

Tomas R. Cotos-Yáñez · F. J. Rodríguez-Rajo ·
M. V. Jato

Short-term prediction of *Betula* airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models

Received: 2 December 2002 / Revised: 17 November 2003 / Accepted: 22 December 2003 / Published online: 10 February 2004
© ISB 2004

Abstract *Betula* pollen is a common cause of pollinosis in localities in NW Spain and between 13% and 60% of individuals who are immunosensitive to pollen grains respond positively to its allergens. It is important in the case of all such people to be able to predict pollen concentrations in advance. We therefore undertook an aerobiological study in the city of Vigo (Pontevedra, Spain) from 1995 to 2001, using a Hirst active-impact pollen trap (VPPS 2000) situated in the city centre. Vigo presents a temperate maritime climate with a mean annual temperature of 14.9 °C and 1,412 mm annual total precipitation. This paper analyses two ways of quantifying the prediction of pollen concentration: first by means of a generalized additive regression model with the object of predicting whether the series of interest exceeds a certain threshold; second using a partially linear model to obtain specific prediction values for pollen grains. Both models use a self-explicative part and another formed by exogenous meteorological factors. The models were tested with data from 2001 (year in which the total precipitation registered was almost twice the climatological average overall during the flowering period), which were not used in formulating the models. A highly satisfactory classification and good forecasting results were achieved with the first and second approaches respectively. The estimated line taking into account temperature and a calm S–SW wind, corresponds to the real line recorded during 2001, which gives us an idea of the proposed model's validity.

Keywords Generalized additive models · *Betula* · Kernel · Partially linear models · Pollen concentration

Introduction

Betula pollen is considered to be one of the main causes of pollinosis in northern and central Europe (Wallin et al. 1991; D'Amato and Spieksma 1992; Wihl et al. 1998). In localities in NW Spain between 13% and 60% of individuals who are immunosensitive to pollen grains respond positively to *Betula* pollen allergens (Aira et al. 2001; Dopazo 2001). There is no unique criterion for establishing the minimum level capable of triggering allergic symptoms since the response is highly individual. However, it has been pointed out that concentrations greater than 30 grains/m³ trigger severe symptoms and values greater than 80 grains/m³ produce allergic symptoms in 90% of patients (Viander and Koivikko 1978; Corsico 1993).

One of aerobiology's objectives is to develop statistical models enabling the short- and long-term prediction of atmospheric pollen concentrations. This would allow allergic individuals to take preventive measures to protect themselves from the severity of the pollen season. Pollen concentration predictions are usually long-term, trying to predict the onset and severity of the pollen season. Such models published recently have used, as prediction values, the sum of temperatures starting from a certain date (Clot 2001; Caramiello et al. 1994; Ruffaldi and Greffier 1991) or different phenological variables such as the chilling units and growing degree days required to trigger the phenological phase of flowering (Andersen 1991; Laaidi 2001; Jato et al. 2002). Another type of model aims at making short-term predictions of concentration, e.g. 1 or 2 days in advance, making use of statistical tools such as time series (Box-Jenkins type). The difficulty of modelling using time series lies in the very nature of the series, since it consists of zero pollen values throughout almost the entire year, interrupted by one or several random intervals, of short duration, of high

T. R. Cotos-Yáñez (✉)
Department of Statistics and Operation Research,
University of Vigo, Campus Universitario As Lagoas,
32004 – Ourense, Spain
e-mail: cotos@uvigo.es
Tel.: +34-988368766
Fax: +34-98836-8701

F. J. Rodríguez-Rajo · M. V. Jato
Department of Plant Biology and Soil Sciences, University of Vigo,
Campus Universitario As Lagoas, 32004 – Ourense, Spain

values with very fast fluctuations. Prediction based on the correlation of pollen levels with meteorological variables involves uni-dimensional non-linear equations using variables with the greatest prediction capacity (Norris-Hill 1995). In the case of *Betula*, the most frequently described models are based on linear parametric statistics (Atkinson and Larsson 1990; Norris-Hill and Emberlin 1991; Spieksma et al. 1995; Aira et al. 1998).

The objective of this study is the short-term prediction of *Betula* pollen levels, with a prediction horizon of 1–3 days starting from the last recorded observation. To that end we applied a nonparametric additive logistic model and a partially linear regression model, which estimate the probability of concentrations greater than 30 grains/m³ and predict future values of pollen grains respectively. Both methods (generalized additive models and semi-parametric regression models) are powerful and increasingly widespread statistical tools that enable the usual structural restrictions to be eliminated from the hypotheses on the model. Within the different nonparametric techniques, generalized additive models are of special importance because of the need to work with flexible multi-variant models that can be adapted to a wide variety of situations. Their main advantages are interpretability and flexibility.

Let (Z_l, Y_l) , $l = 0, \pm 1, \pm 2, \dots$, a $p + 1$ -dimensional time series, where Z_l is a p -dimensional series and Y_l a one-dimensional response series. Given a random sample $[(Z_i, Y_i)]_{i=1}^k$ and for $n = k + \kappa$, we want to predict Y_n given Z_n . Our prediction can be established by a logistic additive model

$$P(Y_l = 1|Z_l) = \text{Logit} \left[\alpha + \sum_{j=1}^p f_j(Z_{lj}) \right], \quad (1)$$

where $\text{Logit}(u) = e^u/(1 + e^u)$ is the link function and the unknown functions f_j are assumed to have mean 0. We applied this model to predict whether the concentration of pollen grains would exceed a certain level with a horizon of $\kappa = 1, 2$ and 3 days. Specifically we established a threshold of 30 grains/m³ since this is the minimum level triggering severe allergy symptoms. We therefore defined the binary response variable of interest, i.e. $Y_l = I\{X_{l+\kappa} > u\}$, $\kappa = 1, 2, 3$, being interested in $P(Y_l = 1|Z_l) = E\{Y_l = 1|Z_l\}$. The usual regression models with a binary solution are generalized linear models (McCullagh and Nelder 1989). These models are restrictive since they specify that the structure of the covariables is linear. Generalized additive models extend linear models by eliminating the linear structure from covariables, only requiring the addition of the covariables' effects.

Partially linear models (Härdle et al. 2000) deserve special attention because of their excellent behaviour in real situations (Prada-Sánchez et al. 2000). This is largely due to the fact that the nonparametric component makes the model highly flexible, enabling the inclusion of variables whose contribution is not completely known and/or in cases of heteroscedasticity. Thus, restructuring the starting series in vector $r + p + 1$ (Z_l, V_l, Y_l) , $l = 0, \pm 1,$

$\pm 2, \dots$, with r -dimensional V_l and supposing that the contribution of variables V_l is linear, results in the partially linear model

$$Y_l = V_l' \beta + \varphi(Z_l) + \varepsilon_l, \quad (2)$$

where the function φ is an unknown smooth function. We applied the previous model to the pollen series to obtain a specific prediction 1, 2 and 3 days in advance, defining the solution as $Y_l = X_{l+\kappa}$, $\kappa = 1, 2, 3$, $Z_l = (X_l, X_{l-1})$ with X_l being the pollen series and V_l the meteorological series.

Materials and methods

The city of Vigo is situated at sea level in the north-west corner of Spain on the shore of the Vigo Estuary (42°14'15''N; 8°43'30''W). Its climate is temperate maritime with a mean temperature of 18.8 °C, a mean minimum temperature of 11 °C and 1,412 mm of total annual precipitation.

Birch is represented in north-west Spain by one species, *Betula alba* L. (Moreno 1990). It is widely distributed and is the dominant tree in altimontane oro-Cantabrian acidophilic forests, with a clearly Euro-Siberian distribution. Such forests are found above a height of 1,150 m and are the last tree formations of the altitudinal series, with montane thermo-climates and hyper-humid ombro-climates (Costa et al. 1990). In this same area, but on siliceous soils with a greater Mediterranean influence, there are also birch forests in the Galician-Portuguese altimontane layer and the Ourense-Sanabria and supra-Mediterranean layer. In the Euro-Siberia region, birch may also form part of riparian forests, along with *Alnus glutinosa*, *Salix atrocinera* and *Frangula alnus*.

This study was based on the monitoring of atmospheric pollen concentrations from 1995 to 2001. Sampling was carried out by means of a 7-day Lanzoni pollen trap (Hirst 1952) placed on a terrace of the town hall 15 m above ground level. The standard sampling procedures proposed by Domínguez et al. (1992) were used to obtain the pollen counts. The meteorological data were supplied by the National Institute of Meteorology from the Vigo station located in the proximity of the pollen trap. Data from March, April and May 1995–2001 were used in the models' estimations and from the corresponding months in 2001 to evaluate the behaviour of the models.

The following group of variables was available for the construction of both models: the number of pollen grains (X_l) and meteorological variables (rainfall, humidity, temperatures, hours of sunshine, wind speed and direction). We broke the temporal dependence among the observations in order to process the series, considering them as observations independent of time, which enables processing from a regression point of view.

Additive logistic model

There are no standard criteria for selecting the variables included in the model. Our selection involved the combination of the most important variables affecting development, release and dispersion of pollen grains and the results obtained when they were included in the model. Therefore the following variables were chosen for the model (Eq. 1) with the objective of $P(Y_l = 1|Z_l) = E\{Y_l = 1|Z_l\}$, with $Y_l = I\{X_{l+\kappa} > u\}$, $\kappa = 1, 2, 3$, $u = 30$: T_l^{min} , T_l^{med} , daily minimum and mean temperature in l ; $T_l^{max} - T_{l-1}^{max}$, difference in daily maximum temperature between l and $l - 1$; t_l^{calm} , the daily mean time without wind in l ; the V_l^3 , daily mean S–SW wind velocity in l . We also added the variable of pollen grains at instant l and with a 1-day

delay (X_l, X_{l-1}) in order to distinguish between upward and downward phases. Thus, the model being considered is

$$P(Y_l = 1|\mathbf{Z}_l) = \text{Logit}[\alpha + f_1(X_l, X_{l-1}) + f_2(T_l^{\max} - T_{l-1}^{\max}) + f_3(T_l^{\min}) + f_4(T_l^{\text{med}}) + f_5(t_l^{\text{calm}}) + f_6(V_l^3)] \quad (3)$$

On the basis of observations and using kernel-type estimators in the estimations of the additive functions, we applied the local scoring algorithm (Hastie and Tibshirani 1990), initialising the additive functions starting from a linear logistic regression model, i.e. $f_1(u, v) = \beta_1 u + \beta_2 v$; $f_2(u, v) = \beta_3 u + \beta_4 v$; $f_j(u) = \beta_{j+2} u$, $j = 3, 4, 5, 6$. The convergence of the algorithm was attained when the adjusted relative variations of the additive functions were small.

Depending on the data point to be predicted we determined smoothing parameters (locally) using the cross-validation method for each variable at each internal step of the backfitting algorithm.

Partially linear model

We considered the data as a whole, as in the previous section, except that the solution response here is $Y_l = X_{l+\kappa}$, $\kappa = 1, 2, 3$, and the covariables included in the model are produced by the vectors:

$$\mathbf{V}_l = (T_l^{\min}, T_l^{\text{med}}, T_l^{\max} - T_{l-1}^{\max}, t_l^{\text{calm}}, V_l^3) \quad \text{and} \quad \mathbf{Z}_l = (X_{l-1}, X_l),$$

i.e. a nonparametric self-explicative part and a linear explanatory part formed by meteorological variables.

We followed the mechanism proposed by Speckman (1988) in the estimation of the model. If we subtract $E\{Y_l/\mathbf{Z}_l\}$ from both sides of Eq. (2)

$$Y_l - E\{Y_l/\mathbf{Z}_l\} = (\mathbf{V}_l - E\{\mathbf{V}_l/\mathbf{Z}_l\})^t \beta + \varepsilon_l,$$

we obtain a linear model relating Y_l and \mathbf{V}_l after adjustment for their expected values given \mathbf{Z}_l . Given the smoother matrix in the non-parametric context \mathbf{Z}_H , where $(\mathbf{Z}_H)_{ij} = w_j^{\mathbf{H},k}[\mathbf{Z}_i, (\mathbf{Z}_1, \dots, \mathbf{Z}_k)]$, $\{w_j^{\mathbf{H},k}\}$ is a set of kernel-generated weights and \mathbf{H} a square matrix, writing $\mathbf{Y} = (\mathbf{I} - \mathbf{Z}_H)\mathbf{Y}$ and $\tilde{\mathbf{V}} = (\mathbf{I} - \mathbf{Z}_H)\mathbf{V}$ we obtain

$$\hat{\beta} = (\tilde{\mathbf{V}}^t \tilde{\mathbf{V}})^{-1} \tilde{\mathbf{V}}^t \tilde{\mathbf{Y}}$$

and the predictor produced by

$$\hat{\phi}(\mathbf{V}_n, \mathbf{Z}_n) = \mathbf{V}_n^t \hat{\beta} + \sum_{j=1}^k w_j^{\mathbf{H},k}(\mathbf{Z}_n, (\mathbf{Z}_1, \dots, \mathbf{Z}_k))(Y_j - \mathbf{V}_j^t \hat{\beta}).$$

The smoothing parameters were selected using cross-validation, i.e. \mathbf{H} was selected so as to minimise the expression

$$\text{CV}(\mathbf{H}) = \sum_{i \in I} [y_i - \hat{\phi}^{-i}(\mathbf{V}_i, \mathbf{Z}_i)]^2 \tilde{w}(\mathbf{Z}_i)$$

where $\hat{\phi}^{-i}(\mathbf{V}_i, \mathbf{Z}_i)$ is the prediction produced previously without the i th point at that point and $\{\tilde{w}(\cdot)\}$ is a set of kernel-generated weights.

Results

Betula pollen is found in Vigo's atmosphere during 3 months in the summer and represents between 1% and 5% of the total annual pollen. The pollination period is short – less than 40 days – except in the last 2 sampling years, lasting 59 and 70 days respectively. It generally begins during March and ends during the first days in May, except in 1997 when it ended on 21 April. The highest concentrations are normally attained in mid-April, although the maximum peaks of 1997 and 1998 occurred earlier, on 28 and 27 March respectively.

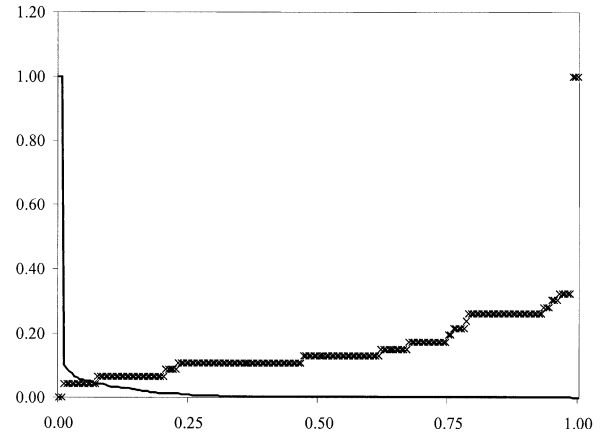


Fig. 1 Power function (solid line) versus type I error function (star)

There were also fluctuations in the maximum values during the years under study, with a minimum of 29 grains/m³ (27 March 1998) and a maximum of 247 grains/m³ (7 April 1999).

The wide variations in the *Betula* pollen concentrations make it difficult to obtain prediction models. We can illustrate the behaviour of the predictor of the model defined by Eq. 3 with $\kappa = 1$ with the design data in Fig. 1 (analogous figures were obtained for the cases $\kappa = 2$ and 3). We quantify the different types of error (type I error and type II error) that are made in the decision

$$\begin{cases} H_0: Y_n = 1|\mathbf{Z}_n \Leftrightarrow P(X_{n+1} \geq 30|\mathbf{Z}_n) \geq p \\ H_1: Y_n = 0|\mathbf{Z}_n \Leftrightarrow P(X_{n+1} \geq 30|\mathbf{Z}_n) < p \end{cases}$$

given p lies between 0 and 1. Thus, we define

$$\text{– Type I error} = \frac{\#\{(\hat{P}(Y_n = 1|\mathbf{Z}_n) < p) \cap (Y_n = 1)\}}{\#\{Y_n = 1\}},$$

i.e. the prediction at κ instants of the probability that the model will detect a level equal to or greater than 30 grains/m³ when this level was actually exceeded.

– Global type I error

$$= \frac{\#\{(\hat{P}(Y_n = 1|\mathbf{Z}_n) < p) \cap (Y_n = 1)\}}{k},$$

i.e. the prediction at κ instants of the probability that the model will detect a level equal to or greater than 30 grains/m³.

$$\text{– Type II error} = \frac{\#\{(\hat{P}(Y_n = 1|\mathbf{Z}_n) > p) \cap (Y_n = 0)\}}{\#\{Y_n = 0\}},$$

i.e. the prediction at κ instants of the probability that the model will detect a level greater than 30 pollen grains/m³ when this level was not actually exceeded.

– Global type II error

$$= \frac{\#\{(\hat{P}(Y_n = 1|\mathbf{Z}_n) > p) \cap (Y_n = 0)\}}{k},$$

i.e. the prediction at κ instants of the probability that the model will detect a level greater than 30 pollen grains/m³.

– Global error

$$= \frac{\#\{(\hat{P}(Y_n = 1|\mathbf{Z}_n) < p) \cap (Y_n = 1)\} + \#\{(\hat{P}(Y_n = 1|\mathbf{Z}_n) > p) \cap (Y_n = 0)\}}{k}$$

i.e. assigning equal importance to both types of error, the error of prediction at k instants in the decision.

It should be pointed out that all of the errors behave in a similar way in both the design data and prediction data

as a whole. The global error is 13.44%, 16.2%, 36.45 with $p = 0.5$ and for $\kappa = 1, 2$ and 3, respectively.

Choosing this as the cutting point gives us the list of errors during the 3 months of 2001, which are shown in Table 1.

The prediction errors are the same for $\kappa = 1$ and 2 because the limit of 30 grains/m³ was only exceeded on 10 days. Figure 2 shows the real pollen values during the months under study from 1995 to 2000 and the predictions 1, 2 and 3 days in advance that were obtained with the partially linear model of Eq. 2. It should be noted that

Table 1 Errors during the 3 months of 2001

Parameters	Type I error	Global type I error	Type II error	Global type II error	Global error
$\kappa = 1; p = 0.5$	0.10	0.011	0.012	0.011	0.112
$\kappa = 2; p = 0.5$	0.10	0.011	0.012	0.011	0.112
$\kappa = 3; p = 0.5$	0.20	0.022	0.024	0.022	0.224

Fig. 2 Daily pollen concentrations (area) for *Betula* from March to May, along with the predictions for 1 (thick solid line), 2 (fine solid line) and 3 (broken line) days ahead (1995–2000)

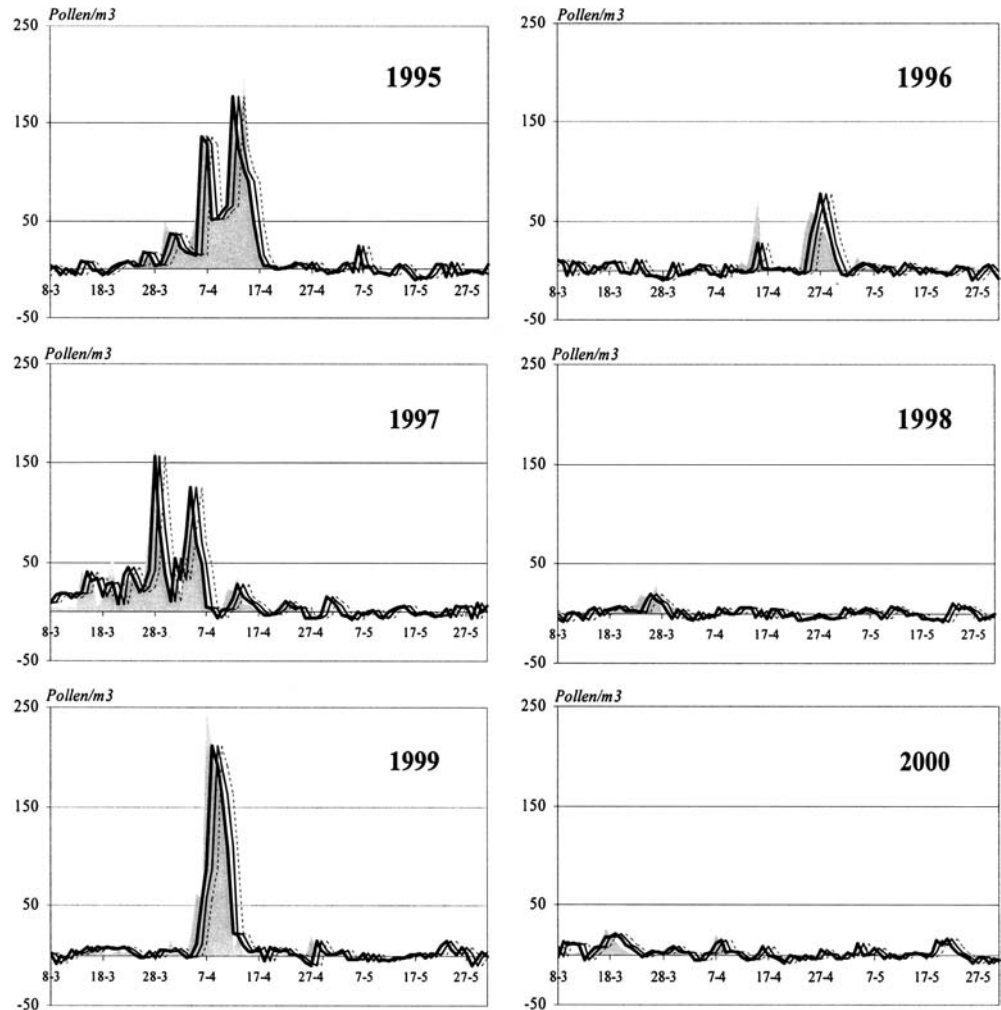
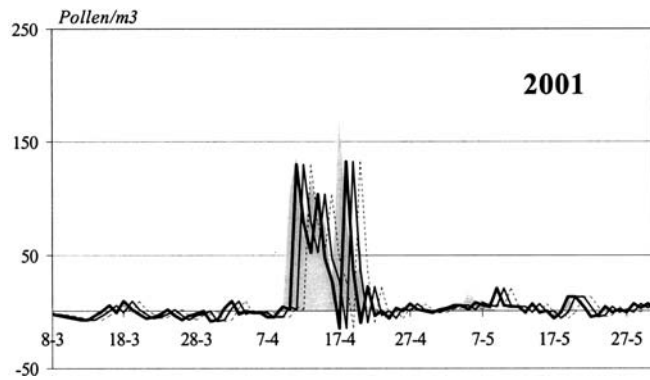


Table 2 Standard error according to year

Year	Means and SD of standard errors in pollen count (grains/m ³)					
	$\kappa = 1$		$\kappa = 2$		$\kappa = 3$	
	Average	SD	Average	SD	Average	SD
1995	186.63	1,136.49	617.29	2,289.12	943.51	2,888.95
1996	100.54	316.74	216.73	719.79	302.11	921.96
1997	104.55	292.41	585.38	2,002.22	917.43	2,944.04
1998	27.94	48.64	29.05	44.80	35.67	62.20
1999	380.54	2,766.19	899.40	4,403.93	1681.11	7,152.53
2000	30.91	51.36	44.39	73.25	55.44	89.03
2001	695.15	4,012.77	949.29	3,515.76	886.13	3,085.13

Table 3 Relative absolute error according to year

Year	Means and SD of relative absolute errors in pollen count (grains/m ³)					
	$\kappa = 1$		$\kappa = 2$		$\kappa = 3$	
	Average	SD	Average	SD	Average	SD
1995	0.92	1.47	1.35	1.70	1.64	2.59
1996	1.40	1.56	2.06	2.06	3.33	5.41
1997	1.83	2.68	2.70	3.14	2.97	3.96
1998	2.21	2.58	2.38	2.94	1.98	3.83
1999	2.36	2.54	3.37	4.72	3.95	6.17
2000	2.34	2.80	3.61	4.06	3.58	4.25
2001	2.64	3.84	2.45	2.67	2.60	3.53

**Fig. 3** Verification of the model during 2001. (These data were not used to formulate the model.) Daily pollen concentrations (area) for *Betula* from March to May, along with predictions for 1 (thick solid line), 2 (fine solid line) and 3 (broken line) days ahead

the real series is displaced so that the comparisons can be made on the same vertical line. Tables 2 and 3 quantify different types of mean errors per year:

$$SE = [y_t - \hat{\phi}(\mathbf{V}_{t-\kappa}, \mathbf{Z}_{t-\kappa})]^2, \quad \kappa = 1, 2, 3$$

$$\left| \frac{y_t - \hat{\phi}(\mathbf{V}_{t-\kappa}, \mathbf{Z}_{t-\kappa})}{y_t} \right| \quad \text{if } y_t \neq 0, \quad \kappa = 1, 2, 3$$

Finally, Fig. 3 shows the corresponding graph with data from 2001, values that were not taken into account when formulating the model. These values were used to check the validity and prediction capacity of the proposed model.

Discussion

The highest concentrations of *Betula* pollen in Spain are recorded in the north-west Iberian Peninsula (Jato et al. 1999). The values found in Vigo are similar to those described by Méndez (2000) for Ourense but lower than those found for Santiago de Compostela, where a total 24-h pollen count of 4,000 grains (annual average) is attained during some years (Aira et al. 1998). The concentrations of this pollen type are higher in the north of Europe and are produced during May (Atkinson and Larsson 1990; Spieksma et al. 1995).

Betula pollen values higher than 30 grains/m³, the quantity considered sufficient to trigger severe allergy symptoms, were attained during 16 days during years of high concentrations and on 7 days for low concentrations. There were never more than 6 days on which a count of 80 grains/m³ or more was reached; this is the value cited as sufficient to produce symptoms in 90% of patients (Viander and Koivikko, 1978; Detandt and Nolard 1996). In studies carried out in London, Norris-Hill and Emberlin (1991) indicate that these maximum values were generally attained on days with temperatures greater than 18 °C. In the city of Vigo, they coincided with temperatures higher than 20 °C.

During the study period, the annual average 24-h sum of *Betula* pollen fluctuated from year to year, oscillating between 1,540 in 1995 and 157 in 1998, describing biannual behaviour with an alternation of years in which the tree gives priority to reproduction and others in which it focuses on vegetative growth (Nilsson and Persson 1981; Emberlin et al. 1993; Jäger et al. 1991; El-Ghazaly et al. 1993; Spieksma et al. 1995; Latorre 1999). This biannual behaviour is also influenced by the meteorological conditions prevailing during the pollination period, since the intense flowering during 1997 coincided with limited precipitation in March and April. On the other hand, the concentrations recorded in 1996 and 1998 may be related to the low temperatures and especially the intense rainfall during the flowering period.

Linear logistic models are usually used in aerobiological studies aimed at predicting pollen concentrations. The regression lines proposed by this methodology, and that taking into account the meteorological values with the highest correlation coefficient in relation to pollen concentrations, explain up to 57% of the *Betula* pollen data in

our area (Rodríguez-Rajo 2000; Méndez 2000). In this study, the proposed additive logistic model for predicting *Betula* pollen concentrations starts from the regression model that is generally used in aerobiological studies. The additive logistic model's prediction behaviour is therefore superior to that described by these linear logistic models, since the final equation starts from such models and thereafter improves them.

The models for predicting pollen concentrations that use only meteorological variables as prediction variables produce 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, the pollen concentrations of previous days reflect these factors and therefore substantially improve the prediction capacity when included as a variable in prediction models. In this case, the prediction behaviour of the additive logistic model attained an acceptable error level during the months of 1995–2000 (Table 1). It is important to note that, despite the previously described biannual behaviour in which years of high and low concentrations alternate, the proposed model adapted perfectly to both cases, accurately describing the curve found in the behaviour of *Betula* pollen grains (Fig. 2).

In 2001 there were few peaks greater than 30 grains/m³, so that the values of the errors are merely informative. As in the case of previous years, the estimated curve accurately describes the behaviour of the *Betula* pollen grains. It is worth mentioning that the partially linear model behaved well, although as the prediction horizon increases we observe a time lag in the prediction series of the same longitude as the prediction horizon (see days 7–14 to 24–4 in Fig. 3). During the model-verification year of 2001 we found three concentration peaks. The first was detected by the proposed model, which predicted a mean concentration of 138 pollen grains/m³ air. The following two were also predicted with similar values although with a delay of 1 day, which is logical in view of the importance of the “previous day's pollen” variable as a predictor in the established model.

As previously discussed, this situation suggests that the currently available meteorological variables are not sufficiently explicative. It is the series itself that has to provide almost all of the information, thereby resulting in this delay.

Conclusions

In Vigo *Betula* pollen is recorded between March and May, attaining concentrations capable of triggering allergy symptoms. Meteorological parameters, along with the previous day's pollen concentrations, are good tools for establishing models predicting pollen concentrations, since they reflect the entire series of factors that affect the plant and on which its pollen production and release

depends. The integration of both variables substantially improves the percentage of variability explained by the regression lines.

The models were tested with data from 2001 and the predicted curve very closely followed the observed variations of the daily mean concentrations, in spite of the special meteorological conditions registered during this year. Even though temperatures were similar to the climatological average, precipitation values were almost double overall during the month of *Betula* flowering. It is worth highlighting the good prediction behaviour of the models used, especially the partially linear model. The results are better than those obtained by classical methodology, although their main inconvenience is their low prediction horizon, since the pollen data are not available until 24 h before the day for which a determination of the pollen concentration is required.

The model obtained would be applicable to different geographical areas, but its adjustment, using the available and adequate variables, would be necessary in every case.

One of the main goals in aeropalynological models is to predict concentration levels that could trigger allergic symptoms. The proposed model offers great reliability since, by predicting 1 day before, the probability that the model will detect a level equal to or greater than 30 pollen grains/m³ is higher than 90%. In the same way the possibility of raising a false alarm for the allergenic people will be 1.2%.

Acknowledgements This study was financed by the Xunta de Galicia's Environment Department (Project XUGA-PGDIT01-PAT38301PR). The authors would like to thank Javier Roca-Pardiñas for his useful comments on local-scoring algorithm programming.

References

- Aira MJ, Jato V, Iglésias I (1998) *Alnus* and *Betula* pollen content in the atmosphere of Santiago de Compostela. North-Western Spain (1993–1995). *Aerobiologia* 14:135–140
- Aira MJ, Ferreiro M, Iglesias I, Jato V, Marcos C, Varela S, Vidal C (2001) Aeropalynología de cuatro ciudades de Galicia y su incidencia sobre la sintomatología alérgica estacional. In: Moreno E, Renduelles B, Moreno J (eds) Proceedings of the XIII Simposio de la A.P.L.E. University of Cartagena, Cartagena, Spain, pp 105–114
- Andersen TB (1991) A model to predict the beginning of the pollen season. *Grana* 30:269–275
- Atkinson H, Larsson KA (1990) A 10 year record of the arboreal pollen in Stockholm, Sweden. *Grana* 29:229–237
- 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
- Clot B (2001) Airborne birch pollen in Neuchâtel (Switzerland): onset, peak and daily patterns. *Aerobiologia* 17:25–29
- Corsico R (1993) L'asthme allergique en Europe. In: Spieksma FTM, Nolard N, Frenguelli G, Van Moerbeke D (eds) Pollens de l'air en Europe. UCB, Braine-l'Alleud pp 19–29
- Costa M, Higuera J, Morla C (1990) Abedulares de la Sierra de San Mamed (Orense, España). *Acta Bot Malacitana* 15:253–265
- D'Amato G, Spieksma FT (1992) European allergenic pollen types. *Aerobiologia* 8:447–450

- Detandt M, Nolard N (1996) The Belgian Pollen Service: immediate and direct information to hay fever sufferers. *Aerobiologia* 12:197–200
- Domínguez E, Galán C, Villamandos F, Infante F (1992) Handling and evaluation of the data from the aerobiological sampling. *Monogr Rea* 1:1–18
- Dopazo A (2001) Variación estacional y modelos predictivos de polen y esporas aeroalergénicos en Santiago de Compostela. Tesis Doctoral, Universidad de Santiago de Compostela
- El-Ghazaly G, El-Ghazaly P, Larsson K, Nilsson S (1993) Comparison of airborne pollen grains in Huddinge and Stockholm, Sweden. *Aerobiología* 9:53–67
- Emberlin J, Savage M, Woodman R (1993) Annual variations in the concentrations of *Betula* pollen in the London area, 1961–1990. *Grana* 29:301–309
- Härdle W, Liang H, Gao J (2000) Partially linear models. Springer, Berlin Heidelberg New York
- Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman & Hall, London
- Hirst J (1952) An automatic volumetric spore-trap. *Ann Appl Biol* 36:257–265
- Jäger S, Spieksma FT, Nolard N (1991) Fluctuations and trends in airborne concentrations of some abundant pollen types, monitored at Vienna, Leiden and Brussels. *Grana* 30:309–312
- Jato V, Aira MJ, Iglésias MI, Alcázar P, Cervigón P, Fernández D, Recio M, Ruíz L, Sbai L (1999) Aeropalynology of birch (*Betula* sp.) in Spain. *Polen* 10:39–49
- Jato V, Rodríguez-Rajo FJ, Méndez J, Aira MJ (2002) Phenological Behaviour of *Quercus* in Ourense (N.W. Spain) and its relationship with the atmospheric pollen season. *Int J Biometeorol* 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
- Latorre F (1999) Differences between airborne pollen and flowering phenology of urban trees with reference to production, dispersal and interannual climate variability. *Aerobiologia* 15:131–141
- McCullagh P, Nelder I (1989) Generalized linear models. Chapman & Hall, London
- Méndez J (2000) Modelos de comportamiento estacional e intradiurno de los pólenes y esporas de la ciudad de Ourense y su relación con los parámetros meteorológicos. Tesis Doctoral. Universidad de Vigo
- Moreno G (1990) XLII Betulaceae. In: Castroviejo S (ed) *Flora Ibérica*, vol II. Real Jardín Botánico. C.S.I.C., Madrid, pp 37–49
- Nilsson S, Persson S (1981) Tree pollen spectra in the Stockholm region (Sweden), 1973–1980. *Grana* 20:179–182
- Norris-Hill J (1995) The modelling of daily Poaceae pollen concentrations. *Grana* 34:182–188
- Norris-Hill J, Emberlin J (1991) Diurnal variation of pollen concentration in the air of north-central London. *F Grana* 30:229–234
- Prada-Sánchez JM, Febrero-Bande M, Cotos-Yáñez TR, González-Manteiga W (2000) Prediction of SO₂ pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors. *Environmetrics* 11:209–226
- Rodríguez-Rajo FJ (2000) El polen como fuente de contaminación ambiental en la atmósfera de la ciudad de Vigo. Tesis Doctoral. Universidad de Vigo
- Ruffaldi P, Greffier F (1991) Birch (*Betula*) pollen incidence in France (1987–1990). *Grana* 30:248–254
- Speckman P (1988) Kernel smoothing in partial linear models. *J R Statist Soc B* 50:413–436
- Spieksma FT, Emberlin JC, Hjelmroos M, Jäger S, Leuschner RM (1995) Atmospheric birch (*Betula*) pollen in Europe: trends and fluctuations in annual quantities and the starting dates of the seasons. *Grana* 34:51–57
- Viander M, Koivikko A (1978) The seasonal symptoms of hyposensitized and untreated hay fever patients in relation to birch pollen counts: correlations with nasal sensitivity, prick tests and RAST. *Clin Allergy* 8:387–396
- Wallin JE, Segerström V, Rosenhall L, Bergmann E, Hjelmroos M (1991) Allergic symptoms caused by long distance transported birch pollen. *Grana* 30:256–268
- Wihl JA, Ipsen B, Nüchel PB, Munch EP, Janniche EP, Lövenstein H (1998) Immunotherapy with partially purified and standardized tree pollen extracts. *Allergy* 43:363–369