ORIGINAL ARTICLES AAEM

Ann Agric Environ Med 2007, 14, 259-264

DEFINITION OF MAIN POLLEN SEASON USING A LOGISTIC MODEL

Helena Ribeiro[1, 2], Mário Cunha[1, 3], Ilda Abreu[2, 4]

[1]Autonomous Section of Agricultural Sciences, Faculty of Sciences, University of Porto, Vairão, Portugal
[2]Department of Botany, Faculty of Sciences, University of Porto, Porto, Portugal
[3]Centre for Investigation in Geo-Spatial Sciences, Faculty of Sciences, University of Porto, Porto, Portugal
[4]IBMC, Institute for Molecular and Cell Biology, Porto, Portugal

Ribeiro H, Cunha M, Abreu I: Definition of main pollen season using a logistic model. *Ann Agric Environ Med* 2007, **14**, 259-264.

Abstract: This paper proposes a method to unify the definition of the main pollen season based on statistical analysis. For this, an aerobiological study was carried out in Porto region (Portugal), from 2003-2005 using a 7-day Hirst-type volumetric spore trap. To define the main pollen season, a non-linear logistic regression model was fitted to the values of the accumulated sum of the daily airborne pollen concentration from several allergological species. An important feature of this method is that the main pollen season will be characterized by the model parameters calculated. These parameters are identifiable aspects of the flowering phenology, and determine not only the beginning and end of the main pollen season, but are also influenced by the meteorological conditions. The results obtained with the proposed methodology were also compared with two of the most used percentage methods. The logistic model fitted well with the sum of accumulated pollen. The explained variance was always higher than 97%, and the exponential part of the predicted curve was well adjusted to the time when higher atmospheric pollen concentration was sampled. The comparison between the different methods tested showed large divergence in the duration and end dates of the main pollen season of the studied species.

Address for correspondence: Ilda Abreu, Departamento de Botânica, Faculdade de Ciência U.P., Rua do Campo Alegre, 1191, 4150 Porto, Portugal.
E-mail: ianoronh@fc.up.pt

Key words: airborne pollen, logistic regression, main pollen season, pollen emission model.

INTRODUCTION

Pollen dispersal is an important natural event of large biological significance and its modelisation allows the understanding of its principles and processes.

The exact determination of the main pollen season is essential i) for forecasting models of crop production based on aerobiological data, to establish the effective pollination period and calculate the regional pollen index [1, 2]; ii) for pollen dispersion models, to identify the quantity of pollen coming from the studied region and which percentage may be attributed to deposition and recirculation; iii) for allergy purposes, to identify the time of year with highest concentrations of airborne allergenic pollen, and to assist with desirable preventive medical treatments [3].

Several methods to determine the MPS have been proposed [4, 5, 6, 7, 8]. Some of these use a given percentage of the annual total pollen and others establish threshold values above and below which the start and end respectively of the pollen season is delineated [9]. However, the methodology used to determine the MPS differ considerably between studies and is dependent upon the type of study and aims of each investigation [9].

Depending on the method used, sometimes a bias can be introduced when studying the influence of the meteorological parameters in pollen dispersion, due to the low concentrations found at the beginning and end of the MPS [7].

The characteristics of the MPS are also largely affected by atmosphere washing due to rain, long distance transport, or even ressuspension phenomena. As a result, changes in

Received: 22 January 2007
Accepted: 21 October 2007

the pollen counts and in the results of aerobiological studies are observed. In addition, if different criteria are used to define the MPS the comparison between studies becomes impossible.

The aim of our study is to establish a new method to define the MPS, based on statistical analysis and to permit a methodological unification. With this method the MPS will be characterized by some estimated parameters which may allow interregional comparisons and trend analysis.

MATERIAL AND METHODS

Study area. Our study was conducted in Porto region, northwest Portugal. Due to the maritime influence, the Porto region has mild temperatures and no really cold season experienced, January being the coldest month and July the hotest. Over the last 50 years, the thermal range has been very low with average minimum temperatures of 10±0.7ºC and an average maximum of 19±0.6ºC. The rainfall is irregularly distributed throughout the year, mainly concentrated in winter and spring.

Ornamental and non-ornamental trees, shrubs and herbaceous species can be found in the surroundings of the sampling site and several of these species are allergenic pollen producers.

Pollen sampling. Airborne pollen monitoring was continuously performed from January 2003–December 2005, using a 7-day Hirst-type volumetric spore trap (model Lanzoni VPPS-2000). The sampler was set on the roof of the Informatics Centre of the Faculty of Sciences in Porto (41º11' N, 8º39' W), approximately 20 m above ground level, and calibrated to sample air at 10 litres per minute. Pollen grains were trapped on a 7-day Melinex tape coated with silicone oil, which was then cut into daily segments. The slides, with the adhesive segments, were covered with fucsin-stained glycerol jelly and a cover glass. The daily mean concentration of the number of pollen grains was carried out using an optical microscope at a magnification of ×400 along 4 full lengthwise traverses.

Pollen counts were expressed as the sum of the average number per cubic meter of air for a 24-hour period. Pollen taxa selected corresponded to those which were considered allergological for the Portuguese Immunological and Allergological Society, and are *Acer*, *Alnus*, *Castanea*, Cupressaceae, *Fraxinus*, Pinaceae, *Platanus*, Poaceae and Urticaceae.

Pollen season definition. A pollen emission model (PEM) was established to define the beginning (x_i) and end (x_s) of the main pollen season (MPS). This model extends the approach proposed by Pathirane [7] of a sigmoid curve with 2 bends, one for the start and the other for the end of the MPS, and was reformulated by Cunha *et al.* [1] to identify the MPS of *Vitis vinifera* in Portugal.

A non-linear logistic regression model (a) was fitted to the values of the accumulated sum of daily airborne pollen

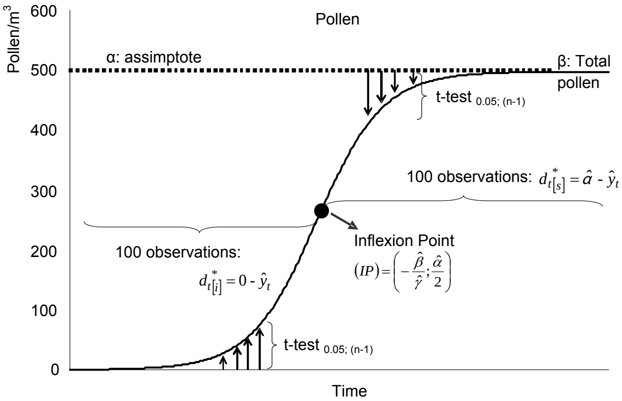

Figure 1. Adjusted Pollen Emission Model.

concentration along a year:

$$y_t = \alpha\left[1 + e^{-(\beta+\gamma,x)_t}\right]^{-1} + e_t;\ t = 1, \ldots, n, \qquad \text{(a)}$$

y_t is the amount of accumulated pollen up to day x_t; α is the distance between the 2 asymptotes, related with the total airborne pollen sampled during the year; β is a position parameter; γ is a parameter related with the pollen emission rate; t is the time: et is the error of the model (Fig. 1).

For *Alnus*, Cupressaceae and *Fraxinus* pollen, the flowering starts at the end of one year and is extended until the next year. Thus, the accumulated sum was calculated in 2004 and 2005 from the 1 September of the previous year until the 31 August of the following year. For all the other pollen types the time considered was from the 1 January until the 31 December.

The logistic model was fitted, applying the Levenberg-Marquardt method, and using the "regression non-linear" function of software SPSS 14.0.

After adjusting the model to each pollen type and year, a one-sided t-test modality was used at the 5% level in order to verify the thresholds where the daily difference between the pollen emission model and its superior asymptote $\left(d^*_{t_{[s]}} = \hat{\alpha} - \hat{y}_t\right)$ and inferior asymptote $\left(d^*_{t_{[i]}} = 0 - \hat{y}_t\right)$ were significant. This allowed to delimitation of the MPS, $d_{t_{[i]}}$ being the beginning date and $d_{t_{[s]}}$ the end date.

The d_t^* differences are a function of x, and the determination of the points where a significant difference is observed between the asymptote values and the adjusted function depends on the initial value and the number of observations considered. The day with the maximum pollen concentration was defined as the initial point, and 100 observations prior to and after this date, with a one day interval, was considered. The value 100 was chosen because no plant family has a flowering period longer than 200 days.

The first derivative of the model equation gives the value of an inflection point $(IP) = \left(-\frac{\hat{\beta}}{\hat{\gamma}}; \frac{\hat{\alpha}}{2}\right)$ that separates an interval at the rate of pollen increase with exponential type of an interval with decreasing increments of the pollen emission in the region.

A comparison was performed between the start and end dates and duration of the MPS achieved by the logistic

Table 1. Parameters estimates for main pollen season definition calculated using a non-linear logistic regression model.

	Parameters estimates				MPS (days)			P.I.		Meteorology		
Pollen	α	β	γ	R^2	x_i	x_s	D	x	y	rain	Tmean	Tsum
2003												
Pinaceae	288.1	-9.5	0.11	0.98	73	101	28	87	144	132.9	13.5	661.2
Poaceae	496.0	-7.5	0.04	0.99	141	195	54	168	248	74.5	19.6	1549.5
Urticaceae	1418.6	-3.3	0.02	0.99	80	194	114	137	709	343.6	13.2	1161.2
Acer	143.8	-15.1	0.18	0.99	73	93	20	83	72	52.6	13.6	491.2
Platanus	30.9	-13.4	0.11	0.99	109	137	28	123	15	113.6	14.6	761.6
Castanea	179.3	-24.2	0.14	0.99	162	186	24	174	90	48.6	19.5	858.0
Cupressaceae	182.1	-3.8	0.07	0.99	36	74	38	55	91	247.8	12.2	733.3
Alnus	57.0	-3.2	0.09	0.97	21	53	32	37	29	333.4	9.6	432.2
Fraxinus	28.1	-4.9	0.11	0.99	32	60	28	46	14	199.3	9.9	424.9
2004												
Pinaceae	1124.5	-6.8	0.07	0.99	75	111	36	93	562	188.3	11.8	733.5
Poaceae	591.5	-5.8	0.03	0.99	131	203	72	167	296	89.5	20.0	1722.5
Urticaceae	1696.5	-2.5	0.02	0.98	49	190	141	119	848	225.9	13.7	1298.5
Acer	505.2	-13.4	0.15	0.99	75	97	22	86	253	143.8	11.8	470.4
Platanus	630.0	-14.9	0.18	0.99	73	93	20	83	315	143.8	11.3	417.5
Castanea	185.3	-33.7	0.17	0.99	183	203	20	193	93	14.5	20.3	772.7
Cupressaceae	327.8	-9.4	0.05	0.99	27	73	46	172	164	200.4	10.3	753.8
Alnus	220.0	-22.7	0.15	0.99	18	41	23	152	110	109.6	10.7	439.6
Fraxinus	64.8	-24.7	0.20	0.99	357	10	18	123	32	82.4	10.8	366.0
2005												
Pinaceae	638.9	-5.7	0.05	0.98	82	128	46	105	319	137.6	14.4	678.9
Poaceae	500.0	-9.3	0.06	0.99	140	184	44	162	250	22.7	19.7	886.8
Urticaceae	1542.4	-4.5	0.04	0.99	82	144	62	113	771	173.8	14.6	922.8
Acer	243.6	-17.2	0.21	0.99	74	92	18	83	122	102.5	13.7	453.5
Platanus	907.0	-34.3	0.41	0.99	77	89	12	83	454	102.5	14.8	310.5
Castanea	102.9	-18.2	0.10	0.99	161	191	30	176	52	17.6	20.5	1109.5
Cupressaceae	532.1	-8.5	0.05	0.99	17	65	48	163	266	43.1	8.9	656.4
Alnus	258.7	-18.7	0.11	0.99	26	54	28	162	130	21.3	7.8	379.8
Fraxinus	46.3	-15.4	0.09	0.99	25	55	30	162	23	31.1	8.2	472.8

MPS – main pollen season; Tmean – mean temperature; Tsum – sum of mean temperature; P.I. – inflexion point; R^2 – determination coefficient

model, and by two other criteria, based on percentages (representing the accumulated sum of pollen between 5-95%–10-90% of the annual total pollen).

RESULTS

In this study, we obtained the parameters of the logistic model adjusted to each species and year, the determination coefficient (R^2) values, the inflection point, the $d_{t[i]}$ and $d_{t[s]}$ points referred in the methodology and necessary for the determination of the main pollen season (MPS) and some meteorological observations registered during this period for each species (Tab. 1).

The model fitting to the accumulated airborne pollen concentration (APC) was very good, with R^2 values always higher than 0.97.

During the 3-year study, no considerable interannual variation in the start date of the MPS of *Acer*, *Alnus*, Pinaceae and Poaceae were observed. The α parameter represents the total annual amount of pollen sampled for each pollen type. Interannual variations were observed, with the year 2003 registering the lowest airborne pollen concentration, with the exception of *Castanea*.

A highly negative significant relationship between the duration of the MPS, expressed by the D parameter, and the pollen emission rate, expressed by the γ parameter, was observed. Generally, years where high precipitation values and low mean temperatures were recorded during MPS are associated with longer D values. The opposite is observed for years with inverse conditions.

The graphical examples in Figure 2 show the annual atmosphere distribution of pollen types and the accumulated

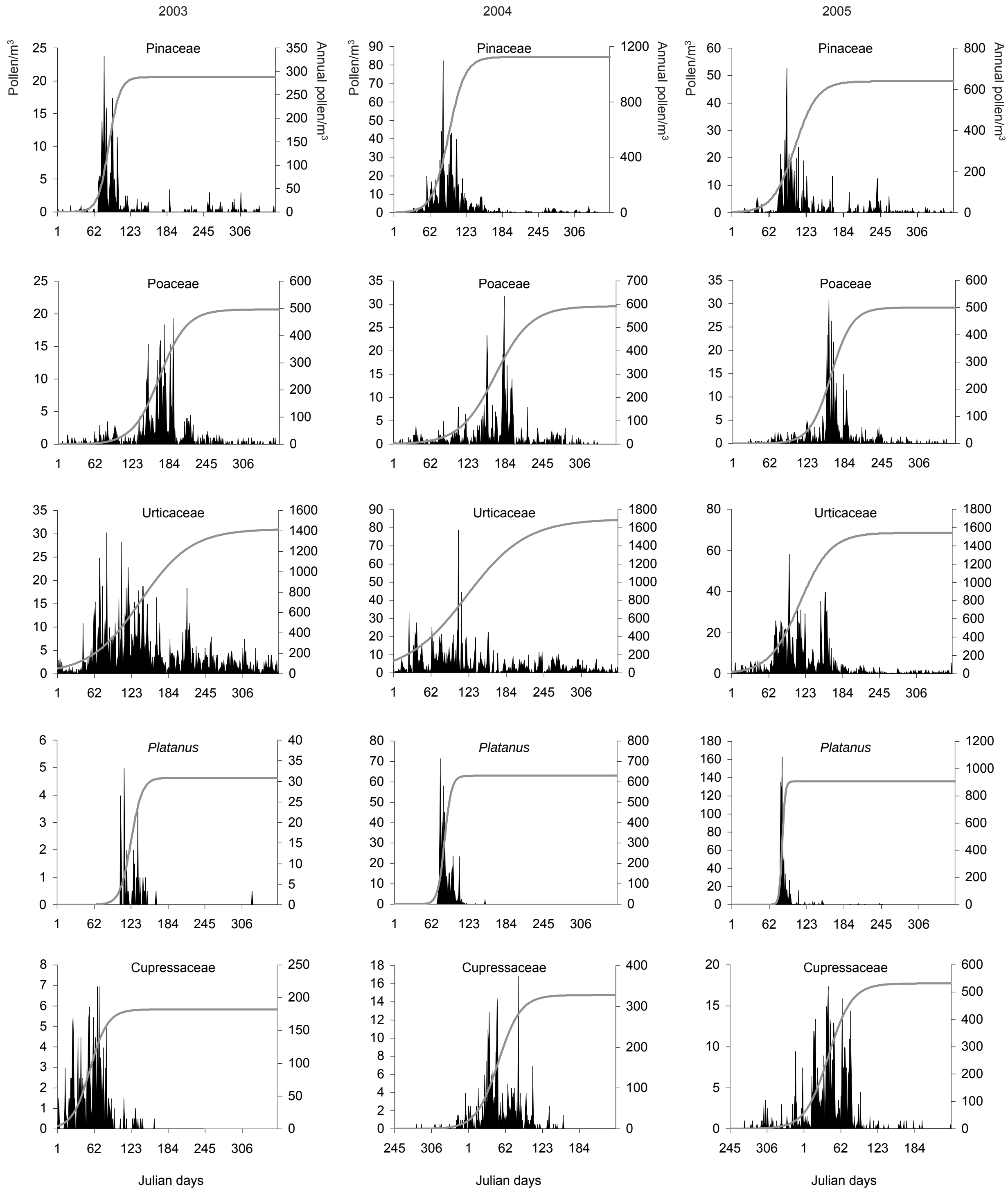

Figure 2. Daily pollen concentration and adjusted accumulated curve using the logistic model.

sums of pollen curve predicted by the logistic model. The exponential part of this curve, which corresponds to the highest pollen emission rates, is well adjusted to the time when higher atmospheric pollen concentration was sampled for all pollen types considered. This shows that the logistic model can be calculated for tree, shrub or herbaceous species, with the shape of the logistic curve adapting to the different interannual changes. For example, this is observed comparing the curves for *Platanus*, *Castanea*, Urticaceae or Cupressaceae pollen. In these cases, the variations in the exponential part of the logistic curve are adjusted to the highest concentrations of pollen in the atmosphere.

The comparison between the length and dates of the MPS defined by the logistic model and by the two percentage criteria tested is shown in Figure 3. In the case for the beginning dates of the MPS, the three methods achieved

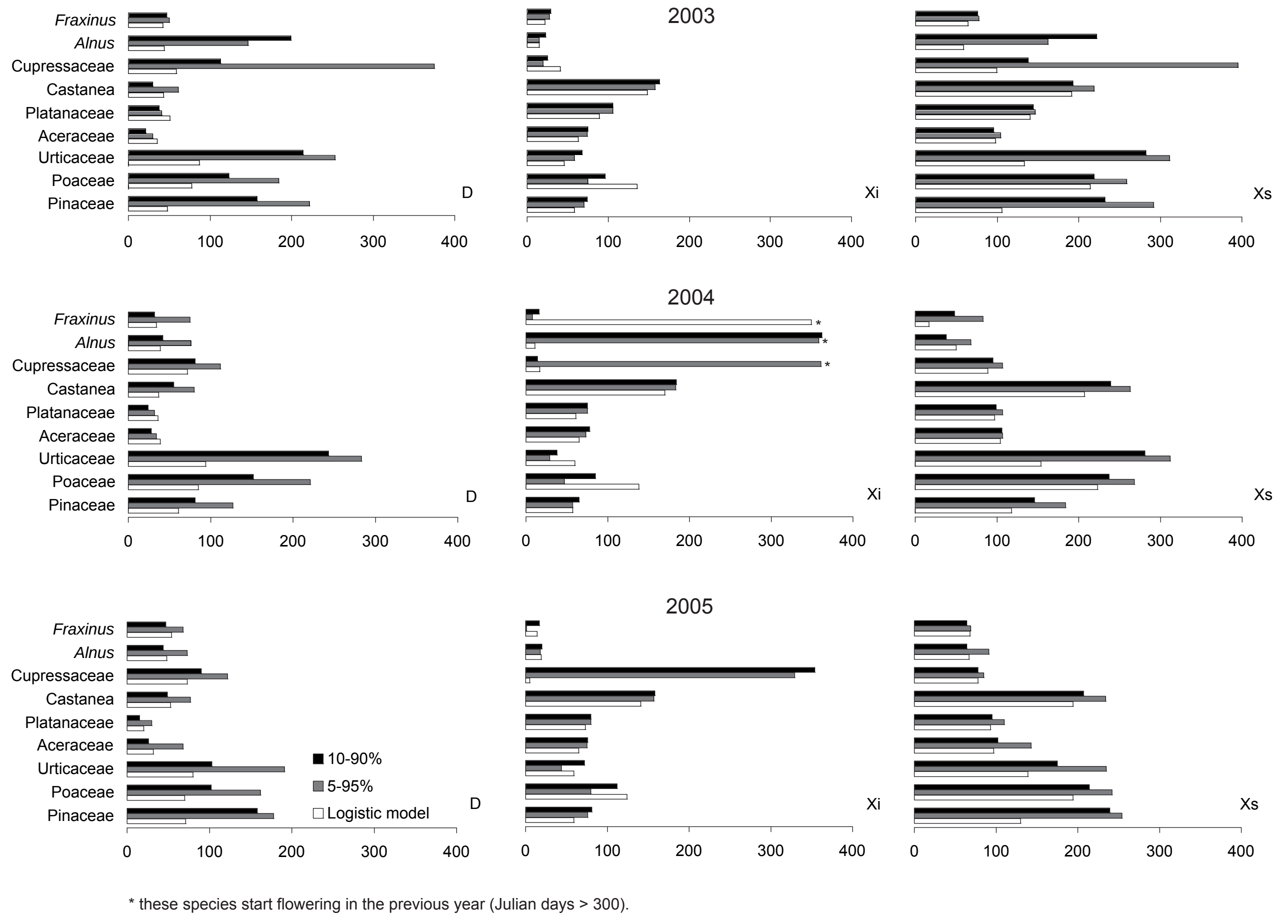

Figure 3. Comparison between duration (D), start (x_i) and end (x_s) dates of main pollen season for different species, using the logistic model and two methods based on percentage.

similar results when tested for the pollen from trees that flower in spring. However, the common tendency was for later beginning dates when the logistic method is used. Great differences for this date were observed for the pollen of herbaceous plants. Generally, the percentage methods gave longer MPS with longer tails for the end dates.

Regarding the extent of the MPS, the logistic method calculated always lower lengths compared with the percentage methods. The biggest variations were observed for herbaceous species, such as those belonging to Poaceae and Urticaceae.

DISCUSSION

The fitting between the values of the accumulated sum of daily airborne pollen concentration during a year and the curve calculated using the logistic model was almost perfect. This model was able to adjust well to the interannual variations observed, being the explained variance higher than 99% in 25 out of the 27 models estimated. Also, the overlap observed between the exponential part of the fitted curve and the timing of highest daily pollen concentration reveals a good fit. These features point out the explanatory capacity of the logistic model to describe and adapt to the diverse pollen emission pattern of trees, shrubs and herbs.

With the logistic model, the MPS of each species and year will be represented by the model parameters that can be interpreted as identifiable aspects of the flowering phenology and prevailing meteorological conditions. As a result, a better comparison between years and even regions can be accomplished. This is an advantage when compared with the methods based on percentage and on thresholds that only give the exact dates of the start and end of the MPS.

The pollen emission and its distribution are greatly influenced by the climatic conditions of the sampling site [10, 11]. The same happens for the model parameters since they describe the behaviour of the pollen season. In rainy weather, a reduced pollen emission rate (given by γ) is observed as well as pollen dispersion. Therefore, the pollen season is extended (given by D) and the amount of pollen reduced (related with α). If the temperatures are high during the blooming period then the number of open flowers with dehiscent anthers at the same time is higher and the MPS has a shorter duration (D). Therefore, the model parameters can be used to characterize the pollen dispersion pattern of each taxa according with the climatic characteristics of each season, such as rain occurrence or higher temperatures.

The α parameter is related with the total APC sampled and is also dependent on the height of the sampler, on the local species diversity and density, and on the climate. This parameter can be converted, dividing it by the duration of the MPS, giving the daily emission rate. This will allow performing the comparison between different sampling sites and, subsequently, between total airborne pollen concentrations in different regions. An extension in the MPS was observed when the percentage methods were used. This fact is probably related with pollen reflotation rather than pollen emission.

The methodology proposed in our study also allows calculation of the regional pollen index through the sum of the amount of pollen obtained during the MPS. This pollen index represents the effective value of the airborne pollen concentration (APC) corresponding to the period where the most pollen is sampled, without taking into consideration the pollen recirculation due to the wind. This is an important feature for the definition of the MPS. For instance, the Urticaceae pollen in the Porto region is present all year long in the atmosphere, but the flowering occurs during the spring-summer months, according to the species. Its constant presence in the atmosphere is related with the aerodynamic nature of this pollen type (small size 2-8 μm) that facilitates its resuspension. Consequently, the accumulated value of 90% or 95% of the annual total pollen concentration are reached at the end of the year, but the time of intensive pollen emission, with serious implication in allergenic sensitization, is observed much earlier.

Interannual variations were observed in the amount of pollen sampled, in the duration and at the start and end dates of the MPS for the different pollen types. According to one study [9], the method giving the shortest or longest MPS varied with the sampling site, the taxa, and even the year. This was also observed for the percentage criteria used in our study, where yearly differences in the duration, start and end dates were observed for several taxa, differing from the more constant pattern obtained when the logistic model was used.

Finally, the yearly variation between the start dates, duration of the MPS and the pollen production during a season, can be evaluated using the logistic model parameters. These parameters can be used to perform a statistical test in longer data sets (more than 15 years) in order to study the existence of trends in the pollen season and relate them with climatic change.

CONCLUSION

The logistic model proposed to define the characteristics of the main pollen season, independently of its duration and yearly variations, presented high adjustment, whatever the species. This model enables the definition of the main pollen season based on statistical analysis and makes possible a unification of the methodology for the main pollen season definition.

Also, the parameters estimated (α, D and γ) are closely related with the events during the main pollen season, such as sampling site, meteorological factors, annual pollen production and pre-flowering conditions, and therefore the interannual variations and long-run trends can be evaluated using these parameters.

Acknowledgements

The first author is grateful to Fundação para a Ciência e Tecnologia, SFRH/BD/13148/2003. This work was partially supported by a project from the Fundação Calouste Gulbenkian (ref.: 77161).

REFERENCES

1. Cunha M, Abreu I, Pinto P, Castro R: Airborne Pollen Samples for Early-Season Estimates of Wine Production in a Mediterranean Climate of Northern Portugal. *Am J Enol Vitic* 2003, **54(3)**, 189-194.
2. Orlandi F, Romano B, Fornaciari M: Effective pollination period estimation in olive (*Olea europaea* L.): a pollen monitoring application. *Sci Hortic* 2005, **105**, 313-318.
3. Rodríguez-Rajo FJ, Valencia-Barrera RM, Vega-Maray AM, Suárez FJ, Fernández-González D, Jato V: Prediction of airborne *Alnus* pollen concentration by using ARIMA models. *Ann Agric Environ Med* 2006, **13**, 25-32.
4. Galán C, Emberlin J, Domínguez E, Bryant RH, Villamandos F: A comparative analysis of daily variations in the Gramineae pollen counts at Córdoba, Spain and London, UK. *Grana* 1995, **34**, 189-198.
5. Nilsson S, Persson S: Tree pollen spectra in the Stockholm region (Sweden) 1973-1980. *Grana* 1981, **20**, 179-182.
6. Jäger S, Nilsson S, Berggren B, Pessi AM, Helander M, Ramfjord H: Trends of some airborne tree pollen in the Nordic countries and Austria, 1980-1993. A comparison between Stockholm, Trondheim, Turku and Vienna. *Grana* 1996, **35**, 171-178.
7. Pathirane L: Graphical determination of the main pollen season. *Pollen et Spores* 1975, **17**, 609-610.
8. Giorato M, Lorenzoni F, Bordin A, De Biasi G, Gemignani C, Schiappoli M, Marcer G: Airborne allergenic pollens in Padua: 1991-1996. *Aerobiologia* 2000, **16**, 453-454.
9. Jato V, Rodríguez-Rajo FJ, Alcázar P, De Nuntiis P, Galán C, Mandrioli P: May the definition of pollen season influence aerobiological results? *Aerobiologia* 2006, **22(1)**, 13-25.
10. Weryszko-Chmielewska E, Puc M, Piotrowska K: Effect of meteorological factors on Betula, Fraxinus and Quercus pollen concentrations in the atmosphere of Lublin and Szczecin. *Ann Agric Environ Med* 2006, **13**, 243-249.
11. García-Mozo H, Galán C, Jato V, Belmonte J, de la Guardia CD, Fernández D, Gutiérrez M, Aira MJ, Roure JM, Ruiz L, Trigo MM, Domínguez-Vilches E: Quercus pollen season dynamics in the Iberian Peninsula: response to meteorological parameters and possible consequences of climate change. *Ann Agric Environ Med* 2006, **13**, 209-224.