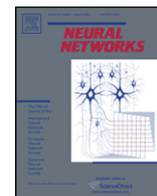




ELSEVIER

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Evaluation of atmospheric Poaceae pollen concentration using a neural network applied to a coastal Atlantic climate region

F.J. Rodríguez-Rajo^{a,*}, G. Astray^b, J.A. Ferreiro-Lage^b, M.J. Aira^c, M.V. Jato-Rodríguez^a, J.C. Mejuto^b

^a Department of Plant Biology and Soil Sciences, Faculty of Sciences, University of Vigo, 32004, Ourense, Spain

^b Department of Physical Chemistry, Faculty of Sciences, University of Vigo, 32004, Ourense, Spain

^c Department of Botany, Faculty of Pharmacy, University of Santiago of Compostela, 15782, Santiago of Compostela, Spain

ARTICLE INFO

Article history:

Received 29 January 2009

Received in revised form 27 April 2009

Accepted 20 June 2009

Keywords:

Neural networks

Pollen

Allergy

Aerobiology

Poaceae

Anova

ABSTRACT

In the South of Europe an important percentage of population suffers pollen allergies, being the Poaceae pollen the major source. One of aerobiology's objectives is to develop statistical models enabling the short- and long-term prediction of atmospheric pollen concentrations to take preventative measures to protect allergic patients from the severity of the atmospheric pollen season. The implementation of a computational model based on supervised MLP neural network was applied for the prediction of the atmospheric Poaceae pollen concentration. There is a good correlation between the values predicted by the ANN for the training cases in comparison with the real pollen concentrations. A high coefficient of linear regression (R^2) of 0.9696 was obtained. The accuracy of the neural network developed was tested with data from 2006 and 2007, which was not taken into account to establish the aforementioned models. Neural networks provided us a good tool to forecasting allergenic airborne pollen concentration helping the automation of the prediction system in the aerobiological information diffusion to the population suffering from allergic problems.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The Poaceae family is the most diverse and prolific family of the herbaceous plants in urban areas. The most abundant genera causing pollinosis in Galicia are *Dactylis*, *Avena*, *Arrhenatherum*, *Vulpia* and *Bromus*. It is well known that the prevalence of allergic diseases has increased in recent decades in the industrialised world. Poaceae pollen is considered as one of the most important aeroallergen, and leading to allergic symptoms in more of 80% of the people sensitive to some pollinic type in Europe (D'Amato, Cecchi, Bonini, Nunes, Annesi-Maesano & Behrendt, 2007). In Spain, 22% of the population suffers pollen allergies, being the Poaceae pollen the major source. Positive skin test values range from 97% in northern Spain to 48% in coastal south-eastern Spain; an average of 80% of the allergy sufferers are affected by Poaceae pollen (Subiza, 2003). Its pollen is a factor of risk for patients with respiratory diseases and the most important pollen responsible for the pollinosis in Galicia (Arenas, González, Tabarés, Iglesias, Méndez & Jato, 1996).

Recent studies report that pollen allergies are found in 80%–90% of childhood asthmatics and 40%–50% of adult onset asthmatics

(Taylor, Jacobson, House, & Glovsky, 2007). Grass pollens have been credited with an etiological role in localised asthma epidemics associated with weather perturbations, particularly thunderstorms and rainfall (Green, Dettmann, Yli-Panula, Rutherford, & Simpson, 2004). Different thresholds are indicated in the bibliography as triggering allergy symptoms: pollen counts of over 50 grains/m³ are reported to cause clinical symptoms in all allergy sufferers (Negrini, Voltolini, & Arobba, 1992), while levels of less than 30 grains/m³ are considered sufficient to produce nasal hypersensitivity (Rantio-Lehtimäki, Koivikko, Kupias, Makinen, & Pohjola, 1991).

Phenological fluctuations in Poaceae are, in general, minimum, but climatic conditions can modify their behaviour. Due to its superficial radical system, pasture ecosystems are especially sensitive to changes in temperature, rainfall (French & Sauer, 1974) and water availability (Dickinson & Dodd, 1976).

One of aerobiology's objectives is to develop statistical models enabling the short- and long-term prediction of atmospheric pollen concentrations. This enables allergic individuals to take preventative measures to protect themselves from the severity of the pollen season. Pollen concentration predictions are usually made long term, trying to predict the onset and severity of the pollen season. In recent literature such models use, as prediction values, the sum of temperatures starting from a certain date (Caramiello, Siniscalco, Mercalli, & Potenza, 1994; Clot, 2001; García-Mozo, Chuine, Aira, Belmonte, Bermejo &

* Corresponding author. Tel.: +34 988387193; fax: +34 988387001.

E-mail address: javirajo@uvigo.es (F.J. Rodríguez-Rajo).

de la Guardia, 2008; Rodríguez-Rajo, Valencia-Barrera, Vega-Maray, Suarez, Fernandez-Gonzalez & Jato, 2006) or different phenological variables such as the chilling units and growing degree days required to trigger the phenological phase of flowering (Andersen, 1991; Jato, Rodríguez-Rajo, Méndez, & Aira, 2002; Laaidi, 2001). Another type of model aims at predicting short-term concentrations, e.g. one or two days in advance, making use of statistical tools such as time series analysis (Cotos-Yañez, Rodríguez-Rajo, & Jato, 2004) or other models based on the correlation of pollen levels with meteorological variables by means of uni-dimensional nonlinear equations using variables with the greatest prediction capacity (Norris-Hill, 1995). In the case of Poaceae, the most frequent models are based on linear parametric statistics (Chuine & Belmonte, 2004; Laaidi, 2001; Sánchez-Mesa, Galán, & Hervás, 2005).

The aim of this research is the short-term prediction of Poaceae pollen concentrations, by using artificial neural networks, in order to alert allergists and population with allergic problems to a potential risk situation. Artificial Neural Networks (ANNs) are a complete statistical tool for data analysis (Bishop, 1995). The ANN origin dates back to the middle of the last century when an interdisciplinary group of biologists, psychologists and engineers with interest in understanding the functioning of the human brain was created (Rosenblatt, 1958). The ANNs try to reproduce artificially the human ability of taking decisions simulating the human brain's basic unit, the neuron, and the interconnections between the neurons that allow them to work together and save experience's information (Haykin, 1994; Xu, Xie, Tang, & Ho, 2003).

In recent years, ANNs have been extended successfully to very different fields, from hydrology to finances (Castillo, Gutiérrez, Hadi, & Lacruz, 2001). Neural networks have also been used in aerobiological studies in order to achieve predicting models to improve daily pollen concentration forecast (Arizmendi, Sánchez, Ramos, & Ramos, 1993; Castellano-Méndez, Aira, Iglesias, Jato, & González-Manteiga, 2005; Grinn-Gofron & Strzelczak, 2008; Sánchez-Mesa et al., 2005; Sánchez-Mesa, Galán, Martínez, & Hervás, 2002).

2. Material and methods

2.1. Study area

The Spanish city of Vigo with 375 000 inhabitants is located in the northwestern corner of the Iberian Peninsula (Fig. 1). In this area, the annual mean temperature is 13.6 °C, mean maximum temperature is 17.7 °C and mean minimum temperature is 9.6 °C. The annual rainfall is 1909 mm, distributed irregularly throughout the year, with only 43 and 40 mm of rain collected in July and August, respectively. Biogeographically, Vigo is situated in the Atlantic Province of the Euro-Siberian Region. The main crops are vines and forage species. The Poaceae genera more common in the Eurosiberian Region, in which the area of study is located, are *Dactylis*, *Alopecurus*, *Holcus*, *Elymus*, *Poa* and *Phleum* in the humid zones and *Hordeum*, *Avena*, *Secale*, *Lolium*, *Poa*, *Agrostis*, ... in nitrified soils.

2.2. Aerobiological survey

A Hirst-type LANZONI VPPS 2000 volumetric 7-day recording sampler (Hirst, 1952) was used to collect the airborne pollen from 1995 to 2007. The sampler was situated on the left bank of the Vigo estuary (42°14'15" N; 8°43'30" W), in the city centre on the roof of the Town Hall, at a height of 15 m above ground level. The Lanzoni sampler is calibrated to handle a flow of 10 l of air per minute, thus matching the human breathing rate. Pollen grains are impacted on a cylindrical drum covered by a melinex

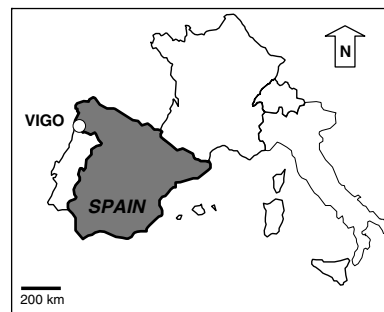


Fig. 1. Location of Vigo in Europe.

film coated with a 2% silicon solution as trapping surface. The drum was changed weekly and the exposed tape was cut into seven pieces, which were mounted on separate glass slides. Daily values were expressed as number of pollen grains per cubic metre of air. Pollen grain identification was performed using a NIKON OPTIPHOT II microscope equipped with a 40X/0.95 lens. Pollen counts were made using the model proposed by the R.E.A. (Spanish Aerobiological Network), consisting in four continuous longitudinal traverses along the 24 h slide (Galan, Cariñanos, Alcazar, & Domínguez, 2007). The following atmospheric pollen season characteristics were analyzed: (1) first and last dates; (2) maximum pollen count and date recorded; and (3) number of days on which pollen concentrations reached values representing allergenic risk. The pollination period was calculated as 95% of the annual total pollen, the first days, producing up to 2.5% of total production, were removed from the calculations as well as the final days (from 97.5% to 100.0% of total production) (Andersen, 1991).

2.3. Meteorological survey

The meteorological data used for the study were recorded by the Spanish Meteorological Agency. The meteorological station is placed next to the Hirst-type LANZONI VPPS 2000 volumetric sampler. This station offers the typical available parameters used for the meteorological forecast: rainfall (mm), humidity (%), maximum temperature (°C), minimum temperature (°C), mean temperature (°C), solar radiation (w/m²), wind direction (h), wind calms (h) and wind speed (m/s).

2.4. Statistical analysis

For the prediction of the atmospheric *Poaceae* pollen concentration in the city of Vigo, it has been chosen the implementation of a computational model based on neural network technology. This computer model has the capacity of "learning" the existing relation among the different input variables or parameters and an already recorded output value (Li, Liao, Wu, & Yu, 2001).

This application could be interpreted as an artificial neural network constituted by a primary neural layer (where the data of the input variables would be collected), an output neural layer (where the collected value would be obtained), and one or various intermediate layers (where the convergence work of the neural network would be facilitated) (Fig. 2).

Due to the great variety of neural networks developed at present, and after studying the case that concerns us, it has been chosen a supervised perceptron multilayer (MLP) neural network model in which the data for the training of the neural network are constituted by both input and output training variables. The fact of knowing the output implies that the training benefits the supervision of a teacher. In the training phase, it has been used a previously known training data or pattern set dated from 1995 to 2005 (which represents the 80% of all data available). These

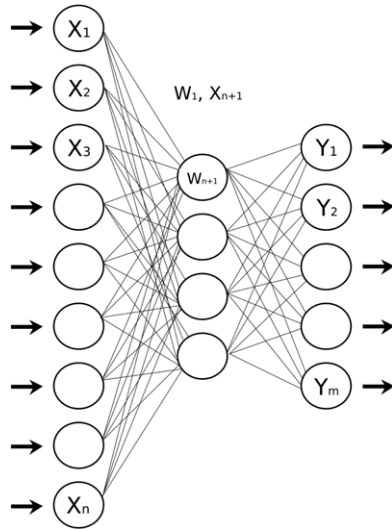


Fig. 2. Diagram of perceptron network constituted by a four-neuron input layer, a hidden three-neuron layer and an output neuron. W represents the weight among the various neurons from each layer.

data were constituted by 18 input variables previously normalized before used in the ANN (day number from 1st January, daily mean pollen concentration (95–07), atmospheric pollen season following the 95% method, accumulated pollen, rainfall, humidity, $T_{\max}$, $T_{\max}$ accumulated, $T_{\min}$, T_{med} , solar radiation, wind direction and calm), this variables have been chosen as they are the typical available parameters of a standard meteorological station used for the meteorological and aerobiological forecast. ANN were constituted with an intermediate layer of 37 neurons and a sole output neuron ([Poaceae]) that corresponds with the pollen value of the third day posterior to the data gathering. As one of the input variables is de Julian Date, a temporal sequence of cases is generated; hence there is a unique possibility of introduction of input cases in the ANN. The number of neurons in the intermediate layer was tested between $n/2 + 1$ and $2n + 1$, where n corresponds with the input variables. Once trained, the ANN correct functioning has been tested with the validation data corresponding to 2006 and 2007. Finally, $2n + 1$ was chosen in the bases of standard deviation and R^2 value for the training data correlation.

The functioning of the neural network could be described as follows: each neuron from the primary layer collects the data of the input variables and presents them according to an input vector

$$x^p = (x_1^p, x_2^p, \dots, x_N^p)^T. \quad (1)$$

This input vector spreads towards the intermediate layer by means of the following propagation rule

$$s_i^p = \sum_{j=1}^N w_{ji} x_j^p + b_i \quad (2)$$

where N is the number of the network input neurons, w_{ji} is the weight value of the connection between the neuron j from the input layer and the neuron i from the intermediate layer, and b_i is the value of the “bias” associated to the neuron i .

If it is assumed that the activation state of the neuron i is the function of the network input vector, then the network output derives from:

$$y_i^p = F_i(s_i^p). \quad (3)$$

Similarly, for any neuron k from the output layer, the equations that determine its activation state are:

$$s_k^p = \sum_{i=1}^L w_{ik} y_i^p + b_k \quad (4)$$

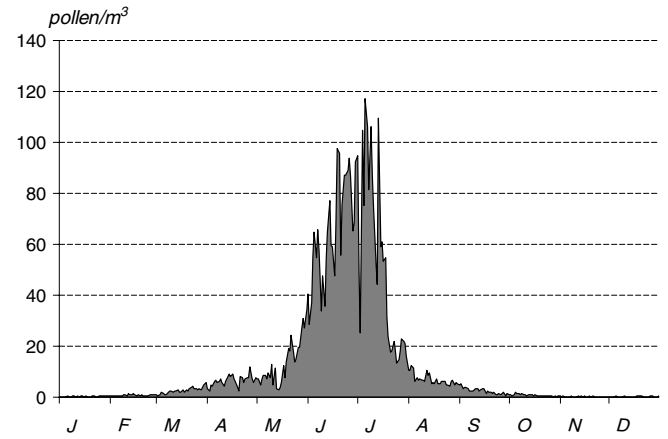


Fig. 3. Average Poaceae pollen concentrations during the period 1995–2007.

$$y_k^p = F_k(s_k^p) \quad (5)$$

where L is the neuron number of the intermediate layer, w_{ik} is the weight value of the connection between the neuron i from the intermediate layer and the neuron k from the output layer, and b_k is the value of the “bias” associated with the neuron k .

The term of error for the output neuron is calculated by means of the following equation

$$E^p = \frac{1}{2} \sum_{k=1}^M (d_k^p - y_k^p)^2 \quad (6)$$

and, if it adjusts to the previously established value, the training of the neural network finishes here. On the contrary, if it does not adjust to the previously established margins of error, the process would be repeated until reaching the desired error value.

The activation equation used in this article is the sigmoid or logistic function (Davoian & Lippe, 2008; Szu, 2007)

$$F_k(s_k^p) = \frac{1}{1 + e^{-s_k^p}}. \quad (7)$$

A Backpropagation rule (BP), which is a typical gradient-based learning algorithm, was used as learning rule in the present work (Chauvin & Rumelhart, 1995; Rumelhart & McClelland, 1986). This learning rule presents an important limitation, which is the large number of input cases for the training process, but, for our study, a large amount of cases are available

Finally, we applied Scheffé’s test to study the homogeneity of the populations under study, which would enabled us to identify the presence of anomalous years, which were out of line with the general behaviour and which could produce alterations in the validation of the prediction models. The analysis was carried out taking into account the different meteorological parameters throughout the years under study.

3. Results and discussion

Poaceae pollen is found in the atmosphere of Vigo during the months of spring and summer (Fig. 3), representing almost a 25% of the total annual pollen.

The period of pollination is very long, of approximately 141 days, due to the inclusion in this pollen type of a high number of species with a flowering period ranging from March to September (Table 1). The beginning of the atmospheric pollen season took place as an average the day April 6th, while the end took place in August 29th as a mean value. The highest concentrations are normally attained during the second fortnight of June or the first

Table 1

Characteristics of *Poaceae* atmospheric pollen season calculated according to 95% method (Andersen, 1991): date of the start and end of the pollen season, length of the pollen season, maximum pollen concentration registered and its date, total annual *Poaceae* pollen grains and number of days in which pollen concentrations up to 25 or 50 pollen/m³ have been attained.

Pollen season	Start date	Final date	Length (days)	Maximum value (pollen/m ³)	Maximum value date	Total pollen	Days 25 pollen/m ³	Days 50 pollen/m ³
1995	31-Mar	2-Sep	155	298	6-Jul	4377	43	25
1996	15-Apr	27-Aug	134	437	25-Jun	7523	54	31
1997	12-Mar	4-Sep	176	282	17-Jul	4467	35	24
1998	28-Mar	20-Aug	145	453	4-Jul	6855	52	33
1999	6-Apr	19-Aug	135	437	7-Jul	4871	40	28
2000	9-Apr	25-Aug	138	288	24-Jun	4377	29	21
2001	9-Apr	2-Sep	146	195	2-Jul	3365	30	16
2002	9-Apr	26-Sep	139	509	14-Jul	4059	31	16
2003	4-Apr	27-Aug	144	338	4-Jul	3890	36	19
2004	10-Apr	25-Aug	127	223	13-Jun	2559	19	13
2005	30-Apr	23-Aug	115	326	20-Jun	3662	31	18
2006	12-Apr	26-Aug	136	230	6-Jun	4791	46	31
2007	29-Mar	25-Sep	170	231	6-Jul	4364	43	23
Average	6-Apr	27-Aug	143	327	29-Jun	4551	38	23

Table 2

Importance of variables considered for ANN.

Value	Importance	Value	Importance
Accumulated pollen	148.3056	Winds coming from SE–N direction	102.9063
Day number from 1st January	148.1465	Relative humidity	102.3167
Average daily mean pollen concentration (95–07)	129.7562	$T_{\max}$	101.6779
Daily mean pollen concentration	120.9281	Winds coming from SE–N direction	100.6219
$T_{\min}$	118.9431	Wind speed	93.6525
Daily mean pollen concentration dia 95%	117.3036	Solar radiation	90.0841
$T_{\max}$ accumulated	114.6796	Winds coming from S–SE direction	79.8752
Rainfall	104.6394	T_{mean}	74.9217
Winds coming from N–NE direction	104.4028	Wind calm	73.8833

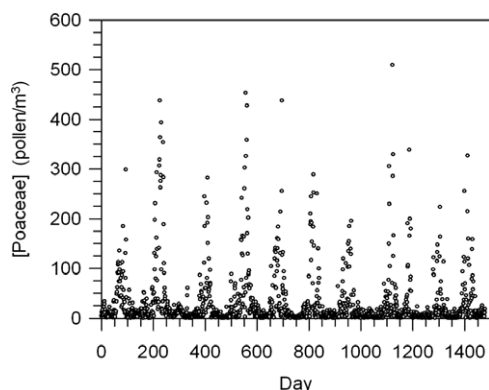


Fig. 4. Temporal distribution of the training data used for the training of the network.

of July. The maximum peak during the years of study occurred on 2002 July 14th with a value of 509 pollen/m³.

The annual total pollen registered during the period of study was of 4566 pollen grains. Davies and Smith (1973) and Ong, Taylor, and Knox (1997) pointed out that concentrations of 25 pollen/m³ induce moderate allergy phenomena. These concentrations were reached in Vigo up to 54 days in 1996. Also, values superior to 50 grains/m³ are able to produce clinical symptoms in all patients. The number of days in which the pollen type reached this value ranged between 13 in 2004, and 33 days in 1998 (Fig. 4). Similar values are attained in other areas from Galicia (Rodríguez-Rajo, 2000), a region where, in general, the 80% of the patients suffering from pollinosis are sensitive to the pollen of the *Poaceae* family (Arenas et al., 1996; Ferreiro, Núñez, Rico, Soto, & López, 1998).

These wide aforementioned variations described for *Poaceae* pollen concentrations make it difficult to obtain prediction models. Previous studies have tackled this subject in aerobiology producing

Table 3

Adjust for the predicted values after training and validation of ANN. (n is the number of values employees, b and a the slope and the origin is the regression line, the coefficient of determination R^2 , RMSE root mean squared error and RE relative error.)

	n	a	b	R^2	RMSE	RE
Training	1600	−5.7288	0.9982	0.9696	16.1228	0.5177
Validation 2006	129	−5.3723	1.0010	0.9380	18.898	0.5394
Validation 2007	147	−0.412	0.95	0.9858	1.2916	0.1149

ANN models to predict pollen data (Arizmendi et al., 1993; Castellano-Méndez et al., 2005; Sánchez-Mesa et al., 2005). Therefore, it has been chosen the implementation of a neural network whose purpose was the prediction of the *Poaceae* pollen.

The training of the ANN and the obtaining of the weight between neurons, were followed by the verification of the obtained results (Fig. 4). Table 2 shows the importance of the variables selected for the ANN, this value is the sum of weights of the input neurons to all intermediate neurons.

As seen in Fig. 5, the values calculated for the neural network after the training phase adjust in an almost ideal way to the real values. If the values predicted by the ANN for the training cases are represented in comparison with the real values, it can be observed that there is a good correlation between the predicted values and the real ones. In fact, the coefficient of linear regression (R^2) for these values is 0.9696 for a χ^2 of 356142.3249.

Once the good functioning of the ANN was verified, it was also checked by using the previously reserved validation data corresponding to the year 2006 and 2007. As shown in Fig. 6, the data predicted by the ANN were confronted with the previously reserved validation data. The correlation is practically linear when R^2 is 0.9380 and χ^2 is 40892.3782.

The adjust for the predicted values after training and validation of ANN is shown in Table 3.

The accuracy of the neural network developed was tested with data from 2006 and 2007, which was not taken into account to

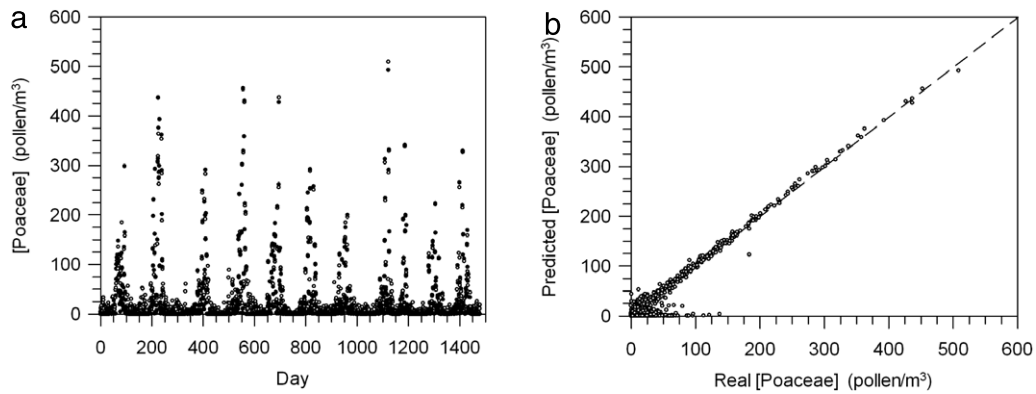


Fig. 5. Real training data: (a) Temporal distribution of the real training data ($\circ$) and the data calculated by the ANN ($\bullet$). (b) Correlation between the real training data and the data calculated by the ANN.

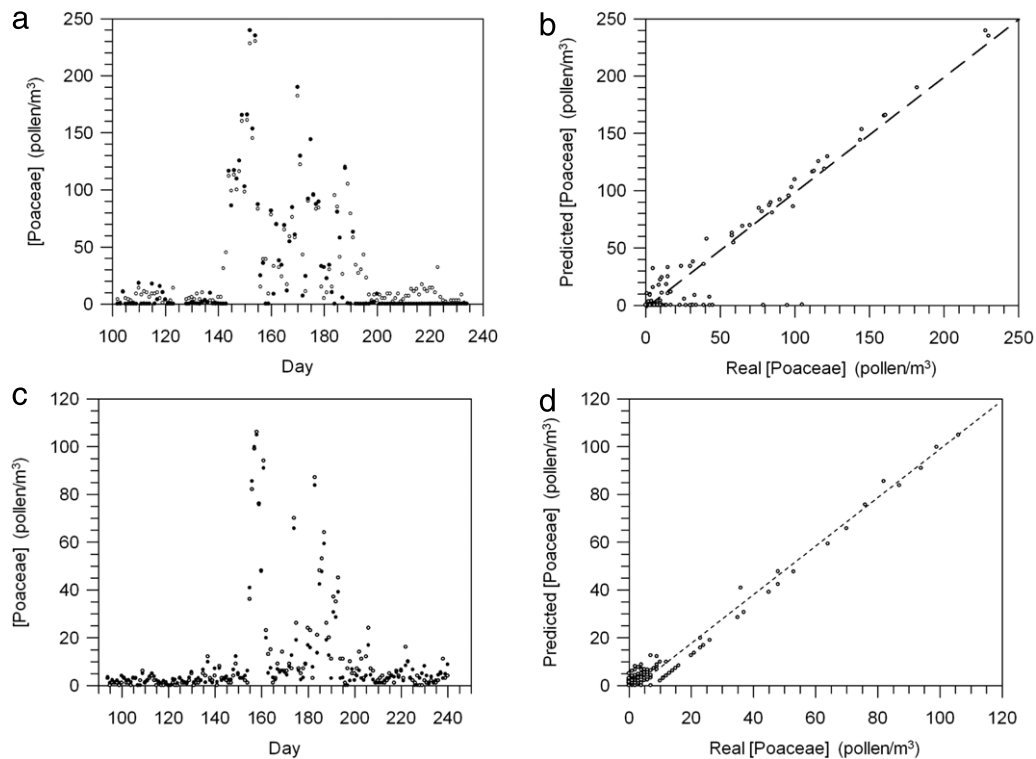


Fig. 6. Validation data: (a) Temporary distribution of the 2006 validation data ($\circ$) and the calculated data ($\bullet$). (b) Correlation between 2006 real validation values and values calculated by the ANN. (c) Temporary distribution of the 2007 validation data ($\circ$) and the calculated data ($\bullet$). (d) Correlation between 2007 real validation values and values calculated by the ANN.

establish the aforementioned models. Scheffé's test was carried out to identify anomalous years that could alter the validation of the general prediction models proposed. Table 4 shows an example of the Scheffé test coefficients considering the average temperature (as the rest of the main meteorological parameters indicate a similar behaviour) and the annual average values of the temperatures, sun hours and humidity. In a general way the statistical analysis reveals significant differences mainly in the case of the year 2006 with the most part of the years considered in our study.

The 2006 is the warmest year of the considered data set, presenting a high average of sun hours and low relative humidity. Even although this variations described could difficult the validation of the ANN prediction model proposed (since in occasions they alter the natural evolution of the cycle of the grass plants and therefore the levels of pollen grains in the air) our elaborated model was able to predict accurately the concentrations Poaceae pollen in the

atmosphere. In this way the predicted Poaceae pollen concentration curve closely follows the observed variations of the daily mean concentrations in the year 2006 (Fig. 7).

It is worth highlighting the good forecast behaviour of the model as the *t*-test of dependent samples conducted shows that there is no difference between predicted data and observed data (Table 5).

One of the main goals in aeropalynological models is to predict concentration levels which could trigger allergic symptomatology. The proposed model presents a great reliability since, when predicting three days in advance, the probability that the model detect a level equal to or greater than 50 pollen grains/ m^3 is higher than 94%. In the same way, the possibility to raise a false alarm to allergenic people with a level equal to or greater than 25 pollen grains/ m^3 will be of 22%. Likewise, some authors pointed out that Neural networks provided us a good tool for forecasting the probability that a given value pollen concentration

Table 4
Scheffé test coefficients taking into account temperature average values. Yearly average values of maximum, mean and minimum temperatures, sun hours and relative humidity.

	Scheffé test coefficients 2006	Scheffé test coefficients 2007	$T_{\max}$	$T_{\min}$	T_{med}	Sun hours	Humidity
1995	0.344	1.000	19.4	12.2	15.7	6.0	73.4
1996	0.076 *	0.999	18.3	11.4	14.9	6.0	75.5
1997	0.003 ***	0.974	19.6	12.2	15.9	5.9	79.8
1998	0.000 ***	0.001 ***	17.8	10.6	14.2	6.5	78.3
1999	0.000 ***	0.000 ***	16.4	9.1	12.7	3.9	83.0
2000	0.000 ***	0.000 ***	16.0	9.0	12.5	5.8	81.3
2001	0.000 ***	0.175	17.8	10.6	14.2	6.4	73.5
2002	0.000 ***	0.112	18.0	11.1	14.6	5.7	76.7
2003	0.151	0.999	16.6	10.2	13.4		71.5
2004	0.001 ***	0.867	18.4	10.0	14.2	4.3	73.6
2005	1.000	0.856	19.6	11.6	15.6	7.3	66.2
2006			20.5	12.6	16.5	7.1	68.3
2007			20.4	12.0	16.2	7.3	66.7

- * $p < 0.1.$
- ** $p < 0.05.$
- *** $p < 0.01.$

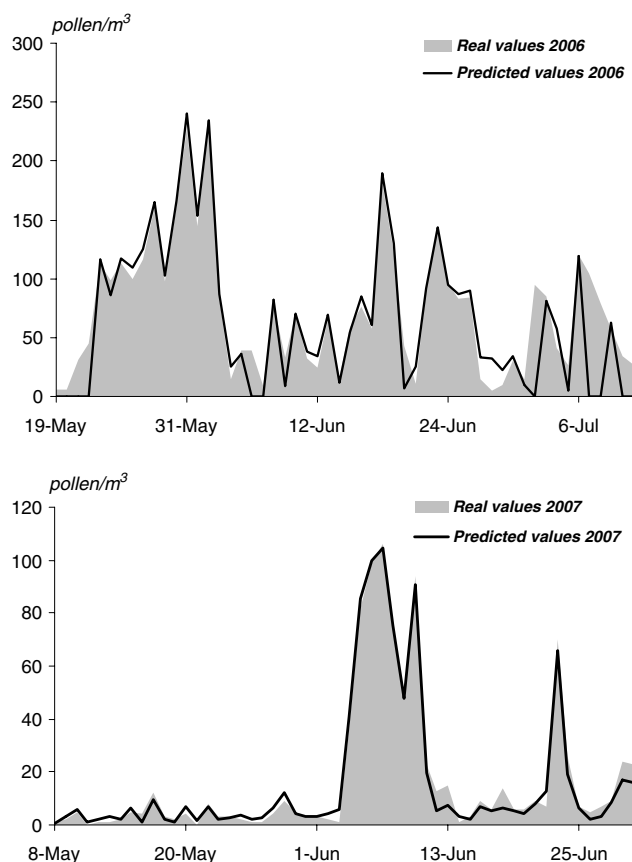


Fig. 7. Predicted versus observed values during the year 2006 and 2007.

Table 5
Test of parallel samples for predicted and observed Poaceae data during the years 2006 and 2007.

	Mean	Std. dv.	<i>N</i>	Diff.	Std. dv. diff.	<i>t</i>	<i>df</i>	<i>p</i>
Predicted data 2006	74.896	61.297	48	4.229	26.253	1.116	47	0.270
Predicted data 2007	28.310	30.850	47	3.300	3.959	5.338	40	0.071

occurs, forecasting with a range from 100% to 83% the prediction in advance of the occurrence of different thresholds pollen concentrations (Castellano-Méndez et al., 2005).

Our results are better than those obtained by classical methodology, both concerning the prediction horizon (since the pollen data is available 72 h before the day whose pollen concentration we want to determine) and the pollen concentration predicted. The most common forecasting short-term concentration models in aerobiology are developed making use of statistical analysis such as time series (Cotos-Yañez et al., 2004) or linear regressions (Chuine & Belmonte, 2004; Laaidi, 2001). Previous studies conducted for Sánchez-Mesa et al. (2005) comparing different methodologies show that the prediction of Poaceae pollen concentration can be made with great accuracy by using ANN (forecasting pollen concentrations with a 90% of probability) than with linear regression (80% of probability).

Neural network models are able to relate different kinds of variables by following a nonlinear path, yielding better results (Sánchez-Mesa et al., 2005). Generally linear regression models for predicting pollen concentrations use only meteorological variables as prediction parameters producing results with a low prediction level, which shows that meteorological variables do not explain such behaviour on their own. Other variables that better reflect the factors affecting the plant, and on which its pollen production and release depend, should be taken into account. In this regard, ANN allows us to consider as forecaster variables other parameters as day number from 1st January or daily mean pollen concentration that substantially improve the prediction capacity.

4. Conclusions

The results obtained in this work show a high coefficient of linear regression between the values predicted by the ANN for the training cases in comparison with the real pollen concentrations. The accuracy of the neural network developed was tested with data from 2006 and 2007, which was not taken into account to establish the aforementioned models. In spite of Scheffé test state the 2006 is an anomalous year that could alter the validation of the forecast model proposed, the predicted Poaceae pollen concentration curve closely follows the observed variations of the daily mean concentrations. The *t*-test of dependent samples conducted shows that there is no difference between predicted data and observed data.

Neural networks provided us a good tool to forecasting Poaceae airborne pollen concentration, and thus could help the automation of the prediction system in the aerobiological information diffusion to the population suffering from allergic problems. The prediction obtained by ANNs represents an improvement against other relationships where the correlation coefficients are much smaller ($R^2 < 0.4$ vs. $R^2 > 0.7$) than that the corresponding one obtained

in the present work (Castellano-Méndez et al., 2005; Cotos-Yañez et al., 2004; Rodríguez-Rajo, Dacosta, & Jato, 2004; Rodríguez-Rajo et al., 2006).

Our proposed model would be applied to different geographical areas; nevertheless, the adjustment of the model by using the available and adequate variables would be realised in each case.

Acknowledgements

The study was supported by different contracts financed by the “Consellería de Medio Ambiente” and “Consellería de Sanidade” of the “Xunta de Galicia”. G. Astray thanks the “Xunta de Galicia” for a training research grant (P.P. 0000 300S 14008).

References

- Andersen, T. B. (1991). A model to predict the beginning of the pollen season. *Grana*, 30, 269–275.
- Arenas, L., González, C., Tabarés, J. M., Iglesias, I., Méndez, J., & Jato, V. (1996). Sensibilización cutánea a pólenes en pacientes afectados de rinoconjuntivitis-asma en la población de Ourense en el año 1994–95. In *Proceedings of the first European symposium on aerobiology* (pp. 93–94).
- Arizmendi, C. M., Sánchez, J. R., Ramos, N. E., & Ramos, G. J. (1993). Time series predictions with neural sets: Application to airborne pollen forecasting. *International Journal of Biometeorology*, 37, 139–144.
- Bishop, M. C. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Caramiello, R., Siniscalco, C., Mercalli, L., & Potenza, A. (1994). The relationship between airborne pollen grains and unusual weather conditions in Turin (Italy) in 1989, 1990 and 1991. *Grana*, 33, 327–332.
- Castellano-Méndez, M., Aira, M. J., Iglesias, I., Jato, V., & González-Manteiga, W. (2005). Artificial neural networks as a useful tool to predict the risk level of *Betula* pollen in the air. *International Journal of Biometeorology*, 49, 310–316.
- Castillo, E., Gutiérrez, J. M., Hadi, A. S., & Lacruz, B. (2001). Some applications of functional networks in statistics and engineering. *Technometrics*, 43, 10–24.
- Chauvin, Y., & Rumelhart, D. E. (1995). *Backpropagation: Theory architectures and applications*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Chuine, I., & Belmonte, J. (2004). Improving prophylaxis for pollen allergies: Predicting the time course of the pollen load of the atmosphere of major allergenic plants in France and Spain. *Grana*, 43, 1–17.
- Clot, B. (2001). Airborne birch pollen in Neuchâtel (Switzerland): Onset, peak and daily patterns. *Aerobiologia*, 17, 25–29.
- Cotos-Yañez, T., Rodríguez-Rajo, F. J., & Jato, V. (2004). Short-term prediction of *Betula* airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models. *International Journal of Biometeorology*, 48, 179–185.
- Davies, R. R., & Smith, L. P. (1973). Forecasting the start and severity of the hay fever season. *Clinical Allergy*, 3, 263–267.
- Davoian, K., & Lippe, W. M. (2008). Exploring the role of activation type in evolutionary artificial neural networks. In *Proceedings of the 2008 international conference on data mining* (pp. 443–449).
- D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., et al. (2007). Allergenic pollen and pollen allergy in Europe. *Allergy*, 62(9), 976–990.
- Dickinson, C. E., & Dodd, J. L. (1976). Phenological pattern in the Shortgrass Prairie. *American Midland Naturalist*, 96, 367–378.
- Ferreiro, M., Núñez, M., Rico, A., Soto, T., & López, R. (1998). Pólenes alergénicos y polinosis en el área de A Coruña. *Revista Española de Alergología e Inmunología Clínica*, 13, 98–101.
- French, N., & Sauer, R. H. (1974). Phenological studies and modelling in grassland. In H. Leith (Ed.), *Phenology and seasonality modelling*. New York: Springer-Verlag.
- Galan, C., Cariñanos, P., Alcazar, P., & Dominguez, E. (2007). *Spanish aerobiology network: Management and quality manual*. University of Córdoba.
- García-Mozo, H., Chuine, I., Aira, M. J., Belmonte, J., Bermejo, D., de la Guardia, C. D., et al. (2008). Regional phenological models for forecasting the start and peak of the *Quercus* pollen season in Spain. *Agricultural and Forest Meteorology*, 148, 372–380.
- Green, B. J., Dettmann, M., Yli-Panula, E., Rutherford, S., & Simpson, R. (2004). Atmospheric Poaceae pollen frequencies and associations with meteorological parameters in Brisbane, Australia a 5-year record, 1994–1999. *International Journal of Biometeorology*, 48, 172–178.
- Grinn-Gofron, A., & Strzelczak, A. (2008). Artificial neural network models of relationships between *Cladosporium* spores and meteorological factors in Szczecin (Poland). *Grana*, 47, 305–315.
- Haykin, S. (1994). *Neural networks*. New York: Macmillan College Publishing Company.
- Hirst, J. (1952). An automatic volumetric spore-trap. *Annals Applied Biology*, 36, 257–265.
- Jato, V., Rodríguez-Rajo, F. J., Méndez, J., & Aira, M. J. (2002). Phenological behaviour of *Quercus* in Ourense (NW Spain) and its relationship with the atmospheric pollen season. *International Journal of Biometeorology*, 46, 176–184.
- Laaidi, M. (2001). Regional variations in the pollen season of *Betula* in Burgundy: Two models for predicting the start of the pollination. *Aerobiologia*, 17, 247–254.
- Li, C., Liao, X., Wu, Z., & Yu, J. (2001). Complex functional networks. *Mathematical Computer Simulation*, 57, 355–365.
- Negrini, A. C., Voltolini, S., & Arobba, C. T. (1992). Comparison between Urticaceae (*Parietaria*) pollen count and hay fever symptoms: Assessment of a “threshold-value”. *Aerobiologia*, 8, 325–329.
- Norris-Hill, J. (1995). The modelling of daily Poaceae pollen concentrations. *Grana*, 34, 182–188.
- Ong, E. K., Taylor, P. E., & Knox, R. B. (1997). Forecasting the onset of the grass pollen season in Melbourne (Australia). *Aerobiologia*, 13, 43–48.
- Rantio-Lehtimäki, A., Koivikko, A., Kupias, R., Makinen, Y., & Pohjola, A. (1991). Significance of sampling height of airborne particles for aerobiological information. *Allergy*, 46, 68–76.
- Rodríguez-Rajo, F. J. (2000). *El polen como fuente de contaminación ambiental en la atmósfera de la ciudad de Vigo*. Ph.D. University of Vigo.
- Rodríguez-Rajo, F. J., Dacosta, N., & Jato, V. (2004). Airborne olive pollen in Vigo (Northwest Spain): A survey to forecast the onset and daily concentrations of the pollen season. *Grana*, 43, 101–110.
- Rodríguez-Rajo, F. J., Valencia-Barrera, R. M., Vega-Maray, A. M., Suarez, F. J., Fernandez-Gonzalez, D., & Jato, V. (2006). Prediction of airborne *Alnus* pollen concentration by using Arima models. *Annals of Agricultural and Environmental Medicine*, 13, 25–32.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Exploration in the microstructure of cognition*, Vols. 1–2. Cambridge, USA: MIT Press.
- Sánchez-Mesa, J. A., Galán, C., & Hervás, C. (2005). 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. *International Journal of Biometeorology*, 49, 355–362.
- Sánchez-Mesa, J. A., Galán, C., Martínez, J. A., & 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. *Clinical Experimental Allergy*, 32, 1606–1612.
- Subiza, J. (2003). Gramíneas: Aerobiología y polinosis en España. *Alergología e Inmunología Clínica*, 18, 7–23.
- Szu, H. (2007). Theories of neural networks leading to unsupervised learning. In *IEEE international conference on neural networks – Conference proceedings* (pp. 3116–3123).
- Taylor, P. E., Jacobson, K. W., House, J. M., & Glovsky, M. M. (2007). Links between pollen, atopy and the asthma epidemic. *International Archives of Allergy and Immunology*, 144, 162–170.
- Xu, K., Xie, M., Tang, L., & Ho, S. L. (2003). Application of neural networks in forecasting engine system reliability. *Applied Software Computing*, 2, 255–268.