



Assessment of neural networks and time series analysis to forecast airborne *Parietaria* pollen presence in the Atlantic coastal regions

J. A. Valencia¹ · G. Astray² · M. Fernández-González¹ · M. J. Aira³ · F. J. Rodríguez-Rajo¹

Received: 9 March 2018 / Revised: 8 January 2019 / Accepted: 5 February 2019
© ISB 2019

Abstract

Pollen forecasting models are a useful tool with which to predict episodes of type I allergenic risk and other environmental or biological processes. *Parietaria* is a wind-pollinated perennial herb that is responsible for many cases of severe pollinosis due to its high pollen production, the long persistence of the pollen grains in the atmosphere and the abundant presence of allergens in their cytoplasm and walls. The aim of this paper is to develop artificial neural networks (ANNs) to predict airborne *Parietaria* pollen concentrations in the northwestern part of Spain using a 19-year data set (1999–2017). The results show a significant increase in the length of time *Parietaria* pollen is in the air, as well as significant increases in the annual *Parietaria* pollen integral and mean daily maximum pollen value in the year. The Neural models show the ability to forecast airborne *Parietaria* pollen concentrations 1, 2, and 3 days ahead. A developed model with five input variables used to predict concentrations of airborne *Parietaria* pollen 1 day ahead shows determination coefficients between 0.618 and 0.652.

Keywords *Parietaria* pollen · Artificial neural networks · Modeling · Time series analysis

Introduction

The *Parietaria* genus (Urticaceae family) is made up of individual dicotyledonous perennial plants consisting of many shoots emerging from a common rootstock, which range from 30 to 100 cm in height (Guardia and Belmonte 2004). The species included in the genus present inconspicuous green stalkless flowers clustered in the leaf axils and have wind-pollination strategies (Sousa et al. 2014). The pollen grain is tri-zonoporate with a smaller size-diameter of approximately 13–20 µm than other types of anemophilous weeds (Erdtman 1952; Rodríguez-Damián et al. 2003). *Parietaria judaica* L. is the most common widespread allergenic species of the genus in southern Europe (Masullo et al. 1996). The plant grows in coastal and urban areas on nitrified soils, such as walls or rubble and abandoned plots of land (Guardia and Belmonte

2004). Its pollen allergens are responsible for many severe cases of spring sensitization due to the high production of pollen grains per plant (Guardia and Belmonte 2004), the long flowering period of the plant (Serafini 1957) and the abundant presence of allergens in the cytoplasm and walls (Vega-Maray et al. 2004). *P. judaica* pollen contains at least nine identified allergens, including two major allergens, Par j 1 and Par 2, belonging to the non-specific lipid-transfer protein family—nsLTP (Asturias et al. 2003; Costa et al. 1994; Suárez-Cervera et al. 2008). The sensitization rates of *Parietaria* prevalence have been reported as 25–50% in Spain and southern France (Masullo et al. 1996), 14.7% in Portugal, 20.6% in Greece, and 30.7% in Italy (Burbach et al. 2009). In the area of northwestern Spain, prevalence of between 6 and 25% of the pollen-sensitized people have been reported (Subiza et al. 2007).

Over the last few decades, the incidence of respiratory diseases has increased worldwide (Portier et al. 2010). Nowadays, the monitoring of airborne pollen is an effective tool for identifying the local pollen types that cause respiratory allergies and is used in the prevention, diagnosis and treatment of allergic diseases (Reyes et al. 2016). The production and the annual distribution of the pollen grains in the atmosphere are closely tied to meteorological conditions both before and during the flowering stage of plants (Dahl et al. 2013; Fuhrmann et al. 2016). The

✉ F. J. Rodríguez-Rajo
javirajo@uvigo.es

¹ Department of Plant Biology and Soil Sciences, Faculty of Sciences, University of Vigo, 32004 Ourense, Spain

² Physical Chemistry Department, Faculty of Science, University of Vigo, 32004 Ourense, Spain

³ Botany Department, Faculty of Pharmacy, University of Santiago, 15782 Santiago Compostela, Spain

detected trend in the airborne *Parietaria* pollen variations are particularly remarkable, depending on both the local geographical features and the weather parameters (Recio et al. 2009; Sabo et al. 2016). In this regard, some authors have identified air temperature as being the most important variable that influences the rise of Urticaceae annual pollen integral together with low air humidity (Sabo et al. 2016).

The research about the role of biological particles has importance due to the possible adverse medical, environmental, or biological effects (Sánchez-Mesa et al. 2005). For this reason, forecasting airborne pollen concentrations is a hot topic (Aznarte et al. 2007). In the last decades, different models have been developed to predict pollen concentrations, at either a regional-to-continental or a local scale (Ritenberga et al. 2016). In this sense, artificial neural networks (ANNs) have become one of the most used methodologies for modeling non-linear behaviors in recent years (Astray et al. 2016a; Desai et al. 2008). ANNs are a computational method formed by individual cells (neurons, or cells) present in the biological neural system (with multiple layers) (Akintunde et al. 2015) to simulate data process as the human brain does (Bas and Boyaci 2007). ANN models can learn from incomplete/imprecise and noisy data or ANNs can model non-linear behaviors (Akintunde et al. 2015; Perpetuo et al. 2012). Neural networks have been used in aerobiology to predict atmospheric pollen concentrations (Rodríguez-Rajo et al. 2010; Puc 2012; Astray et al. 2016b) or the risk of pollen concentration exceeding a specific level (Castellano-Méndez et al. 2005) as well as levels of airborne fungal spores (Harteveld et al. 2017; Astray et al. 2010; Grinn-Gofrón and Strzelczak 2008).

This paper aims to forecast airborne concentrations of *Parietaria* pollen by means of ANN and time series analysis. The importance of the results is reinforced by the fact that this plant has not been traditionally considered as an important aeroallergen along the Atlantic coast of Spain. There are indications that *Parietaria* pollen is emerging as an aeroallergen in this area.

Material and methods

Aerobiological survey

The study was conducted in Vigo (42° 14' 11.4"N 8° 43' 35.9" W), located in the northwestern corner of the Iberian Peninsula (Fig. 1). The area is situated at the Eurosiberian region (Rivas-Martínez 1987) under a moist oceanic climate, with a Mediterranean influence. The main bioclimatological characteristics show a mean annual temperature of 14.9 °C, a mean maximum temperature of 21.3 °C, a mean minimum temperature of 8.5 °C, and an annual rainfall amount of 811 mm on average,

irregularly spread throughout the year, as July and August are the driest months (AEMET 2012).

Pollen sampling was carried out by using a Hirst type LANZONI VPPS 2000 volumetric trap working continuously from 1999 to 2017. The device was calibrated with a suction flow rate of 10 L/min (Hirst 1952). The sampler was located on the roof of the town hall building in Vigo, approximately 15 m above ground level. The capture media of the pollen grains was a Melinex tape coated with an adhesive solution. The Spanish Aerobiology Network methodology (Galán et al. 2007) was applied for the counting and identification of the pollen grains by using an optical microscope with a magnification of $\times 40$. Airborne pollen concentrations were expressed as grains per cubic meter of air (Galán et al. 2017). For the data management, the main pollen season (MPS) is considered to be the period during which a 95% of the annual total accumulated sum of pollen is recorded. MPS starts when the accumulated sum of pollen reaches a 2.5% of the annual total and finishes the day a 97.5% of the annual total is reached (Andersen 1991). In order to observe the tendencies of the aerobiological pollen parameters the data were fitted to a simple linear regression line considering the slopes of the regression equations, the determination coefficients (R^2), and the significance levels (p).

The meteorological data were supplied by the AEMET Port of Vigo meteorological station located next to the Hirst pollen sampler. The daily, monthly, annual, and seasonal (spring and autumn) values of the main meteorological parameters were considered: mean, maximum, and minimum temperatures (degree Celsius), sunshine hours (hours), rainfall (millimeter), relative humidity (percent), frequency of wind from the quadrants centered in NE, SE, SW, and NW, and the xi) frequency of calm weather.

Methodology and software used to model *Parietaria* pollen

A first study was conducted using a time series analysis to find an auto-correlation or cross-correlation function between the different available variables. According to the slope obtained in each analysis, it is possible to interpret the memory effect (time to dissipate) (Sahu et al. 2009; Lee and Lee 2000). For this study, *Parietaria* pollen and weather variables as rainfall and temperatures (maximum, minimum, and mean) were used because it was necessary to have the completed data series over a period at regular time intervals.

Once the study of time series had been carried out (years 1999 to 2017), the database was divided into three groups. The first group, which comprises the years between 1999 and 2013, was used to develop the ANN models and the second group, that ranges from the years 2014 to 2015, was used to choose the best neural model and to determine the five most important variables to implement new simplified

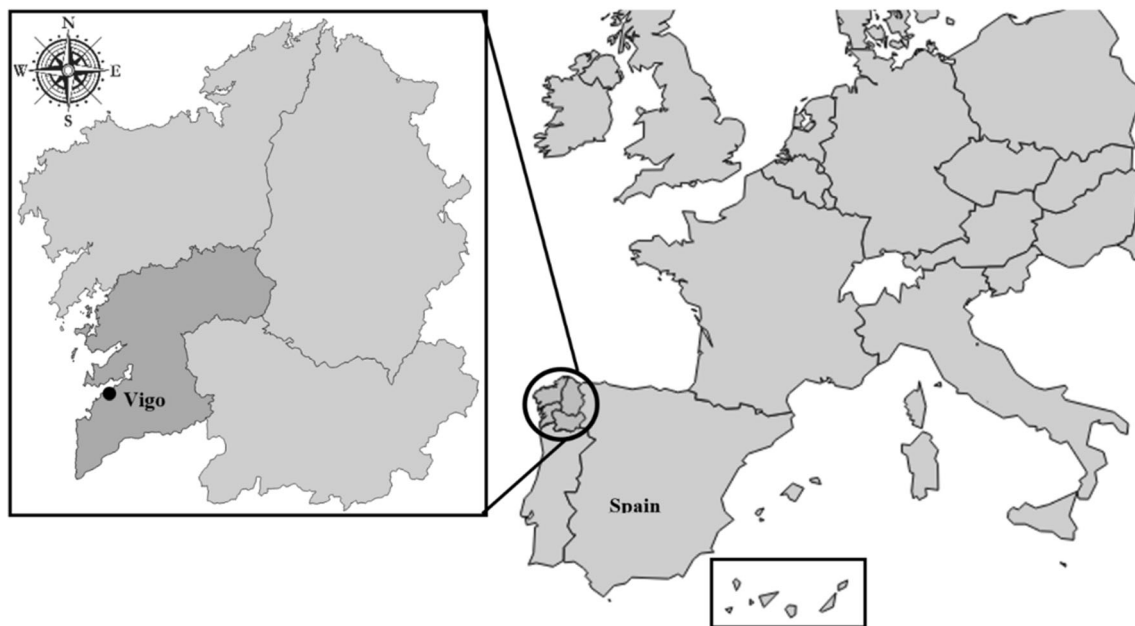


Fig. 1 Location of Vigo in Europe

models. Once the new simplified models were implemented (between the years of 1999 and 2013), the best models were chosen using the years 2014 and 2015. Finally, the most recent years, 2016 and 2017 were used as an external validation of the best simplified models obtained.

Neural network requires a huge implementation of models with different topologies, that is, a different number of neurons in hidden layers. To determine the best topology, the trial and error approach can be used (Astray et al. 2016a, b; Govindaraju 2000). During the training phase, weights (w) and bias value (b) are modified to minimize the error between outputs and targets (Govindaraju 2000). In this research, the logistic function was used in the hidden and output neurons. When all models have been developed, the validation group is used to choose the five models with higher determination coefficient (R^2) and then the best model is chosen, among the five models, according to the lowest root mean square error (RMSE). The multiple model is chosen from the best five models that presents the most representative behavior for each parameter.

PAST v2.17 software (Hammer et al. 2001) and EasyNN plus v14.0d from the Neural Planner Software Ltd. (UK) were used for time series analysis and neural model development, respectively. Data were fitted using Microsoft Excel from Microsoft (USA). Figures were made with Sigmaplot 13 from Systat Software Inc. (USA), Power Point and Microsoft Excel from Microsoft (USA).

Results

Parietaria pollen represents on average a 14.2% of the total airborne annual pollen integral detected in the city

of Vigo during the study period. The timing of the appearance of *Parietaria* pollen in the air varies from year to year, depending on the meteorological conditions. As a general rule, and taking into account the data after these years of research, the main pollen season starts the first few days of March and ends in the middle of October (Table 1). Regarding the phenological differences detected in this genus of plants, a significant ($p = 0.000$) tendency to a delay of the end in the main pollen season was observed (5 days per year) (Table 1, Fig. 2). The pollination period of *Parietaria* is very long, with an average duration of 225 days. A significant ($p = 0.025$) tendency towards an increase of the flowering period (3 days per year) was detected as a consequence of the delay in the end of the main pollen season (Table 1, Fig. 2).

The presence of *Parietaria* pollen grains was not constant in the atmosphere during the year. In general, the highest concentrations were observed during the middle of June, with a maximum peak attained on June 26, 2012 (Table 1). During the period of the study, significant trends towards an increase of the annual *Parietaria* pollen integral ($R^2 = 0.50$) and the concentration of the mean daily maximum pollen value in the year ($R^2 = 0.27$) were detected (Fig. 2). Important inter annual fluctuations on the annual pollen integral were detected during the years of study. The minimum annual pollen integral was registered during 2005 with 1033 pollen grains, while the maximum total annual pollen integral was observed during 2012 with 7180 pollen grains (Table 1).

On the one hand, if we analyzed the behavior of the meteorological variables, we could observe a tendency towards an increase in temperature (maximum, minimum, and average). On the other hand, a negative trend was detected for rainfall

Table 1 Start, end date, and length of the Main pollen season (MPS). Annual *Parietaria* pollen integral (pollen), peak pollen (pollen/m³), and pollen peak date. Lineal trend analysis, R^2 , p level and slope during the study period (1999–2017)

	MPS			Total	Pollen	Date
	Start	End	Length	Pollen	Peak	Peak day
1999	14-Mar	11-Sep	182	2895	112	25-May
2000	18-Feb	2-Oct	228	2091	73	13-Jun
2001	12-Mar	27-Aug	169	1859	116	28-May
2002	25-Feb	25-Oct	243	1554	51	13-Jun
2003	20-Feb	22-Sep	215	1373	66	20-Jun
2004	19-Feb	24-Sep	219	2142	100	26-Apr
2005	23-Mar	13-Oct	205	982	43	3-Jul
2006	14-Mar	27-Oct	228	4361	158	25-May
2007	26-Mar	4-Nov	224	5932	144	29-Jul
2008	27-Feb	9-Nov	257	6825	189	28-Jun
2009	17-Mar	16-Oct	214	6525	159	31-May
2010	24-Mar	21-Oct	212	4548	99	28-Apr
2011	17-Mar	24-Nov	253	5095	131	11-May
2012	18-Feb	20-Nov	277	6152	225	26-Jun
2013	4-Apr	3-Nov	214	5275	187	25-Jun
2014	31-Mar	31-Oct	215	5202	142	13-Jun
2015	16-Mar	12-Dec	272	4462	94	28-May
2016	09-Apr	10-Dec	247	5429	129	27-Jun
2017	01-Feb	20-Nov	293	5769	146	17-Jun
Average	10-Mar	25-Oct	230	4130	124	8-Jun
R^2	0.055	0.758	0.408	0.499	0.241	0.017
p level	0.332	0.000	0.003	0.000	0.032	0.598
Slope	0.728	4.547	3.566	252.58	4.232	0.539

(10 days per year) whereas no tendency was observed for the relative humidity (Fig. 3).

Time series analysis

In order to forecast airborne concentrations of *Parietaria* pollen, auto-correlation analysis and cross-correlation analysis were applied. The auto-correlation for *Parietaria* pollen decreases from 1 to 0.688 after 1 day, and reaches zero value in 94 days. The location of consecutive maximum peaks suggests that *Parietaria* pollen presents a seasonality close to 365 days (with a small displacement probably due to the effect of all the other variables). Auto-correlation function for rainfall decreases faster than *Parietaria* pollen, as expected, and does not show seasonality, while auto-correlation functions for temperatures decrease slower and show a clear seasonality close to 365 days.

To determine the most acceptable prediction window for *Parietaria* pollen the cross-correlation functions for *Parietaria* pollen were studied. The cross-correlogram shows

a positive correlation coefficient between *Parietaria* pollen and *Parietaria* pollen + 1 with a maximum cross-correlation value of 0.688 ($p = 0.000$) for lag 0, and 0.586 ($p = 0.000$), and 0.467 ($p = 0.000$) for lag one, two, and three, respectively. These values reveal that there is a good correlation between *Parietaria* pollen and *Parietaria* pollen + 1, but it clearly decreases with days. These results show that *Parietaria* pollen concentration can be predicted with a maximum accuracy of 68.8% 1 day ahead using pollen concentration.

According to the purpose of this research, artificial neural models were implemented with 13 input variables to predict pollen concentration at 1, 2, and 3 days ahead of using three individual models.

Individual neural models (13 input neurons)

As mentioned earlier, neural network modeling requires a huge implementation of models, using a trial and error approach to obtain a model which is capable of predicting atmospheric *Parietaria* pollen concentrations 1, 2, and 3 days ahead. Over five hundred neural models with different topologies and training cycles were developed. The required time to develop each neural model depends on the number of input variables, the complexity of the topology, and the number of training cycles, among others. All models were developed with 13 input neurons. When the models were implemented, it was necessary to calculate the different adjustments, the determination coefficient, and the root mean squared error (Table 2). The best individual models present a R^2_T between 0.678 for 1 day ahead, and 0.476 for 3 days ahead. These values correspond with a $RMSE_T$ between 11.72 pollen/m³ and 14.89 pollen/m³, 1 or 3 days ahead respectively (Table 2). For the validation phase, these best models presented a determination coefficient between 0.647 and 0.389, 1 or 3 days ahead. Figure 4 shows the adjustments for 1 day ahead (top) are comparable to 2 and 3 days ahead (middle and bottom) for the validation phase. It is necessary to emphasize that the maximum *Parietaria* pollen concentration is underestimated when the time window is expanded to 2 and 3 days ahead (Fig. 4, middle and bottom).

The importance of each variable used can be determined by means of the weights values between input neurons and intermediate neurons. The most important variables for *Parietaria* pollen 1 day ahead were as follows: (i) *Parietaria* daily pollen concentration (pollen concentration at zero day), (ii) day, (iii) maximum temperature, (iv) northwest wind, and (v) minimum temperature. The main variables for *Parietaria* pollen 2 and 3 days ahead are similar to the variables obtained for pollen 1 day ahead. A slight difference was detected regarding the wind variable which leaves the group of the five most important being replaced by mean temperature. In order to choose which five variables could be used for the new simplified models, the average of each variable is calculated. It was

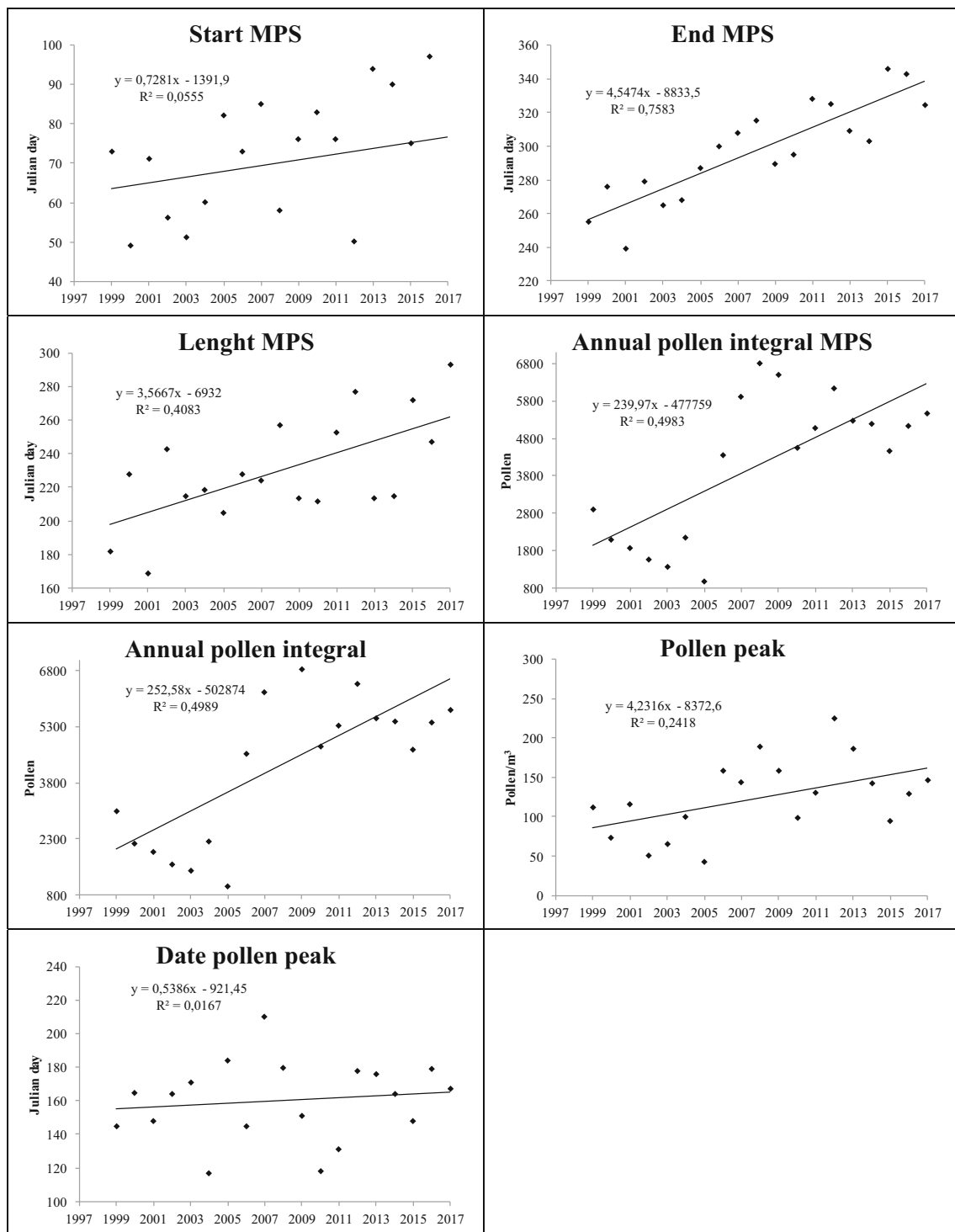


Fig. 2 Trends of the Main Pollen Season characteristics

determined that the new variables were as follows: (i) *Parietaria* pollen concentration; (ii) day; and (iii–v) maximum, mean, and minimum temperature.

Nevertheless, at this time, models with one output (individual model) and one model with three outputs (1, 2, and 3 days ahead, multiple ANN) were developed. About 300 more neural models with the selected five input variables were conducted.

Neural models (five input neurons)

In Table 3, training adjustments for the best neural models (1, 2, and 3 days ahead) using five input variables are shown. Individual models present more unfavorable results for the training phase than the models with 13 input variables with a determination coefficient, between 0.600 for the case of

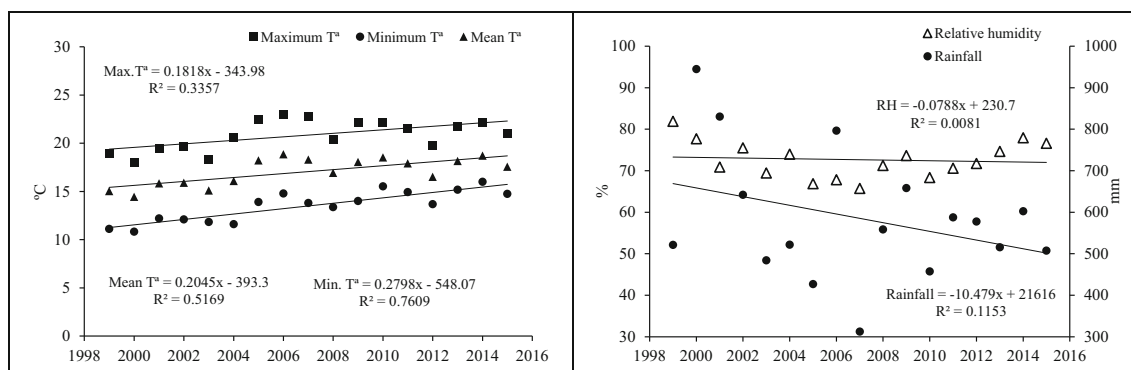


Fig. 3 Trends of the main meteorological variables

1 day ahead, and 0.452 for 3 days ahead. In terms of root mean square error for the training phase, the values vary between 13.15 pollen/m³ for 1 day ahead and 15.22 pollen/m³ for 3 days ahead. On the other hand, the multiple model presents better results than individual models with five input variables and similar results to the individual models with 13 variables. The best multiple model to predict 1 day ahead presents a R^2_T of 0.652 with a RMSE_T of 12.20 pollen/m³. As expected, adjustments are worse when the time window is increased, 3 days ahead presents a R^2_T of 0.456 with a RMSE_T of 15.16 pollen/m³.

Good results for the validation phase (2014 to 2015) are according to the results obtained for the training phase (Table 3). In this case, developed individual neural models to predict *Parietaria* concentration present a determination coefficient for the validation phase between 0.624 and 0.367 for 1 and 3 days ahead, respectively, with RMSEs between 11.90 and 15.53 pollen/m³. For the multiple model, the adjustments present more favorable results than individual models with 13 input variables, in terms of RMSE_V (Tables 2 and 3).

Finally, for the external validation (years 2016 and 2017), the simplified models with five variables offer similar results to the validation phase. The individual models present a R^2 of

0.590, 0.390, and 0.333 for the *Parietaria* 1 day ahead, 2 days ahead, and 3 days ahead, respectively. These values correspond with errors between 13.71 and 17.54 pollen/m³. A multiple neural network offers slightly better results than the individual models, with an R^2 between 0.618, and 0.332 for the *Parietaria* 1 day ahead and 3 days ahead, that correspond with errors between 13.22 and 17.57 pollen/m³.

We can see that predicted values for the *Parietaria* peaks using the multiple model are underestimated (Fig. 5) in the same way as the individual models with 13 variables (Fig. 4).

Discussion

In the studied Atlantic area of Spain, *Parietaria* pollen represents an average of 14.2% of the airborne annual pollen integral. Similar results were found by different authors for Urticaceae pollen in different areas of the Mediterranean Basin, ranging from a 14.6% (Belmonte et al. 1998) to 15.7% (Sabo et al. 2015). *Parietaria* is considered one of the most important allergen sources across the Mediterranean countries (Makris et al. 2010; Fotiou et al. 2011), it can also be found sparsely in temperate regions of

Table 2 Fits for training and validation phase of the best individual neural model developed with 13 input neurons. Topology is the model structure, cycles are the number of training cycles used to develop the neural model, R^2 is the coefficient of determination, and RMSE is the root

Training phase (1999–2013)-individual neural networks-13 input neurons							
Topology	Cycles	R^2_{T+1}	R^2_{T+2}	R^2_{T+3}	RMSE _{T+1}	RMSE _{T+2}	RMSE _{T+3}
13-(5)-1	64,000	0.678			11.72		
13-(3)-1	32,000		0.533			14.13	
13-(6)-1	8000			0.476			14.89
Validation phase (2014–2015)-individual neural networks-13 input neurons							
Topology	Cycles	R^2_{V+1}	R^2_{V+2}	R^2_{V+3}	RMSE _{V+1}	RMSE _{V+2}	RMSE _{V+3}
13-(5)-1	64,000	0.647			12.73		
13-(3)-1	32,000		0.456			14.96	
13-(6)-1	8000			0.389			16.46

mean square error (pollen/m³). The subscripts, T and V correspond to training and validation phase, respectively, while +1, +2, and +3 correspond to prediction window 1, 2, and 3 days ahead, respectively

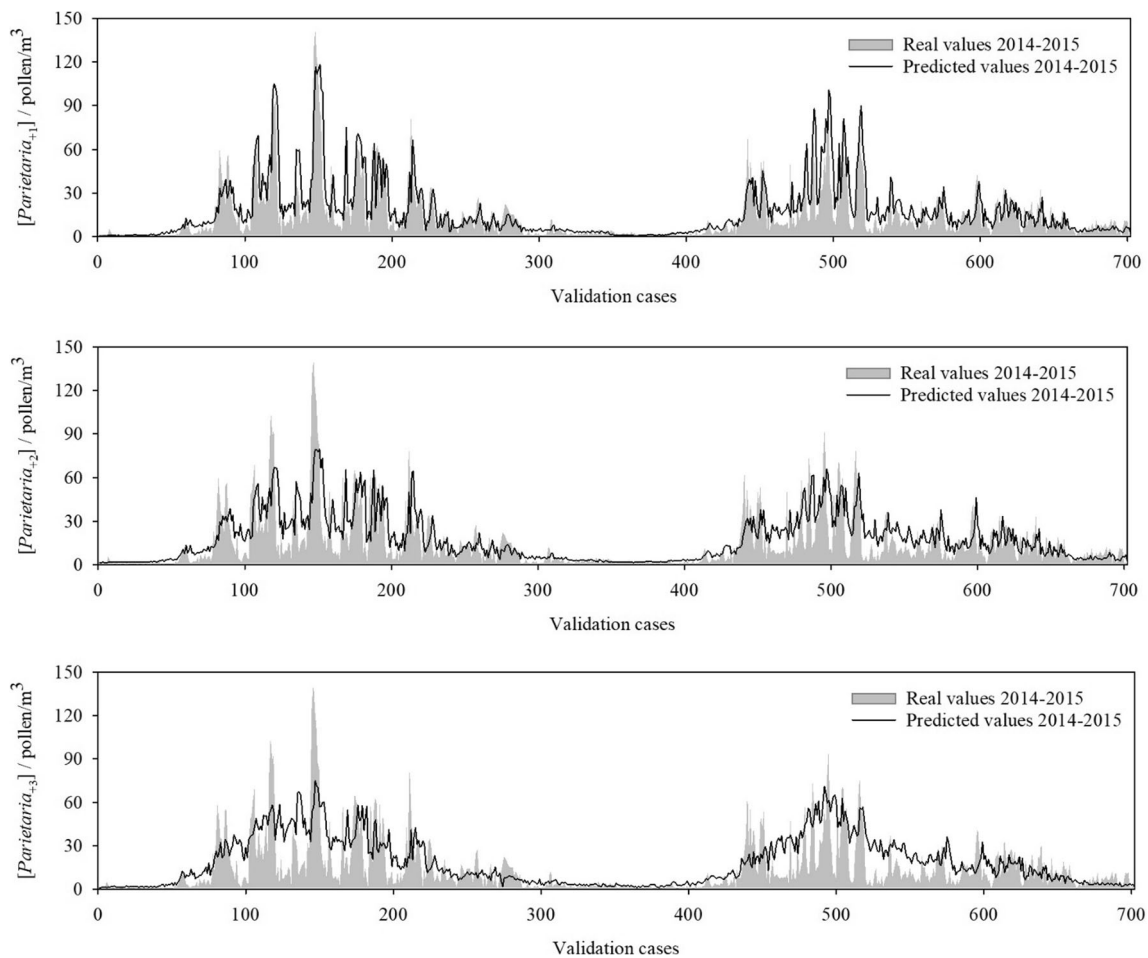


Fig. 4 Predicted and real values of *Parietaria* pollen during validation years 2014 and 2015 to 1 day ahead (top), 2 days ahead (middle), and 3 days ahead (bottom) using individual models with 13 input variables

central and eastern Europe (Holgate et al. 1988), Australia (Ford et al. 1986), and California (Kaufman 1990).

Meteorological variables influence the *Parietaria* pollination period, which takes place from spring to autumn and reaches maximum peaks from May to June (D'Amato 2002). Similar findings were recorded in our study area, with a main pollen season length of 225 days. The detected trend to a delay of the main pollen season end date (5 days per year) was observed in previous studies which noted a tendency towards lower mean minimum temperatures during spring in the coastal Atlantic area (Recio et al. 2009). Temperature and water availability are known to be the determining factors for both the beginning and the end of the pollination period, which explains the prolongation of flowering during the summer months in Vigo (Recio et al. 2009) as a consequence of the warm temperatures registered and the highest water availability by the plant. Temperature and water availability during the beginning or the end of the pollination period show a different behavior for the prediction of airborne *Parietaria* pollen. It can be observed in the list of chosen variables that water (considered in our case as precipitation or relative

humidity) was not selected. Precipitation shows a medium-low importance for the three models, varying between positions six and 11 for the model 3 days ahead and 2 days ahead respectively (on average it is located in the 8 of 13). Relative humidity exerts low importance for the three models (between eight and thirteen) given the fact that it was the last variable in importance as average. This behavior exerts an allergological consequence in the sensitized people due to the increase of the period of time in which they could suffer from pollinosis symptoms. Different authors noted that the long pollination period was the cause of up to 80% of patients with pollinosis in the Mediterranean region and these people are sensitized against *Parietaria* allergens (D'Amato and Labetalol 1989; Pérez and Gorgues 1986; Bousquet et al. 1986). About 52% of these patients suffer rhinitis and asthma (D'Amato et al. 1992; D'Amato 2002).

Nonetheless, the increase of the pollen quantity in the atmosphere during the last few years may be influenced by the meteorological variables. In the present study, we have observed an increased tendency in temperatures and a decreased tendency in rainfall. These differences may be due to climate

Table 3 Fits for training phase and validation phases of the best individual and multiple neural models developed with five input neurons. Topology is the model structure, cycles are the number of training cycles used to develop the neural model, R^2 is the coefficient of determination, and RMSE is the root mean square error (pollen/m³). The subscripts, T and V correspond to training and validation phases, respectively, while + 1, + 2, and + 3 correspond to prediction window 1, 2, and 3 days ahead, respectively

Training phase (1999–2013)-individual neural networks-5 input neurons							
Topology	Cycles	R^2_{T+1}	R^2_{T+2}	R^2_{T+3}	RMSE $_{T+1}$	RMSE $_{T+2}$	RMSE $_{T+3}$
5-(4)-1	2000	0.600			13.15		
5-(3)-1	2000		0.483			15.15	
5-(3)-1	128,000			0.452			15.22
Training phase (1999–2013)-multiple neural networks-5 input neurons							
Topology	Cycles	R^2_{T+1}	R^2_{T+2}	R^2_{T+3}	RMSE $_{T+1}$	RMSE $_{T+2}$	RMSE $_{T+3}$
5-(8)-3	256,000	0.652	0.533	0.456	12.20	14.12	15.16
Validation phase (2014–2015)-individual neural networks-5 input neurons							
Topology	Cycles	R^2_{V+1}	R^2_{V+2}	R^2_{V+3}	RMSE $_{V+1}$	RMSE $_{V+2}$	RMSE $_{V+3}$
5-(4)-1	2000	0.624			11.90		
5-(3)-1	2000		0.446			14.68	
5-(3)-1	128,000			0.367			15.53
Validation phase (2014–2015)-multiple neural networks-5 input neurons							
Topology	Cycles	R^2_{V+1}	R^2_{V+2}	R^2_{V+3}	RMSE $_{V+1}$	RMSE $_{V+2}$	RMSE $_{V+3}$
5-(8)-3	256,000	0.640	0.458	0.373	11.68	14.39	15.48
External validation phase (2016–2017)-individual neural networks-5 input neurons							
Topology	Cycles	R^2_{V+1}	R^2_{V+2}	R^2_{V+3}	RMSE $_{V+1}$	RMSE $_{V+2}$	RMSE $_{V+3}$
5-(4)-1	2000	0.590			13.71		
5-(3)-1	2000		0.390			16.67	
5-(3)-1	128,000			0.333			17.54
External validation phase (2016–2017)-multiple neural networks-5 input neurons							
Topology	Cycles	R^2_{V+1}	R^2_{V+2}	R^2_{V+3}	RMSE $_{V+1}$	RMSE $_{V+2}$	RMSE $_{V+3}$
5-(8)-3	256,000	0.618	0.412	0.332	13.22	16.47	17.57

change patterns occurring in the recent years, since the release of the pollen grains in the Urticaceae family requires suitable conditions of temperature and humidity (González Minero et al. 1997). Temperature is most important on airborne pollen concentrations during spring (Recio et al. 2010; Puc 2012; Ianovici et al. 2013), while humidity has a predominant role in autumn (Galán et al. 2000). No tendency was observed for the relative humidity as consequence of the coastal character of the studied area. Therefore, despite being considered a typically Mediterranean taxon, our study reinforces the allergenic importance of the *Parietaria* pollen grains around the Atlantic coast of northwestern Spain, which may increase over the coming years. The colonization of new habitats by *Parietaria* populations have increased over the last few years with the subsequent elevation of their pollen concentrations in the atmosphere. It has been demonstrated that the atmospheric pollen levels of many species, including Urticaceae, tended to rise during the last decades (Recio et al. 2009) in different European regions (García-Mozo et al. 2006; Rodríguez-Rajo et al. 2004; Tedeschini et al. 2006). A concomitant general increase in pollinosis cases during the same period has also been detected (Beggs 2004).

The aforementioned findings reinforce the need of the short-term prediction of airborne *Parietaria* pollen concentrations in order to check whether allergic individuals could take preventative measures to protect themselves from the severity

of the pollen season. Statistical models enabling the short-term prediction of atmospheric pollen concentrations were mainly conducted making use of statistical tools such as additive models (Cotos-Yañez et al. 2004), Time series analysis (Rodríguez-Rajo et al. 2006) or models based on the correlation of pollen levels with meteorological variables by means of uni-dimensional linear equations using variables with the greatest prediction capacity (Rodríguez-Rajo et al. 2004). ANN models have demonstrated a higher efficiency to predict airborne pollen concentrations than these traditional aerobiological forecasting methodologies (Sánchez-Mesa et al. 2005; Astray et al. 2016b).

The prediction models should be approached from a point of view of reliability and ease of implementation. As can be seen in models with the most variables, the results for the training phase are better in terms of R^2 ; however, the determination coefficient for the multiple model with five variables has remained similar in the validation phase and, what is more important, for this phase, the RMSE has been improved. The possible reason for this can be the noise suppression when the input variables are reduced. As stated above, our research group developed another three different neural models to predict *Parietaria* pollen (1, 2, and 3 days ahead) in other inland areas (around 75 km away) of northwestern Spain (Iglesias-Otero 2013). In this case, the models presented low R^2 for 1 day ahead (0.327), 2 (0.251), and 3 days ahead (0.248) for

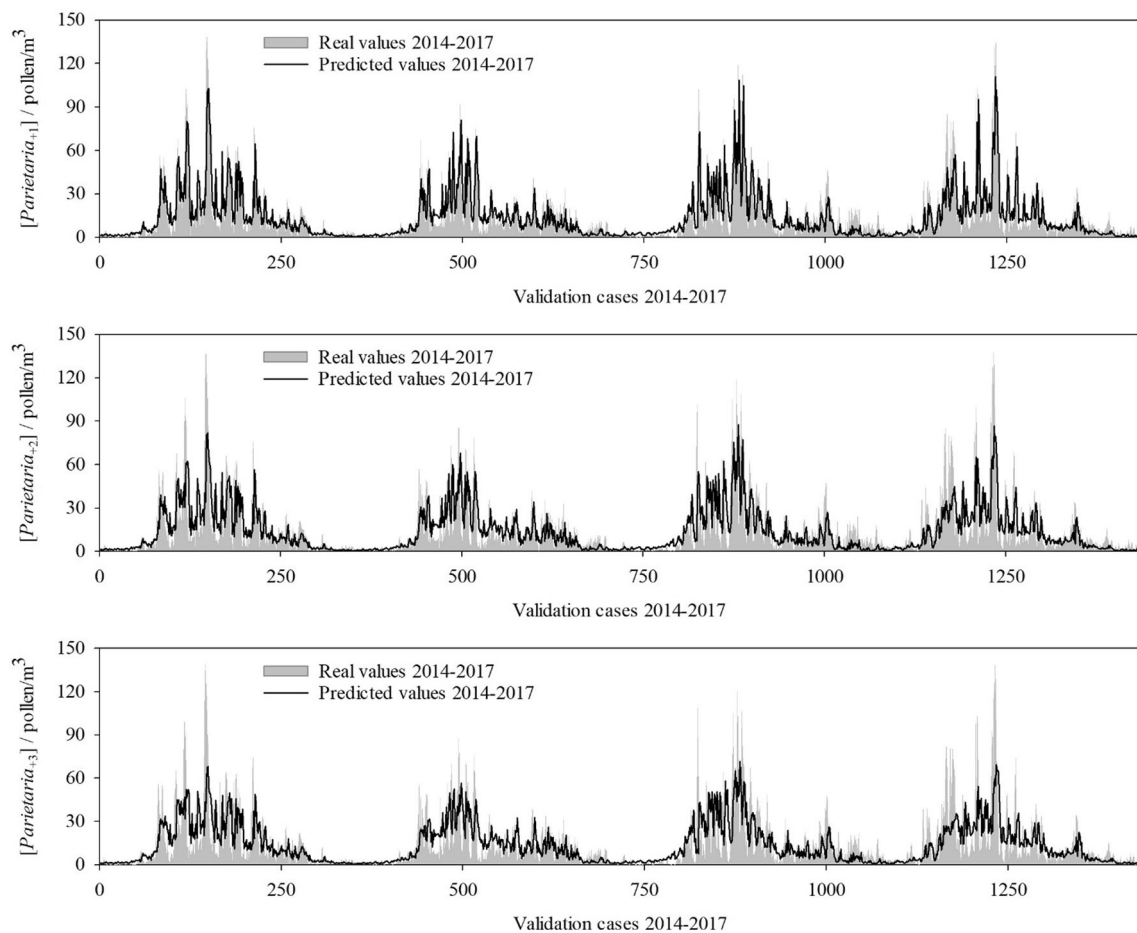


Fig. 5 Real and predicted values of *Parietaria* pollen for validation years 2014 and 2015 and for external validation years 2016 and 2017 to 1 day ahead (top), 2 days ahead (middle), and 3 days ahead (bottom) using multiple model with 5 input variables

a validation period (years 2009 and 2010). The input variables used by Iglesias-Otero (2013) were as follows: (i) day, (ii) rainfall, (iii) relative humidity, (iv) sunshine hours, and (v) *Parietaria* daily pollen concentration. It is for this reason that we can affirm that a good variable selection is fundamental to predict the concentration of *Parietaria* pollen. Moreover, it allowed us to eliminate one of the biggest drawbacks of artificial neural networks, the long implementation time due to their computationally intensive procedure (large number of inputs and cases) (Tu 1996).

The proposed neural models have demonstrated a higher efficiency to forecast *Parietaria* allergenic risk periods with a prediction horizon for 1, 2, and 3 days ahead. Mild symptoms were registered with *Parietaria* concentrations above 10–15 pollen/m³ and severe symptoms (over 90% of patients suffer from pollinosis) occurred with airborne concentrations exceeded 80 pollen/m³ (Negrini et al. 1992). Even the 3-day ahead model is able to accurately predict the days with the possibility of mild allergy symptoms as well as the days with risk of severe pollinosis symptoms. A strength of the models proposed in our study was the use of observed meteorological data of a given day to predict the pollen values during the next

3 days. Investigations conducted by Ritenberga et al. (2016) pointed out that the models based on meteorological parameters obtained by Numerical Weather Prediction forecasts can potentially be a roadblock because the uncertainty of the weather forecast will add to the uncertainty of the pollen model itself and, in case of significant issues, disqualify the pollen prediction. This shortcoming could be detected for several pollen prediction models in literature as many of them use future meteorological data that should be estimated.

Conclusions

The current paper presents a local-scale ANN statistical model for *Parietaria* pollen forecasting calibrated with observed time series of pollen concentrations and meteorological variables. A specific application is demonstrated for Vigo, but the methodology is applicable for other Atlantic coastal locations. Neural models have demonstrated a higher efficiency to predict atmospheric *Parietaria* pollen concentrations (for 1, 2, and 3 days ahead) than traditional aerobiological forecasting methodologies such as time series analysis.

This study provides a new approach in areas such as the Atlantic coast of Spain where this weed has not been considered an important aeroallergen. We show a significant increase in the length of time *Parietaria* pollen is in the air, as well as significant increases in the annual *Parietaria* pollen integral and mean daily maximum pollen value in the year. This circumstance could exert an important impact in public health due to the tendency to the aggravation and the increase of the type I allergy cases detected in recent years. We recommend further improvement of the model to make the prediction model much robust with new experimental data.

Funding information This study was supported by the contract CO-0082-16 on behalf of Xunta of Galicia, Consellería de Sanidade (Dirección Xeral Salud Pública). Astray G. received the postdoctoral grant (Plan 12C), POS-A/2012/164 and the postdoctoral grant B, POS-B/2016/001 which is from the Xunta de Galicia, Consellería de Cultura, Educación e Ordenación Universitaria.

References

- Agencia Estatal de Meteorología – AEMET (2012) Valores climatológicos normales. Ourense. <http://www.aemet.es/es/serviciosclimaticos/datosclimatologicos/valoresclimatologicos?l=1690A&k=gal>. Accessed 10 Feb 2018
- Akintunde AM, Ajala SO, Betiku E (2015) Optimization of *Bauhinia monandra* seed oil extraction via artificial neural network and response surface methodology: a potential biofuel candidate. *Ind Crop Prod* 67:387–394
- Andersen TB (1991) A model to predict the beginning of the pollen season. *Grana* 30(1):269–275
- Astray G, Rodríguez-Rajo FJ, Ferreira-Lage A, Fernandez-Gonzalez M, Jato V, Mejuto J (2010) The use of artificial neural networks to forecast biological atmospheric allergens or pathogens only as *Alternaria* spores. *J Environ Monitor* 12(11):2145–2152
- Astray G, Gullón B, Labidi J, Gullón P (2016a) Comparison between developed models using response surface methodology (RSM) and artificial neural networks (ANNs) with the purpose to optimize oligosaccharide mixtures production from sugar beet pulp. *Ind Crop Prod* 92:290–299
- Astray G, Fernández-González M, Rodríguez-Rajo FJ, López D, Mejuto JC (2016b) Airborne *Castanea* pollen forecasting model for ecological and allergological implementation. *Sci Total Environ* 548–549: 110–121
- Asturias JA, Gómez-Bayón N, Eseverri JL, Martínez A (2003) Par j1 and par j2, the major allergens from *Parietaria judaica* pollen, have similar immunoglobulin E epitopes. *Clin Exp Allergy* 33(4):518–524
- Aznarte JL, Benítez Sánchez JM, Lúgilde D, de Linares C, de la Guardia CD, Sánchez F (2007) Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. *Expert Syst Appl* 32(4):1218–1225
- Bas D, Boyaci IH (2007) Modeling and optimization II: comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction. *J Food Eng* 78(3): 846–854
- Beggs PJ (2004) Impacts of climate change on aeroallergens: Past and future. *Clin Exp Allergy* 34(10):1507–1513
- Belmonte J, Roure JM, March X (1998) Aerobiology of Vigo, North-Western Spain: atmospheric pollen spectrum and annual dynamics of the most important taxa, and their clinical importance for allergy. *Aerobiologia* 14(2–3):155–163
- Bousquet J, Hewitt B, Guerin B, Dhivert H, Michel FB (1986) Allergy in the Mediterranean area. II: cross-allergenicity among Urticaceae pollens (*Parietaria* and *Urtica*). *Clin Exp Allergy* 16(1):57–64
- Burbach GJ, Heinzerling LM, Edenharter G, Bachert C, Bindslev-Jensen C, Bonini S, Bousquet J, Bousquet-Rouanet L, Bousquet PJ, Bresciani M, Bruno A, Canonica GW, Darsow U, Demoly P, Durham S, Fokkens WJ, Giavi S, Gjomarkaj M, Gramiccioni C, Haahtela T, Kowalski ML, Magyar P, Muraközi G, Orosz M, Papadopoulos NG, Röhnelt C, Stingl G, Todo-Bom A, von Mutius E, Wiesner A, Wöhrl S, Zuberbier T (2009) GA(2)LEN skin test study II: clinical relevance of inhalant allergen sensitizations in Europe. *Allergy* 64(10):1507–1515
- Castellano-Méndez M, Aira MJ, 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. *Int J Biometeorol* 49(5):310–316
- Costa MA, Colombo P, Izzo V, Kennedy H, Venturella S, Cocchiara R, Geraci D (1994) cDNA cloning, expression and primary structure of par j1, a major allergen of *Parietaria judaica* pollen. *FEBS Lett* 341(2–3):182–186
- Cotos-Yañez T, Rodríguez-Rajo FJ, Jato V (2004) Short-term prediction of *Betula* airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models. *Int J Biometeorol* 48:179–185
- D’Amato G (2002) Urban air pollution and respiratory allergy. *Monaldi Arch Chest Dis* 57(2):136–140
- D’Amato G, Labetalol G (1989) Allergenic pollens in the southern Mediterranean area. *J Allergy Clin Immun* 83(1):116–122
- D’Amato G, Ruffilli A, Sacerdoti G, Bonini S (1992) *Parietaria* pollinosis: a review. *Allergy* 47(5):443–449
- Dahl A, Galán C, Hajkova L, Pauling A, Sikoparija B, Smith M, Vokou D (2013) The onset, course and intensity of the pollen season allergenic pollen: a review of the production, release, distribution and health impacts. Springer, Amsterdam, pp 29–70
- Desai KM, Survase SA, Saudagar PS, Lele SS, Singhal RS (2008) Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan. *Biochem Eng J* 41(3): 266–273
- Erdtman G (1952) Pollen morphology and plant taxonomy angiosperms (an introduction to palynology). In Almqvist and Wiksell (Ed), (pp 445–446) Stockholm, Sweden
- Ford SA, Baldo BA, Geraci D, Bass D (1986) Identification of *Parietaria judaica* pollen allergens. *Int Arch Allergy Appl Immunol* 79(2): 120–126
- Fotiou C, Damialis A, Krigas N, Halley JM, Vokou D (2011) *Parietaria judaica* flowering phenology, pollen production, viability and atmospheric circulation, and expansive ability in the urban environment: impacts of environmental factors. *Int J Biometeorol* 55(1):35–50
- Fuhrmann CM, Sugg MM, Konrad CE (2016) Airborne pollen characteristics and the influence of temperature and precipitation in Raleigh, North Carolina, USA (1999–2012). *Aerobiologia* 32(4): 683–696
- Galán C, Alcázar P, Cariñanos P, García H, Domínguez-Vilches E (2000) Meteorological factors affecting daily urticaceae pollen counts in Southwest Spain. *Int J Biometeorol* 43(4):191–195
- Galán C, Cariñanos P, Alcázar P, Domínguez E (2007) Spanish aerobiology network: management and quality manual. University of Córdoba, Córdoba
- Galán C, Ariatti A, Bonini M, Clot B, Crouzy B, Dahl A, Fernandez-González D, Frenguelli G, Gehrig R, Isard S, Levetin E, Li DW, Mandrioli P, Rogers CA, Thibaudon M, Sauliene I, Skjoth C, Smith M, Sofiev M (2017) Recommended terminology for aerobiological studies. *Aerobiologia* 33:293–295

- García-Mozo H, Galán C, Jato V, Belmonte J, Díaz De La Guardia C, Fernández D, Domínguez-Vilches E (2006) *Quercus* pollen season dynamics in the Iberian Peninsula: response to meteorological parameters and possible consequences of climate change. *Ann Agr Env Med* 13(2):209–224
- González Minero FJ, Candau P, Tomás C, Morales J (1997) Variación anual y estacional del polen de Urticaceae en el aire de Sevilla y su relación con los factores meteorológicos. *Polen* 8:60–77
- Govindaraju RS (2000) Artificial neural networks in hydrology I: preliminary concepts. *J Hydrol Eng* 5(2):115–123
- Grinn-Gofrón A, Strzelczak A (2008) Artificial neural network models of relationships between *Cladosporium* spores and meteorological factors in Szczecin (Poland). *Grana* 47(4):305–315
- Guardia R, Belmonte J (2004) Phenology and pollen production of *Parietaria judaica* L in Catalonia (NE Spain). *Grana* 43(1):57–64
- Hammer Ø, Harper DAT, Ryan PD (2001) Past: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 4(1):4–9. http://palaeoelectronica.org/2001_1/past/issue1_01.htm. Accessed 10 Feb 2018
- Harteveld DOC, Grant MR, Pscheidt JW, Peever TL (2017) Predicting Ascospore release of *Monilinia vaccinii-corymbosi* of blueberry with machine learning. *Ecol Epidemiol* 107(11):1364–1371
- Hirst JM (1952) An automatic volumetric spore trap. *Annals Appl Biol* 39(2):257–265
- Holgate ST, Jackson L, Watson HK, Ganderton MA (1988) Sensitivity to *Parietaria* pollen in the Southampton area as determined by skin-prick and RAST tests. *Clin Exp Allergy* 18(6):549–556
- Ianovici N, Panaitecu CB, Brudiu I (2013) Analysis of airborne allergenic pollen spectrum for 2009 in Timisoara, Romania. *Aerobiologia* 29(1):95–111
- Iglesias-Otero MA (2013) Bioinformática aplicada á aerobiología, Universidad de Vigo
- Kaufman HS (1990) *Parietaria*: an unrecognized cause of respiratory allergy in the United States. *Ann Allergy* 64(3):293–296
- Lee J, Lee K (2000) Use of hydrologic time series data for identification of recharge mechanism in a fractured bedrock aquifer system. *J Hydrol* 229(3–4):190–201
- Makris M, Koulouris S, Koti I, Aggelides X, Sideri K, Chliva C (2010) Temporal relationship of allergic rhinitis with asthma and other comorbidities in a Mediterranean country: a retrospective study in a tertiary reference allergy clinic. *Allergol Immunopathol* 38(5):246–253
- Masullo M, Mariotta S, Torrelli L, Graziani E, Anticoli S, Mannino F (1996) Respiratory allergy to *Parietaria* pollen in 348 subjects. *Allergol Immunopathol* 24(1):3–6
- Negrini AC, Voltolini S, Troise C, Aroba D (1992) Comparison between Urticaceae (*Parietaria*) pollen count and hay fever symptoms: assessment of a “threshold-value”. *Aerobiologia* 8:325–329
- Pérez OA, Gorgues RRM (1986) Alergia a la *Parietaria*. Bencard Publ, Beecham Pharm Res Div, Madrid
- Perpetuo EA, Silva DN, Avanzi IR, Gracioso LH, Baltazar MPG, Nascimento CAO (2012) Phenol biodegradation by a microbial consortium: application of artificial neural network (ANN) modelling. *Environ Tech* 33(15):1739–1745
- Portier CJ, Thigpen-Tart K, Carter SR, Dilworth CH, Grambsch AE (2010) A human health perspective on climate change: a report outlining the research needs on the human health effects of climate change. Environmental Health Perspectives and the National Institute of Environmental Health Sciences. https://www.niehs.nih.gov/health/materials/a_human_health_perspective_on_climate_change_full_report_508.pdf. Accessed 10 Feb 2018
- Puc M (2012) Artificial neural network model of the relationship between *Betula* pollen and meteorological factors in Szczecin (Poland). *Int J Biometeorol* 56(2):395–401
- Recio M, Rodríguez-Rajo FJ, Jato V, Trigo MM, Cabezero B (2009) The effect of recent climatic trends on Urticaceae pollination in two bioclimatically different areas in the Iberian Peninsula: Malaga and Vigo. *Clim Chang* 97(1):215–228
- Recio M, Docampo S, García-Sánchez J, Trigo M, Melgar M, Cabezero B (2010) Influence of temperature, rainfall and wind trends on grass pollination in Malaga (western Mediterranean coast). *Agric Forest Meteorol* 150(7–8):931–940
- Reyes ES, de la Cruz DR, Sánchez JS (2016) First fungal spore calendar of the middle-west of the Iberian Peninsula. *Aerobiologia* 32(3):529–539
- Ritenberga O, Sofiev M, Kirillova V, Kalina L, Genikhovich E (2016) Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. *Agric For Meteorol* 226–227:96–107
- Rivas-Martínez S (1987) Memoria del mapa de series de vegetación de España. Ministerio de Agricultura, Pesca y Alimentación, Madrid
- Rodríguez-Damián M, Cernadas E, Formella A, González A (2003) Automatic identification and classification of pollen of the Urticaceae family. Proceedings of Acivs (Advanced Concepts for Intelligent Vision Systems), Ghent, Belgium: 38–45 pp.
- Rodríguez-Rajo FJ, 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(2):101–110
- Rodríguez-Rajo FJ, Valencia-Barrera RM, Vega-Maray AM, Suárez FJ, Fernández-González D, Jato V (2006) Prediction of airborne *Alnus* pollen concentration by using ARIMA models. *Ann Agric Environ Med* 13(1):25–32
- Rodríguez-Rajo FJ, Astray G, Ferreiro-Lage JA, Aira MJ, Jato-Rodríguez MV, Mejuto JC (2010) Evaluation of atmospheric Poaceae pollen concentration using a neural network applied to a coastal Atlantic climate region. *Neural Netw* 23(3):419–425
- Sabo NC, Popovic A, Dordevic D (2015) Air pollution by pollen grains of anemophilous species: influence of chemical and meteorological parameters. *Water Air Soil Poll* 9:226–292
- Sabo NC, Kiš T, Janackovic P, Dordevic D, Popovic A (2016) Pollution by urticaceae pollen -influence of selected air pollutants and meteorological parameters. *Environ Sci Pollut R* 23(10):10072–10079
- Sahu P, López DL, Stoertz MW (2009) Using time series analysis of coal mine hydrographs to estimate mine storage, retention time, and mine-pool interconnection. *Mine Water Environ* 28(3):194–205
- Sánchez-Mesa JA, Galán C, Hervás-Martínez 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. *Int J Biometeorol* 49:355–362
- Serafini U (1957) Studies on hay fever: with special regard to pollinosis due to *Parietaria officinalis*. *Allergy* 11(1):3–27
- Sousa R, Ribeiro H, Cruz A, Duque L, Abreu I (2014) *Parietaria judaica* pollen: Aerobiology and allergenicity [Pólen de *parietaria judaica*: Aerobiologia e alergenidade]. *Rev Port Imunoalergol* 22(2):81–92
- Suárez-Cervera M, Vega-Maray A, Castells T, Rodríguez-Rajo FJ, Asturias JA, Le Thomas A, Seoane-Camba JA (2008) An approach to the knowledge of pollen and allergen diversity through lipid transfer protein localisation in taxonomically distant pollen grains. *Grana* 47(4):272–284
- Subiza FJ, Pola J, Feo Brito F, Moral de Gregorio AS (2007) Pólenes de interés en alergología en nuestro medio. In: Peláez A, Dávila IJ (eds) Tratado de alergología vol 1 Madrid, Spain: Ergon; 2007: 425–443. Ergon, Madrid, pp 425–443
- Tedeschini E, Rodríguez-Rajo FJ, Caramiello R, Jato V, Frenguelli G (2006) The influence of climate changes in *Platanus* spp. pollination in Spain and Italy. *Grana* 45(3):222–229
- Tu V (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 49(11):1225–1231
- Vega-Maray AM, Fernández-González D, Valencia-Barrera R, Polo F, Seoane-Camba JA, Suárez-Cervera M (2004) Lipid transfer proteins in *Parietaria judaica* L pollen grains: immunocytochemical localization and function. *Eur J Cell Biol* 83(9):493–497