into the K-means cluster analysis. The mean values of these variables for each class are shown in Table 1. The ANOVA test revealed significant differences between mean values for each class, with a significance level of 0.05, for the following variables: DaysrainDJ, MintDJF, DaysrainFMA, TotPollen, Maximum and Daysbefpeak. Therefore, the value of these variables was enough to match a year with a certain class.

The main meteorological characteristics of the three classes were as follows:

**Class 1.** Comprised 8 years: 1982, 1983, 1987, 1992, 1993, 1994, 1995 and 1999. Weather conditions prior to the pollen season of these were highly unfavorable for the development of annual grasses. Winter rainfall was low, with a mean of 12 rainy days in December and January. The winter was cold, with mean accumulated minimum temperature for the months of December, January and February of 310.65°C. Early spring was also quite dry, with a mean of 12 rainy days from 1 February to 15 April. Late frosts, which irretrievably damage the plant, often occur in spring.

**Class 2.** Comprised 9 years: 1984, 1985, 1988, 1989, 1990, 1991, 1997, 1998 and 2000. This class includes years with typically Mediterranean weather, i.e. very rainy winter and a dry early spring, or years with a dry winter and a very rainy early spring. During the winter it rained a mean number of 20 days during the months of December and January. Mean accumulated minimum temperature for the months of December, January and February was 471.58°C. From 1 February to 15 April, it rained a mean number of 16 days.

**Class 3.** Comprised 4 years: 1986, 1996, 2001 and 2002. This class was characterized by frequent periods of rainfall in both winter and early spring. In winter it rained a mean number of 26 days during the months of December and January. The mean accumulated minimum temperature in

December, January and February was 533.37°C. This high value is due to these being rainy years, making for a milder winter. Early spring was also rainy, with a mean number of 24 rainy days from 1 February to 15 April.

In spite of considering the cold required by grasses in winter, the results suggest the opposite. In fact, class 1, which has the coldest temperatures, is the least favorable to the development of grasses; class 3, which has the warmest temperatures, is the most favorable. These results seem to indicate that rainfall is far more important than temperature. Furthermore, in the Mediterranean climate there is an association between rainfall and temperature. Rainy years are usually characterized by being warm and by an absence of frosts. In contrast, dry years tend to be cold. This apparent contradiction between cold requirement and grass development can also be explained by the plants themselves which have very different pollination periods from one species to another and so possibly different cold requirements.

Pollen-season severity for each class is shown in Fig. 3. In class 1, years were characterized by very low airborne grass pollen concentrations. People sensitive to grass pollen, probably had very few problems in these years. For people sensitive to this pollen type, class 2 years were moderate with several days of high concentrations. During class 3 years, gramineous plants probably produced the greatest number of allergy episodes among the population, related to the large amount of pollen collected during these years.

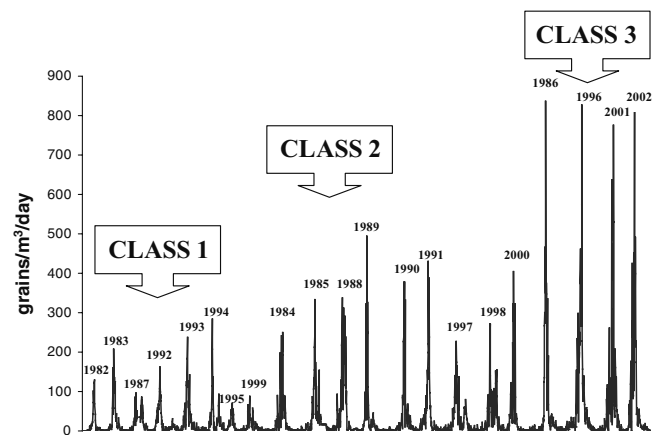
### Discriminant linear analysis

Discriminant formulas obtained were as follows:

$$D_1 = -1.699 \text{ RainFMA} + 2.045 \text{ DaysrainFMA}$$

$$D_2 = 1.151 \text{ RainFMA} - 0.177 \text{ DaysrainFMA}$$

The first function maximizes the differences between the values of the dependent variable. The second function is orthogonal to it (uncorrelated with it) and maximizes



**Fig. 3** The severity of Poaceae pollen seasons for the three classes included in this study

the differences between values of the dependent variable, controlling for the first factor. Though mathematically different, each discriminant function is a dimension which differentiates a case into categories of the dependent (class 1, class 2 or class 3) based on its values on the independents. The first function is the most powerful differentiating dimension, but the second function may also represent additional significant dimensions of differentiation.

Neural network analysis

Neural network models obtained are shown in Fig. 4. For each year we obtained three values as a result of the replacement of variables in the three formulas shown in Fig. 4 (class 1, class 2 and class 3). These three values are included in an interval between 0 and 1. The value closer to 1 determined the expected class according to neural net-

work models. From the formulas in Fig. 4, the following important rules could be deduced:

Rule 1. Wet winter and wet early spring: a wet winter determines an increase in temperature, so MintDJF rises. A wet early spring determines an increase of DaysrainFMA. These two variables are the most important in the  $h_1$  formula.

$$\text{MintDJF and DaysrainFMA } \uparrow \rightarrow h_1 \downarrow \rightarrow C_1 \downarrow$$

$$\text{and } C_3 \uparrow \rightarrow \text{class 3 year}$$

Rule 2. Dry winter and dry early spring: a dry winter determines a decrease in temperature, so MintDJF falls. A dry early spring determines a decrease in DaysrainFMA.

$$\text{MintDJF and DaysrainFMA } \downarrow \rightarrow h_1 \uparrow \rightarrow C_1 \uparrow$$

$$\text{and } C_3 \downarrow \rightarrow \text{class 1 year}$$

Class 2 could be considered as an intermediate class with weather conditions different from those of class 3 and class 1, i.e. wet winter and dry early spring or vice versa. Years classified by the neural network as class 2 are those which the model could not place in class 1 or class 3.

Validation

Table 2 shows the accuracy of analysis each year in both training and validation phases. Percentage accuracy is shown in Table 3. Linear discriminant analysis and neural network models obtained the same results in the training phase. However, in the validation phase, neural network models (50% accuracy) performed better than linear discriminant analysis (16.70%).

Discussion

The meteorological variables used to group the 21 study years were rainfall and minimum temperature. With regard to rainfall, it was observed that the number of rainy days was even more important than the amount collected. This is due to the fact that in Mediterranean countries rain tends to be torrential and almost all is lost as run-off. For herbaceous plants, gradual rainfall is better assimilated. As for temperature, studies have shown that minimum temperature is the most influential factor in the city of Córdoba, whereas in other cities with a different climate, such as London, the most important factor is maximum temperature (Galán et al. 1995).

Taking into account the classification models, discriminant linear equations included only rainfall-related variables from early spring (RainFMA and DaysrainFMA). However, neural network models included two winter variables (DaysrainDJ and MintDJF) as well as one variable from early spring (DaysrainFMA). Previous studies have attempted to predict pollen-season severity using

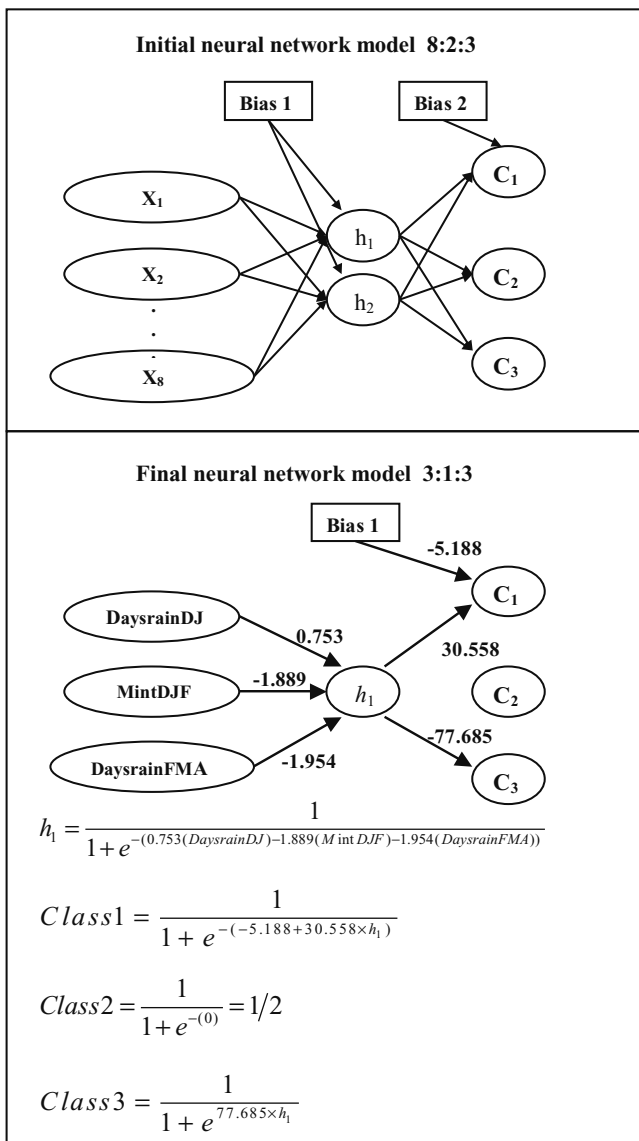


Fig. 4 Neural network model used to classify years

**Table 2** Observed class according to K-mean analysis and expected class taking into account both linear and neural network models

	Year	Observed class (K-mean analysis)	Expected class according to linear models	Expected class according to neural network models
Training years	1982	1	1	1
	1983	1	1	1
	1984	2	2	2
	1985	2	2	2
	1986	3	3	3
	1987	1	1	1
	1988	2	2	2
	1989	2	2	2
	1990	2	1	1
	1991	2	2	2
	1992	1	1	1
	1993	1	1	1
	1994	1	1	1
	1995	1	1	1
Validation years	1996	3	3	3
	1997	2	1	1
	1998	2	2	2
	1999	1	2	1
	2000	2	1	1
	2001	3	2	3
	2002	3	2	2

**Table 3** Percentage of accuracy (%) in both linear and neural network models taking into account training and validation phases

	Linear discriminant models	Neural network models
Training phase	93.34	93.34
Validation phase	16.70	50.00

meteorological variables. In the case of Cupressaceae, cumulative minimum temperature and rainfall in winter seem to be the most influential factors in the southern Iberian Peninsula (Galán et al. 1998). In other pollen types, such as *Olea*, it has been demonstrated that rainfall can affect the intensity of the pollen season (González-Minero and Candau 1996).

Neural network models are able to relate different kinds of variables by following a non-linear path, yielding better results. Previous studies have tackled this subject in aerobiology, producing neural network models to predict pollen data (Arizmendi et al. 1993; Ranzi et al. 2003). In general, a neural network produces complicated equations that are difficult to understand. In this case, neural network formulas are not particularly complex. In fact, as the results show, neural network models can be readily interpreted even when weather-related variables are involved.

Discriminant linear analysis and neural networks showed similar results in the training phase. However, taking into account the validation phase, better results were

obtained with neural networks than with discriminant linear analysis, though they are more complex. Neural network models performed better than linear models when predicting daily airborne grass-pollen concentrations in some studies carried out previously (Sánchez-Mesa et al. 2002; Hidalgo et al. 2002).

To predict the class of a certain year, prior to the onset of the pollen season with over 50% accuracy represents a major challenge. Although conditions prior to the pollen season could be typical of a certain class, weather patterns during the pollen season itself could modify this prediction. For example, in some years weather variables favor grass development, leading to the inclusion of such years in class 3. However, weather patterns during the pollen season might be unfavorable for pollen release and transport (i.e. a rainy pollen season). Thus, the real category could change from class 3 to class 2 or even 1.

As it has been reported in previous studies (Sánchez-Mesa et al. 2002), the classification into three classes does not improve the prediction of pollen concentrations much but the process is considerably simplified. In fact, modelling can be done without the need to train all the years and finally the prediction of pollen concentrations can be made with great accuracy (90% by using neural network and 80% with linear regression).

**Acknowledgements** The authors wish to thank the Fondo de Investigación Sanitaria (FIS) for the BEFI Grant Ref: 00/9406 and the "Comisión Interministerial de Ciencia y Tecnología" TIC2002-04036-C05-02 (Department of Computer Science, University of Córdoba). Additional financial support from FEDER funds was also received

## References

- Angeline PJ, Saunders GM, Pollack JB (1994) An evolutionary algorithm that constructs recurrent neural networks. *IEEE Trans Neural Networks* 1:54–65
- Arizmendi CM, Sánchez JR, Ramos NE, Ramos GJ (1993) Time series predictions with neural nets: applications to airborne pollen forecasting. *Int J Biometeorol* 37:139–144
- Bebis G, Georgiopoulos M, Kasparis T (1997) Coupling weight elimination with genetic algorithms to reduce network size and preserve generalization. *Neurocomputing* 17:167–194
- Brown RW (1995) The water relations of range plants: adaptations of water deficits. In: Bedunah DJ, Sosebee RE (eds). *Wildland plants: physiological ecology and developmental morphology*. Society for Range Management, Denver, Colo., pp 291–413
- D'Amato G, Spiekma F Th M, Liccardi G, Jäger S, Russo M, Kontou-Fili K, Nikkels H, Wüthrich B, Bonini S (1998) Pollen-related allergy in Europe. *Allergy* 53:567–578
- Domínguez E, Galán C, Villamandos F, Infante F (1992) Manejo y Evaluación de los datos obtenidos en los muestreos aerobiológicos. *Monografías REA/EAN* 1:1–13
- Emberlin J (1994) The effects of patterns in climate and pollen abundance on allergy. *Allergy* 49:15–20
- Frenguelli G, Bricchi E, Romano B, Mincigrucci G, Spiekma F Th M (1989) A predictive study on the beginning of the pollen season for Gramineae and *Olea europea* L. *Aerobiologia* 5:64–70
- Galán C, Cuevas J, Infante F, Domínguez E (1989) Seasonal and diurnal variation of pollen from Gramineae in the atmosphere of Córdoba (Spain). *Allergol immunopathol* 17:245–249

- Galán C, Emberlin J, Domínguez E, Bryant R, Villamandos F (1995) A comparative análisis of daily variations in the Gramineae pollen counts at Córdoba, Spain and London, UK. *Ghana* 34:189–198
- Galán C, Fuillerat MJ, Comtois P, Domínguez E (1998) A predictive study of Cupressaceae pollen season onset, severity, maximum value and maximum value data. *Aerobiologia* 14:195–199
- García N, Hervás C, Muñoz J, Covnet A (2003) A cooperative evolutionary model for evolving artificial neural networks". *IEEE Trans Neural Networks* 14:575–596
- González-Minero FJ, Candau P (1996) Prediction of the beginning of the olive full pollen season in South-West Spain. *Aerobiologia* 12:91–96
- Hervás C, Algar JA, Silva M (2000) Correction of temperature variations in kinetic-based determinations by use of pruning computational neural networks in conjunction with genetic algorithms. *J Chem Inf Comput Sci* 40:724–731
- Hidalgo PJ, Mangin A, Galán C, Hembise O, Vázquez LM, Sánchez O (2002) An automated system for surveying and forecasting *Olea* pollen dispersion. *Aerobiologia* 18:23–31
- Jacobs RA (1978) Increased rates of convergence through learning rate adaptation. *Neural Networks* 1:295–307
- Laaidi M (2001) Forecasting the start of the pollen season of Poaceae: evaluation of some methods based on meteorological factors. *Int J Biometeorol* 45:1–7
- Norris-Hill J, Emberlin J (1991) Diurnal variation of pollen concentration in the air of north-central London. *Grana* 30:229–234
- Ranzi A, Lauriola P, Marletto V, Zinoni F (2003) Forecasting airborne pollen concentrations: development of local models. *Aerobiologia* 19:39–45
- Rumelhart DE, McLelland JC (1986) *Parallel distributed processing: explorations in the microstructure of cognition*. MIT, Boston
- Sánchez Mesa JA, Galán C, Martínez JA, Hervás C (2002) The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. *Clin Exp Allergy* 32:1606–1612
- Sánchez Mesa JA, Serrano P, Cariñanos P, Moreno C, Guerra F, Galán C (in press) Pollen allergy in Cordoba city: frequency of sensitization and relation with antihistamines sales. *J Invest Allergol Clin Immunol*
- Spiekma F Th M, D'amato G, Mullins J, Nolard N, Wachter R, Weeke ER (1989) City Spore Concentrations in the European Economic Community (EEC). VI. Poaceae (Grasses), 1982–1986. *Aerobiologia* 5:38–43
- Werbos PJ (1974) *Beyond regression: New tools for prediction and analysis in the behavioural sciences*. Ph D. thesis, University of Harvard
- Whitley LD, Schaffer JD (1992) *International workshop on combinations of genetic algorithms and neural networks*. IEEE Computer Society, Vancouver, Canada
- Widrow B, Winter RG, Baxter RA (1988) Layered neural nets for pattern recognition. *Trans Acoust Speech Signal Process* 36:1109–1118
- Williams PM (1994) Bayesian regularization and pruning using a Laplace prior. *Neural Computation* 7:117–143
- Williams RJ, Minai AA (1990) Back-propagation heuristics: A study of the Extended Delta-Bar-Delta algorithms. *IEEE Intl Joint Conf Neural Networks* 12:595–600
- Yao X (1999) Evolving artificial neural networks. *Proc IEEE* 9:1423–1447