Int J Biometeorol (2011) 55:361–371
DOI 10.1007/s00484-010-0346-9

ORIGINAL PAPER

Forecasting ragweed pollen characteristics with nonparametric regression methods over the most polluted areas in Europe

László Makra · István Matyasovszky · Michel Thibaudon · Maira Bonini

Received: 12 March 2010 / Revised: 25 May 2010 / Accepted: 26 June 2010 / Published online: 13 July 2010

Abstract Nonparametric time-varying regression methods were developed to forecast daily ragweed pollen concentration, and the probability of the exceedance of a given concentration threshold 1 day ahead. Five-day and 10-day predictions of the start and end of the pollen season were also addressed with a nonparametric regression technique combining regression analysis with the method of temperature sum. Our methods were applied to three of the most polluted regions in Europe, namely Lyon (Rhône Valley, France), Legnano (Po River Plain, Italy) and Szeged (Great Plain, Hungary). For a 1-day prediction of both the daily pollen concentration and daily threshold exceedance, the order of these cities from the smallest to largest prediction errors was Legnano, Lyon, Szeged and Legnano, Szeged, Lyon, respectively. The most important predictor for each location was the pollen concentration of previous days. The second main predictor was precipitation for Lyon, and temperature for Legnano and Szeged. Wind speed should be considered for daily concentration at Legnano, and for daily pollen threshold exceedances at Lyon and Szeged. Prediction capabilities compared to the annual cycles for the start and end of the pollen season decreased from west to east. The order of the cities from the lowest to largest errors for the end of the pollen season was Lyon, Legnano, Szeged for both the 5- and 10-day predictions, while for the start of the pollen season the order was Legnano, Lyon, Szeged for 5-day predictions, and Legnano, Szeged, Lyon for 10-day predictions.

Keywords Daily concentration · Pollen threshold exceedance · Pollen season · Start date · End date

Introduction

The main plants that cause pollen allergy in Southern Europe are grasses (Poaceae), birch (Betulaceae), mugwort (*Artemisia* spp.) and olive-tree (Oleaceae) (D'Amato et al. 2007). In the 1980s a new species—the annual, wind-pollinated plant ragweed (*Ambrosia* spp.)—started to spread extremely aggressively. The blooming period of ragweed, which is appearing in more and more countries (Wopfner et al. 2005), lasts a long time (in some regions for up to 3 months) and it produces a lot of pollen (Béres et al. 2005). In Europe, regions most highly polluted with ragweed pollen are the southern part of European Russia (Juhász 1998), the Ukraine (Turos et al. 2009) and the Balkan Peninsula (Šikoparija et al. 2009). By the end of

L. Makra (✉)
Department of Climatology and Landscape Ecology,
University of Szeged,
P.O.B. 653, 6701 Szeged, Hungary
e-mail: makra@geo.u-szeged.hu

I. Matyasovszky
Department of Meteorology, Eötvös Loránd University,
Pázmány Péter st. 1/A,
1117 Budapest, Hungary
e-mail: matya@ludens.elte.hu

M. Thibaudon
RNSA (Aerobiology Network of France),
La Parličre 69610 Saint Genis l'Argentière, France
e-mail: michel.thibaudon@wanadoo.fr

M. Bonini
Department of Medical Prevention, ASL (Local Health Unit) of
the Province of Milan 1, Public Health Service,
via Spagliardi 19,
Parabiago 20015 Milan, Italy
e-mail: maira.bonini@aslmi1.mi.it

the twentieth century ragweed had spread north all over Poland (Kasprzyk 2008; Stepalska et al. 2008), reaching the northernmost point of Szczecin near the Baltic Sea (Puc 2006). However, the three main regions further south in Europe currently affected by *Ambrosia* are the Carpathian Basin (Chrenová et al. 2009; Ianovici and Sîrbu 2007; Peternel et al. 2006; Šikoparija et al. 2009; Štefanič et al. 2005), with peak values in Hungary (Makra et al. 2004, 2005, 2008), the Rhône-Alpes region (Laaidi et al. 2003) in France and the western part of the Po River Plain, mostly in the northwest part of the province of Milan (Bottero et al. 1990; Mandrioli et al. 1998). Bottero et al. (1990) first reported the spread and pollinosis of ragweed over this area of Lombardy.

Studying the characteristics of airborne pollen concentration in terms of meteorological parameters is of great practical importance, since a worldwide increase in respiratory diseases has been observed over the past few decades (Traidl-Hoffmann et al. 2003). A possible reason for increased pollen allergy levels might be global warming, as higher temperatures can raise substantially not only ragweed pollen levels (Wan et al. 2002) but pollen concentrations of other taxa as well (Teran et al. 2009; Frei and Gassner 2008) and induce longer pollen seasons (Wopfner et al. 2005; Teran et al. 2009; Frei and Gassner 2008). Increasing urbanisation, high levels of air pollution of transport origin, and a westernised lifestyle also contribute to an increase in the frequency of pollen-induced respiratory allergy, which is more prevalent in people who live in urban areas compared to those living in countryside (D’Amato and Cecchi 2008). Forecasting airborne pollen concentrations is one of the most studied fields in aerobiology due to its important applications in allergology. Tools commonly used for this task include auto regressive integrated moving average (ARIMA) modelling (Rodríguez-Rajo et al. 2005, 2006; Ocana-Peinado et al. 2008) and multiple regression analysis (Ribeiro et al. 2008; Stach et al. 2008; Rodríguez-Rajo et al. 2009). Other studies have used more advanced techniques such as neural network and neuro-fuzzy models (Castellano-Méndez et al. 2005; Sánchez Mesa et al. 2005; Aznarte et al. 2007; Rodríguez-Rajo et al. 2010). However, there is no evidence that these latter procedures perform better than traditional techniques (Aznarte et al. 2007; Verma and Pathak 2009).

For assessing allergic consequences, it is sufficient to predict the chance of exceeding a specified pollen concentration threshold that will cause severe health risks. Castellano-Méndez et al. (2005), for instance, focussed on the estimation of these risk levels.

The strong allergenicity and the rapid increase in prevalence of *Ambrosia* pollen as well as the extensive spread of respiratory diseases over the regions concerned has made it necessary to develop methods of forecasting the appearance of pollen for use by allergists and allergic sufferers. Using meteorological data, the aim of such forecasts is to predict the start of the pollen season (and also its end) early enough to allow sufferers to take preventive treatments (Laaidi 2001; Laaidi et al. 2003; García-Mozo et al. 2009). Two main approaches are used for this purpose, namely multiple regression analyses involving meteorological variables that influence ragweed development, and the method of summing temperatures above a base temperature for a specified period.

The aim of this study was to predict daily ragweed pollen concentrations and the probability of exceedance of a given pollen concentration threshold 1 day ahead. The prediction of the start and end of the pollen season at 5 and 10 days ahead is also discussed. Specifically, a threshold of 20 grains m^{-3} corresponding to the critical level for health risks (Jäger 1998) is considered. It should be noted, however, that this high threshold is applicable in areas suffering from very high daily ragweed pollen peak levels. This is the case for our study areas, whereas most European countries use a threshold value of 5–10 grains m^{-3} for ragweed pollen (e.g. Banken and Comtois 1992; Dechamp et al. 1997). In order to predict daily pollen concentrations and exceedance probabilities, nonparametric time-varying regression methods were developed. The prediction of the start and end of the pollen season was addressed with a nonparametric regression technique that combines regression analysis with the method of summing temperatures. Predicting ragweed pollen characteristics for such time intervals would make it easier to prepare for heavy pollen episodes and, in this way, prevents the development of serious respiratory diseases.

Predictions concerning ragweed have serious statistical difficulties. Pollen concentrations exhibit very strong annual cycles that have to be removed. However, for simplicity, just considering the daily temperature and ignoring other meterological variables, a given temperature (or temperature minus its annual cycle) may have a very different influence on pollen production (or pollen minus its annual cycle) in late September compared to in mid-August. Hence, the statistical model should be time-dependent, which complicates model building. Our methodology (introduced in the section on Statistical methods below) offers a natural way to address this issue. Normally, an available data set is divided into a learning set and a validation set. The learning set is used to estimate parameters of the statistical model, and this model is then applied to the validation set. However, the length of data sets is typically around 10 years, which is quite short; splitting this into two equal parts worsens the situation still further. A prime example is the start of the pollen season. For instance, Laaidi et al. (2003) used a 13-year learning set and a 2-year validation set. They fitted linear regression models with 3–5 predictors that required 4–6 regression parameters. Obviously, the estimation of 4–6 parameters with 13 items of data has a very high uncertainty, and validation with 2 items of data is practically unacceptable. As expected, the learning set provided prediction errors only of 0–3 days, while the validation set produced errors of 2–5 days. Hence, it is hard to evaluate the accuracy of predictions due to the small size of the data sets and the traditional methodologies used. Moreover, in order to assess the prediction abilities of any given technique, a comparison with some base estimate without any predictors is required. In general, previous studies have simply presented their findings, and assessments of the results relied on a certain degree of subjective judgement. For instance, the question of how well a regression technique can be addressed by its explained variance, given the big uncertainty due to the small amount of data. The approaches presented in the section Verification and validation are intended to provide a solution to these difficulties.

Data and methodology

Study areas and data

Six- to 10-year daily ragweed pollen data (from 1997 to 2006) for the period 1 June–31 October from three European cities, namely Lyon (Rhône Valley, France), Legnano (Po River Plain, Italy) and Szeged (Great Plain, Hungary) were used. These cities, representing the most polluted areas in Europe, differ in their topography and climate as well as in ragweed pollen characteristics. Pollen grains in all three cities were recorded using 7-day recording volumetric spore traps of the Hirst design (Hirst 1952).

Lyon (45.77 N; 4.83 E) lies in the Rhône-Alpes of France. The city is located in the Rhône valley with an elevation of 175 m a.s.l. at the confluence of the Rhône and Saône rivers (Fig. 1). Lyon has the second largest metropolitan area in France, with a population of 1.8 million in the urban area, and 4.4 million in the metropolitan area. In the Köppen system, its climate is of the Cbf type, i.e. it has a temperate oceanic climate with mild winters and cool-to-warm summers, as well as a uniform annual precipitation distribution (Köppen 1931). The trap was placed on the roof of a building in the centre of Lyon, approximately 25 m above ground level. In our study only 6 years of pollen data were used due to missing meteorological data in certain years.

Legnano, in contrast, is 33 km from the centre of Milan and was used to characterise the neighbourhood of Milan as the pollen data set of Milan has many missing data values. Milan (45.43 N; 9.28 E), located in the plains of Lombardy, is Italy's second largest city (Fig. 1). It has a population of about 1.3 million, while the population of the urban area is around 3.1 million. The Milan metropolitan area, by far the largest in Italy, is estimated to have a population of 7.4 million. The city is located 35 km north of the River Po, at an elevation of 122 m a.s.l. In the Köppen climate classification system, Milan is typically classified as a warm-temperate climate with a uniform annual precipitation distribution (Caf). Milan's winters are damp and cold, while its summers are often quite warm and humid (Köppen 1931). The trap was placed on top of a building about 17 m above ground level.

Szeged (46.25 N; 20.10 E), the largest settlement in south-eastern Hungary, is located at the confluence of the rivers Tisza and Maros (Fig. 1). The area is characterised by an extensive flat landscape of the Great Hungarian Plain with an elevation of 79 m a.s.l. The city is the centre of the Szeged region and has 203,000 inhabitants. In the Köppen system, the climate of Szeged is type Ca (warm, temperate climate), with relatively mild and short winters and hot summers (Köppen 1931). The air sampler is located on top of the building of the Faculty of Arts at the University of Szeged, about 20 m above the ground. Only 7 years of pollen data sets were used in our study because a lot of meteorological data for certain years was missing.

The daily mean temperature, daily precipitation amount and daily average wind speed from 1 April to 31 October for the period 1997–2006 were utilised as predictors. Temperature is the most important meteorological parameter influencing pollen counts, since it is the main reason for the increase in pollen concentration in the atmosphere. Wind helps disperse pollen grains in the atmosphere, while rainfall can wash the grains out of the atmosphere (Laaidi 1997). Applying simple statistical analyses, several studies have detected a significant positive correlation between daily ragweed pollen concentration and daily mean temperature (Bartkova-Scevkova 2003; Štefanič et al. 2005; Peternel et al. 2006; Puc 2006; Kasprzyk 2008), daily mean wind speed (Kasprzyk 2008), but a negative correlation with precipitation (Peternel et al. 2006; Kasprzyk 2008). The pollen concentration of the previous day is another useful predictor of the pollen level for the following day (Galán et al. 2001; Makra et al. 2004).

Statistical methods

Daily pollen concentrations are predicted using a method based on that of Cai (2007). Let the daily concentrations from 1 April to 31 October be denoted by y_i, $i=0,1,\ldots,n$ at times $t_{0,}$ $t_1,\ldots,t_n$. These latter values are scaled from 1 April

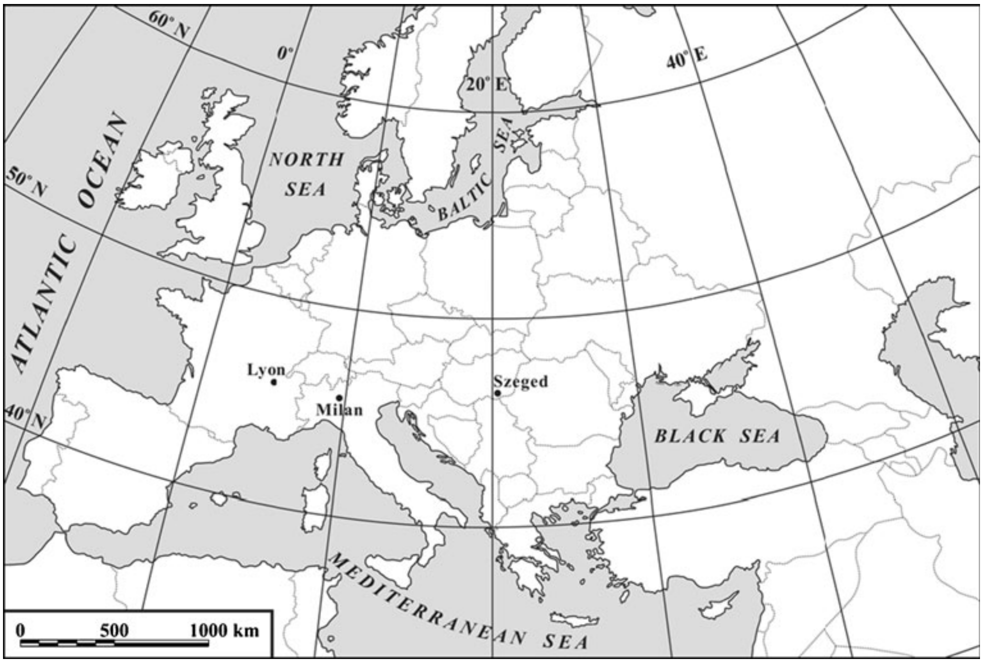

Fig. 1 Geographical locations of Lyon, Milan and Szeged

for each particular year, i.e. $t_{u+vM}=t_u$, u=0,...,M–1, v=1,...,N–1, where N is the number of years and M is the length of the period examined in each year. M for the three locations will be specified in the section Application below. Simultaneous values of p+1 number of predictors are written as $x_{ij}, i=0,1,\ldots,n, j=1,\ldots,p+1$. Our estimate $\hat{y}_i$ of y_i is defined as a time-varying linear regression

$$\hat{y}_i=\sum_{j=0}^{p}a_j(t_i)x_{ij}, \tag{1}$$

where x_{i0}=1, and precipitation (the (p+1)th predictor x_{ip+1}) is in fact omitted. The remaining predictors are the daily mean temperature, daily average wind speed, and pollen concentration of the previous day. As precipitation has temporal intermittency, including it in Eq. 1 is not straightforward. The problem is simplified by making the following reasonable assumption. A high precipitation level probably has essentially the same washout effect on pollen concentrations as even higher precipitation levels. Also, two small but different precipitation levels have again almost the same low washout effects. Hence, a critical precipitation amount c can be defined that distinguishes high and low effects on pollen concentrations. The (p+1)th predictor thus takes values of one or zero depending on whether the corresponding precipitation level exceeds c or not. Assuming that the regression coefficients in Eq. 1 vary "smoothly" in time, they may be approximated locally linearly according to Cai (2007). Applying the method of Li and Racine (2004) for binary predictors, the task is to minimise

$$\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{p}\left(\alpha_j+\beta_j(t_i-t_k)x_{ij}\right)\right)^2 K\left(\frac{t_i-t_k}{h}\right)L_r\left(x_{ip+1}-x_{kp+1}\right) \tag{2}$$

with respect to (w.r.t.) α_j and β_j, j=0,…,p for $1\le k\le n$, and $\hat{\alpha}_j(t_k)=\hat{\alpha}_j=\hat{\alpha}_j(k)$. K is the Epanechnikov kernel (weighting function) defined as $K(u)=3/4(1-u^2)$ for $-1<u<1$ and zero for u outside that range. Here $u=(t_i-t_k)/h$, where h is the window or bandwidth. The function $L_r(u)$ is a kernel function for binary variables satisfying $L_r(u)$=1 if u=0, and $L_r(u)$=r otherwise with $0\le r\le 1$. Note that the smaller the value of r, the larger the role of precipitation. The bandwidth h plays a crucial role in the accuracy of the procedure. Large bandwidths that allow large amounts of smoothing produce small variances with possibly large biases, while small bandwidths provide large variances with small biases. Thus, an optimal bandwidth that provides a trade-off between the bias and variance has to be estimated. The bandwidth h and r are estimated according to Cai (2007) in the approach introduced above. The determination of an optimal c value can be placed in the bandwidth choice task. It should be mentioned that the Epanechnikov kernel has an optimality property as it minimises the mean squared error of nonparametric regression estimates (Fan 1992).

The prediction of daily threshold exceedances was performed via the methodology outlined above. The only

Table 1 Comparing the prediction capabilities for $d=1$ day ahead with estimates obtained from the annual cycles using the root mean squared error (RMSE) and mean absolute error (MAE) for daily pollen concentration

City	Lyon		Legnano		Szeged	
Error (grain m^{-3})	RMSE	MAE	RMSE	MAE	RMSE	MAE
Prediction	36.3	13.3	34.1	13.3	73.0	26.6
Annual cycle	43.2	16.8	38.6	15.5	105.6	42.5

difference was that y_i, i=0,1,...,n took values of one or zero according to whether the corresponding pollen level exceeded the threshold or not. Hence an estimate $\hat{y}_i$ gives us the exceedance probability.

The methodology for predicting the start and end of the pollen seasons requires a few modifications. Here, a predictand y_i represents the normalised cumulative pollen concentrations (NCPCs) defined separately for each year. The normalisation factor is the total pollen number detected during the given year. Hence, the predictand varies from zero to one. Three predictors are used here (p=3), namely, the cumulated daily mean temperature x_{i1}, cumulative daily precipitation amount x_{i2} and cumulative pollen numbers x_{i3}, each calculated from 1 April to the actual day i within each year. For the d-day ahead prediction, the task is to minimise

$$\sum_{i=1}^{n}\left(y_i-\alpha_0-\sum_{j=1}^{p}\beta_j\left(x_{i-dj}-x_{k-dj}\right)\right)^2\bigcap_{j=1}^{p}K\left(\frac{x_{i-dj}-x_{k-dj}}{h_j}\right) \tag{3}$$

w.r.t. α_0 and β_j, j=1,…,p for specific values of k, and $\hat{y}_i=\hat{\alpha}_0=\hat{\alpha}_0(k)$. Here k takes values based on the given periods for the start and end of the pollen season. The period for the start is defined by the date of the earliest non-zero pollen level and the date of the latest NCPCs exceeding 5% during any of the years studied. Similarly, the period for the end is defined by dates corresponding to the earliest and latest NCPCs exceeding 95% and reaching 100%, respectively, during any of the years studied. Only the estimated NCPCs have a practical importance, but an explicit decision on the start/end date of a pollen season can be made by comparing these estimates with thresholds of 5% and 95%. The bandwidths in Eq. 3 are estimated by following Cai (2007), using to the approach introduced above.

Note that cumulated daily mean temperatures are usually calculated by summing temperatures above a base temperature for a specified period, and both the base temperature and the start period are estimated from the data. However, when comparing different methods for estimating the base temperature, it was found that by setting this value to zero, the null method performed surprisingly well (Snyder et al. 1999; Ruml et al. 2009). Also, estimates for the start of the above period are close to 1 April (Laaidi et al. 2003). Therefore our choice of applying the null method (no daily mean temperatures

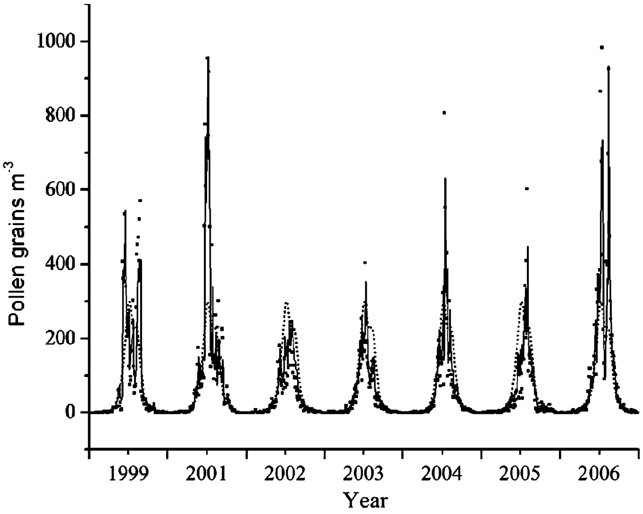

Fig. 2 Daily pollen concentration with 1-day prediction (*solid line*) and annual cycle (*dotted line*) for Szeged. The period examined in each year consists of 121 days

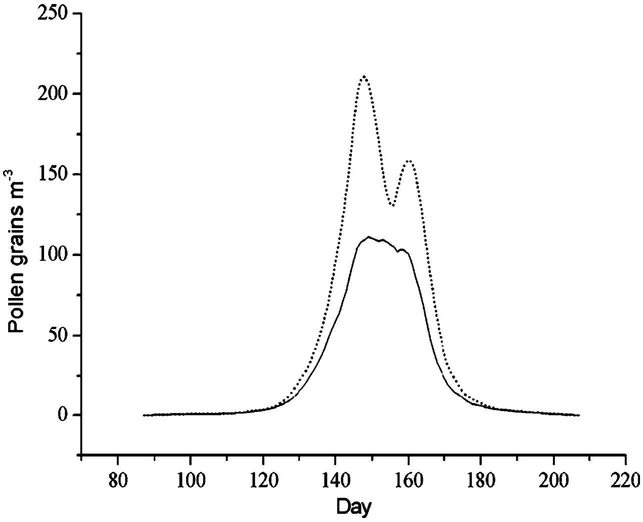

Fig. 3 Mean absolute error of estimates with 1-day prediction (*solid line*) and with annual cycle (*dotted line*) for Szeged. Values on the horizontal axis refer to days after 1 April

Table 2 Comparing the prediction capabilities for d=1 day ahead with estimates obtained from the annual cycles using the RMSE and MAE for daily threshold exceedance probability

City	Lyon		Legnano		Szeged	
Error	RMSE	MAE	RMSE	MAE	RMSE	MAE
Prediction	0.253	0.139	0.165	0.059	0.198	0.089
Annual cycle	0.294	0.179	0.184	0.072	0.330	0.109

below zero appeared in our data sets after 1 April) and choosing 1 April as the start date seem quite reasonable.

Verification and validation

The verification of the prediction capabilities of any technique requires a comparison with a base estimation without any predictors. This base estimate is the annual cycle of the ragweed pollen concentration estimated without any predictors. Namely, Eq. 1 is used with p=0 and the L_r term in Eq. 2 is omitted.

Our methodology requires the estimation of bandwidths. Having N years of data, the proposed methods and a modification of the standard bandwidth estimation make it possible to use an N-year validation set with an (N−1)-year learning set. Taking the lth year, the bandwidth is estimated with data that do not include the lth year, and estimates for this year are then obtained using this bandwidth. The procedure is applied for l=1,…N, thus these estimates for the entire data set can be validated directly.

Application

This section provides measures of the goodness-of-fit of the estimations validated in the manner outlined above.

Daily concentration

The period examined in each year covers days 155–281, 152–304, 177–297 defined from 1 January for Lyon, Legnano and Szeged, respectively. These dates are defined by the earliest and latest dates of days accompanied by non-zero pollen values in the years with available data.

The most important predictor for each location was the pollen concentration of the previous day. The role of precipitation varies, being of great importance in Lyon ($\hat{r} = 0.07$ at $\hat{c} = 0.4$ mm), but having no influence at all ($\hat{r} = 1$ for any c) in Szeged. Its role was also significant for Legnano ($\hat{r} = 0.54$ at $\hat{c} = 0.2$ mm). Temperature should be incorporated into the estimates for Legnano and Szeged (second in rank of importance), but Lyon does not require this element as a predictor. Among the predictors, wind speed has the weakest influence on pollen concentration and should be considered only for Legnano.

Predictor selection is based on minimising the root mean squared error (RMSE). Suppose we have a set of selected predictors and their bandwidths. When including further predictors in the estimation procedure, there is an obvious chance of getting higher optimal bandwidths

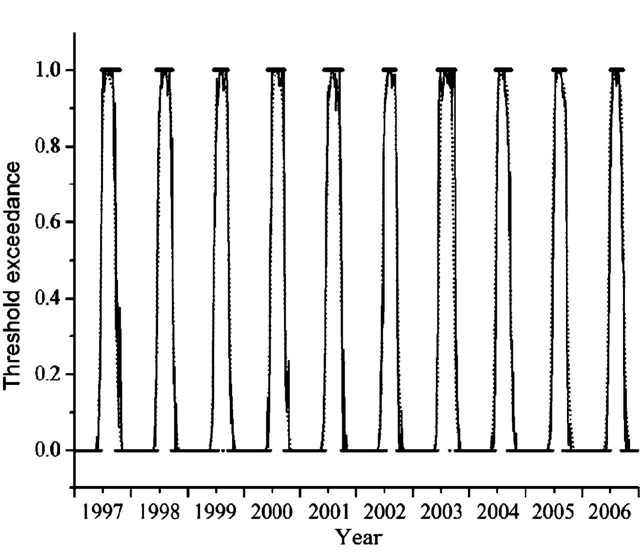

Fig. 4 Daily threshold (20 grains m^{-3}) exceedances with 1-day probability prediction (*solid line*) and annual cycle (*dotted line*) for Legnano. The period examined in each year consists of 153 days

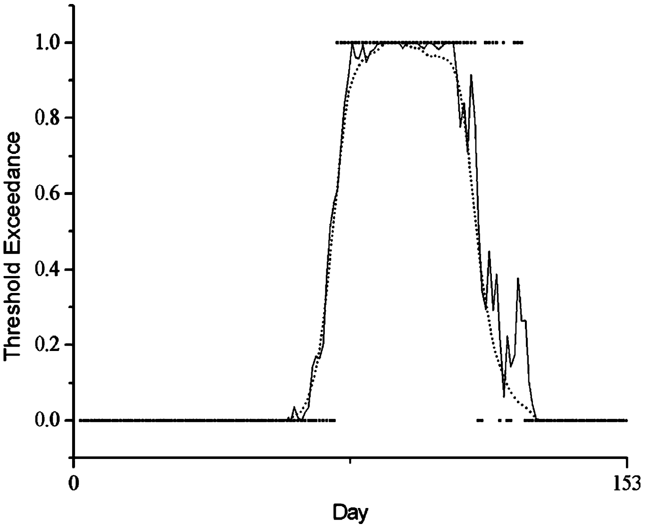

Fig. 5 Daily threshold (20 grains m^{-3}) exceedances with 1-day probability prediction (*solid line*) and annual cycle (*dotted line*) for Legnano for 1997. The value of 153 on the horizontal axis refers to the length of the period examined each year

Table 3 Comparing the prediction capabilities for d=5 days ahead with estimates obtained from annual cycles using the MAE and absolute error range (AER) for the start and end of the pollen season

City	Lyon		Legnano		Szeged	
Error (days)	MAE	AER	MAE	AER	MAE	AER
Start, prediction	2.2	1-4	1.3	0-3	2.9	0-5
Start, annual cycle	3.0	0-5	4.2	1-11	2.6	0-5
End, prediction	2.3	0-6	4.6	1-8	2.4	0-6
End, annual cycle	3.0	1-6	2.0	0-6	1.0	0-3

because higher dimensional prediction surfaces can be represented by larger amounts of data produced with larger bandwidths. However, larger bandwidths yield larger biases of the estimates obtained with the predictors selected above. Thus RMSE values for a larger number of predictors depend on whether the newly added predictors have enough information on the predictand to produce a variance reduction larger than the bias increase of the estimates. Hence the optimal configuration of predictors minimises the RMSE.

Table 1 summarises our results obtained for the three cities. Lyon and Legnano have very similar estimation errors measured either with RMSE or mean absolute error (MAE). The smallest RMSE is observed for Legnano, the lowest MAE is for Lyon and Legnano, while Szeged exhibits errors that are about twice as large. However, the performance of relative RMSE (RRMSE) or relative MAE (RMAE) of the prediction compared to RMSE and MAE obtained just with the annual cycle is best for Szeged and poorest for Legnano. It seems, therefore, that the larger the variance of the pollen concentration, the larger the relative variance reduction produced by the estimation procedure. Figure 2 illustrates the prediction potential of our methodology for Szeged having the largest RMSE (and MAE). Evidently, prediction errors exhibit strong annual cycles, with largest errors at largest concentrations, but these errors are substantially smaller than those obtained from estimates with the annual cycle of pollen concentrations (Fig. 3).

Daily threshold exceedance

In this section, the periods studied were the same as those studied for daily concentrations. The most important predictor for each location was the pollen concentration of the previous day. The role of the other variables was rather different. Precipitation was of great importance in Lyon ($\hat{r} = 0.07$ at $\hat{c} = 0.4\,\mathrm{mm}$), but had no influence at all ($\hat{r} = 1$ for any c) in Legnano and Szeged. Temperature should be included in estimates for Legnano and Szeged (second in order of importance), but Lyon did not require this element as a predictor. Wind speed, with the weakest influence on threshold exceedance among the predictors, should be considered only for Lyon and Szeged.

Table 2 summarises our results obtained for the three cities. When considering RMSE or MAE, it appears that the prediction performance is highest for Legnano and lowest for Lyon. However, RRMSE tells us that Szeged provides the best estimate and Legnano the weakest estimate. RMAE leads to a slightly different conclusion because Lyon and Szeged change their order compared to that obtained for RRMSE. Hence, the larger the variance of the pollen concentration exceedances, the larger the relative variance reduction produced by the estimation procedure. Figure 4 illustrates the prediction potential of our methodology for Legnano, which has the highest RRMSE of the cities examined here. In order to provide a better temporal resolution, Fig. 5 shows just the year 1997.

Pollen season

The pollen season is defined by its start and end dates. These dates have to be defined precisely and calculated for the years with available data. Several different criteria for determining the pollen season can be found in the literature (Cassagne 2009). For instance, Lejoly-Gabriel (1978) observed that plotting cumulated pollen amounts

Table 4 Comparing the prediction capabilities for d=10 days ahead with estimates obtained from the annual cycles using the MAE and AER for the start and end of the pollen season

City	Lyon		Legnano		Szeged	
Error (days)	MAE	AER	MAE	AER	MAE	AER
Start, prediction	4.7	1-8	2.3	0-5	2.7	1-5
Start, annual cycle	3.0	0-5	4.2	1-11	2.6	0-5
End, prediction	2.7	1-6	4.3	1-9	3.7	1-5
End, annual cycle	3.0	1-6	2.0	0-6	1.0	0-3

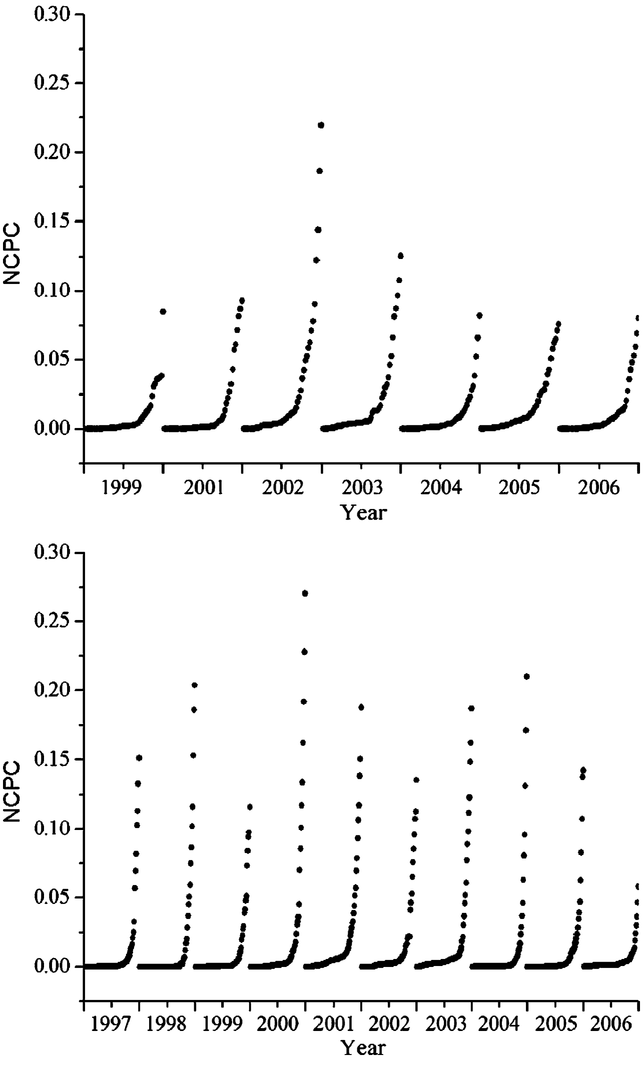

Fig. 6 Normalised cumulative pollen concentration (NCPC) for Szeged (*top*) and Legnano (*bottom*) for the start of the pollen season

against dates yields a sigmoid curve with two bends that may correspond to the start and end of the pollen season. Another approach says that the start of the season is the date on which at least one pollen grain m^{-3} is recorded and all subsequent days (Sánchez Mesa et al. 2005), or at least 5 consecutive days (Galán et al. 2001) with one or more pollen grains m^{-3}. According to García-Mozo et al. (2009), the start date is the 1st day on which at least five pollen grains m^{-3} are recorded, with the 3 subsequent days having five or more pollen grains m^{-3}. Furthermore, it can be seen that typical annual cycles of daily pollen concentrations indicate two inflection points that correspond to the start and end of the pollen season (Laaidi 2001).

As any definition has its apparent subjective elements, in our study the two dates are defined as in the paper by Nilsson and Persson (1981). These authors took into account 90% of the total annual pollen counts after excluding the initial 5% and the final 5%. Therefore, the period value for the method described in section on Statistical methods is defined by the date of earliest non-zero pollen level and the date of latest NCPCs exceeding 5% during any of the years studied for the start. Similarly, the period for the end is defined by dates corresponding to the day before the earliest NCPCs reaching 95% and the latest NCPCs reaching 100% during any of the years studied. Thus, the period examined for each year covers days 155–231, 152–233, 177–227 for the start of the pollen season and days 255–281, 255–304, 257–297 for the end of the pollen season defined from 1 January for Lyon, Legnano and Szeged, respectively.

The most important predictor for both the start and end of the pollen season for each city is the cumulated daily mean temperature. The cumulative daily precipitation plays a smaller role in each case, while the predictor having the smallest significance is the cumulative pollen value. This latter variable should be taken into account only for the end of the pollen season with infinite optimal bandwidths. Such bandwidths indicate (global) linear relationships between the end of the pollen season and the cumulated pollen values observed during the previous period.

The order of the cities from the smallest to the largest MAE of the prediction for the end of the pollen season was Lyon, Legnano and Szeged for both d=5 and d=10 days, for the given periods and cities when the NCPCs were calculated. The order of the cities from the smallest to the largest MAE of the prediction for the start of the pollen season was Legnano, Lyon and Szeged for d=5, for the given periods when the NCPCs were calculated. For d=10, Lyon and Szeged changed their order compared to the latter

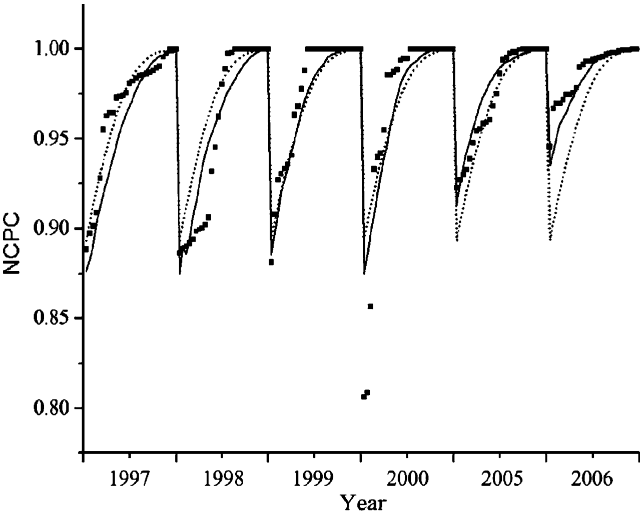

Fig. 7 Normalised cumulative pollen concentration (NCPC) with 10-day prediction (*solid line*) and annual cycle (*dotted line*) for Lyon for the end of the pollen season

list. Tables 3 and 4 show that the proposed prediction method does not provide any improvement over estimation based on the annual cycle for Szeged. This is because the start and end of the pollen season appear quite regularly here. To illustrate this fact, Fig. 6 shows the NCPC for the start of the pollen seasons at Szeged and Legnano. Note that the end of the pollen season appears even more regularly at Szeged; see the MAE for the end of the pollen season in Table 3. For example, the NCPC with 10-day prediction and annual cycle for Lyon for the end of the pollen season is shown in Fig. 7. Quite surprisingly, the prediction for Szeged at $d=10$ seems more accurate than at $d=5$. This may be due to statistical uncertainties because the MAE of predictions with these two different time steps is identical when the validation part is omitted and bandwidths are estimated using the entire data sets. This latter finding, however, highlights the very regular development of the pollen season in Szeged.

Conclusions

For a 1-day prediction of both the daily pollen concentration and daily threshold exceedance, the order of the cities from the smallest to the largest prediction errors was Legnano, Lyon, Szeged and Legnano, Szeged, and Lyon, respectively. However, prediction errors relative to the annual cycles were smallest for Szeged in both cases. Thus the larger the variance of the pollen concentration, the larger the relative variance reduction produced by the estimation procedure. For each location, the most important predictor is the pollen concentration of the previous day. The second main predictor was precipitation for Lyon and temperature for Legnano and Szeged. Why wind speed should have the smallest significance is unclear; this predictor should be considered only for daily concentrations at Legnano and for daily pollen threshold exceedances at Lyon and Szeged.

The prediction capabilities of the method compared to the annual cycles for the start and end of the pollen season decrease from west to east. This decrease is so clear-cut that it is worthwhile, say, making the prediction using just the average annual cycle for Szeged. This is because the annual cycle of Szeged yields a very good result (the individual years are quite similar for the pollination season), i.e. the MAE obtained for the end of the pollination season is only 1 day (Table 3). This is scarcely exceeded by any other prediction schemes. One reason for this may be that ragweed thrives best in environmental conditions similar to its place of origin, namely, the prairies of the United States with their warm and dry climate and poor and loamy or sandy soils, which favour the growth, development and spread of this plant. Comparing the three habitats examined, the climate and soil conditions of the Great Plain in Hungary, represented by Szeged, best fit the above conditions. That is, while the climates of Lyon and Legnano are warm and humid, Szeged is characterised by a warm and dry period (liable to occasional droughts) with abundant sunshine hours, low relative humidity and cloudiness during the summer and early autumn—the pollination season of ragweed.

The order of the cities from the smallest to the largest MAE of the prediction for the end of the pollen season is Lyon, Legnano and Szeged for both $d=5$ and $d=10$ days. Similarly, the order of the cities for the start of the pollen season is Legnano, Lyon and Szeged for $d=5$, and Legnano, Szeged and Lyon for $d=10$.

A reviewer missed an analysis for selecting the temperature threshold and starting date for calculating cumulated daily mean temperatures used as a predictor for the estimation of the start of the pollen season. Therefore, an RMSE criterion (e.g. Ruml et al. 2009) was used to estimate these two quantities. We found the optimal temperature threshold to be 0°C for all three locations. (More exactly, any threshold not exceeding the lowest observed daily mean temperature after 1 April is applicable.) So the null method used in the paper is appropriate. The RMSE was more sensitive to the starting date when calculating cumulated daily mean temperatures, at least for Legnano. However, no improvement was achieved by repeating every calculation for the start of the pollen season with new predictors defined by optimal dates, as both the MAE and AER were identical with MAE and AER values provided in the paper for all three locations. Hence there is no need to modify the starting date of 1 April.

Acknowledgements The authors would like to thank Gilles Oliver for his part in collecting the pollen data for Lyon, Miklós Juhász for providing pollen data of Szeged, and Zoltán Sümeghy for the digital mapping in Fig. 1.

References

Aznarte JL, Sánchez JMB, Lugilde DN, Fernández CDL, de la Guardia CD, Sánchez FA (2007) Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. Expert Syst Appl 32(4):1218–1225

Banken R, Comtois P (1992) Concentration of ragweed pollen and prevalence of allergic rhinitis in 2 municipalities in the Laurentides. Allerg Immunol 24:91–94

Bartkova-Scevkova J (2003) The influence of temperature, relative humidity and rainfall on the occurrence of pollen allergens (Betula, Poaceae, *Ambrosia artemisiifolia*) in the atmosphere of Bratislava (Slovakia). Int J Biometeorol 48(1):1–5

Béres I, Novák R, Hoffmanné Pathy Zs, Kazinczi G (2005) Spreading, morphology, biology, importance of mugwort leaves ragweed and possibilities of protection. [Az ürömlevelű parlagfű (*Ambro-*

sia artemisiifolia L.) elterjedése, morfológiája, biológiája, jelentősége és a védekezés lehetőségei.]. Gyomnövények, Gyomirtás 6(1):1–48, in Hungarian

Bottero P, Venegoni E, Riccio G, Vignati G, Brivio M, Novi C, Ortolani C (1990) Pollinosi da *Ambrosia artemisiifolia* in Provincia di Milano. Folia Allergol Immunol Clin 37(2):99–105

Cai Z (2007) Trending time-varying coefficient time series models with serially correlated errors. J Econometrics 136:163–188

Cassagne E (2009) Revue bibliographique des principaux seuils de détermination et méthodes de prévision de la date de début de pollinisation (DDP). Rev Fr Allergol 49(8):571–576

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

Chrenová J, Mičieta K, Ščevková J (2009) Monitoring of Ambrosia pollen concentration in the atmosphere of Bratislava (Slovakia) during years 2002–2007. Aerobiologia 26:83–88. doi:10.1007/s10453-009-9145-3

D'Amato G, Cecchi L (2008) Effects of climate change on environmental factors in respiratory allergic diseases. Clin Exp Allergy 38(8):1264–1274

D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, Liccardi G, Popov T, van Cauwenberge P (2007) Allergenic pollen and pollen allergy in Europe. Allergy 62(9):976–990

Dechamp C, Rimet ML, Meon L, Deviller P (1997) Parameters of ragweed pollination in the Lyon's area (France) from 14 years of pollen counts. Aerobiologia 13:275–279

Fan J (1992) Design-adaptive nonparametric regression. J Am Stat Assoc 87:998–1004

Frei T, Gassner E (2008) Climate change and its impact on birch pollen quantities and the start of the pollen season an example from Switzerland for the period 1969–2006. Int J Biometeorol 52(7):667–674

Galán C, Cariňanos P, García-Mozo H, Alcázar P, Domínguez-Vilches E (2001) Model for forecasting *Olea europaea* L. airborne pollen in South-West Andalusia, Spain. Int J Biometeorol 45(2):59–63

García-Mozo H, Galán C, Belmonte J, Bermejo D, Candau P, de la Guardia CD, Elvira B, Gutierrez M, Jato V, Silva I, Trigo MM, Valencia R, Chuine I (2009) Predicting the start and peak dates of the Poaceae pollen season in Spain using process-based models. Agric For Meteorol 149(2):256–262

Hirst JM (1952) An automatic volumetric spore trap. Ann Appl Biol 39:257–265

Ianovici N, Sîrbu C (2007) Analysis of airborne ragweed (*Ambrosia artemisiifolia* L.) pollen in Timişoara, 2004. An Univ Oradea Fascicula Biol 14:101–108

Jäger S (1998) Global aspect of ragweed in Europe. In: Spieksma FThM (ed) Satellite Symposium Proceedings: Ragweed in Europe. Proceedings of the 6th International Congress on Aerobiology. Alk-Abelló, Perugia (Italy), pp 6–10

Juhász M (1998) History of ragweed in Europe. In: Spieksma FThM (ed) Ragweed in Europe. Satellite Symposium Proceedings of the 6th International Congress on Aerobiology. Alk-Abelló, Perugia (Italy), pp 11–14

Kasprzyk I (2008) Non-native Ambrosia pollen in the atmosphere of Rzeszow (SE Poland); evaluation of the effect of weather conditions on daily concentrations and starting dates of the pollen season. Int J Biometeorol 52(5):341–351

Köppen W (1931) Grundriss Der Klimakunde. De Gruyter, Berlin

Laaidi M (1997) Influence des facteurs météorologiques sur la concentration du pollen dans l'air. Climat Santé 17:7–25

Laaidi M (2001) Regional variations in the pollen season of Betula in Burgundy: two models for predicting the start of the pollination. Aerobiologia 17(3):247–254

Laaidi M, Thibaudon M, Besancenot JP (2003) Two statistical approaches to forecasting the start and duration of the pollen season of Ambrosia in the area of Lyon (France). Int J Biometeorol 48:65–73

Lejoly-Gabriel HL (1978) Recherches écologiques sur la pluie pollinique en Belgique. Acta Geogr Lovan 13:1–278

Li Q, Racine J (2004) Cross-validated local linear nonparametric regression. Stat Sinica 14:485–512

Makra L, Juhász M, Borsos E, Béczi R (2004) Meteorological variables connected with airborne ragweed pollen in Southern Hungary. Int J Biometeorol 49(1):37–47

Makra L, Juhász M, Béczi R, Borsos E (2005) The history and impacts of airborne Ambrosia (Asteraceae) pollen in Hungary. Grana 44(1):57–64

Makra L, Sz T, Bálint B, Sümeghy Z, Sánta T, Hirsch T (2008) Influences of meteorological parameters and biological and chemical air pollutants to the incidence of asthma and rhinitis. Climate Res 37(1):99–119

Mandrioli P, Di Cecco M, Andina G (1998) Ragweed pollen: The aeroallergen is spreading in Italy. Aerobiologia 14:13–20

Nilsson S, Persson S (1981) Tree pollen spectra in the Stockholm region (Sweden), 1973-1980. Grana 20:179–182

Ocana-Peinado F, Valderrama MJ, Aguilera AM (2008) A dynamic regression model for air pollen concentration. Stoch Environ Res Risk A 22:S59–S63

Peternel R, Čulig J, Hrga I, Hercog P (2006) Airborne ragweed (*Ambrosia artemisiifolia* L.) pollen concentrations in Croatia, 2002–2004. Aerobiologia 22(3):161–168

Puc M (2006) Ragweed and mugwort pollen in Szczecin, Poland. Aerobiologia 22(1):67–78

Ribeiro H, Cunha M, Abreu I (2008) Quantitative forecasting of olive yield in Northern Portugal using a bioclimatic model. Aerobiologia 24(3):141–150

Rodríguez-Rajo FJ, Jato V, Aira MJ (2005) Relationship between meteorology and Castanea airborne pollen. Belg J Bot 138 (2):129–140

Rodríguez-Rajo FJ, Valencia-Barrera RM, Vega-Maray AM, Suarez FJ, Fernandez-Gonzalez 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, Grewling L, Stach A, Smith M (2009) Factors involved in the phenological mechanism of Alnus flowering in Central Europe. Ann Agric Environ Med 16(2):277–284

Rodríguez-Rajo FJ, Astray G, Ferreiro-Lage JA, Aira MJ, Jato-Rodriguez 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

Ruml M, Vukovič A, Milatovič D (2009) Evaluation of different methods for determining growing degree-day threshold in apricot cultivars. Int J Biometeorol 54:411–422. doi:10.1007/s00484-009-0292-06

Sánchez Mesa JA, Galán C, Hervás C (2005) The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate. Int J Biometeorol 49(6):355–362

Šikoparija B, Smith M, Skjøth CA, Radišič P, Milkovska S, Šimič S, Brandt J (2009) The Pannonian plain as a source of Ambrosia pollen in the Balkans. Int J Biometeorol 53(3):263–272

Snyder RI, Spano D, Cesaraccio C, Duce P (1999) Determining degree-day thresholds from field observations. Int J Biometeorol 42:177–182

Stach A, Smith M, Baena JCP, Emberlin J (2008) Long-term and short-term forecast models for Poaceae (grass) pollen in Poznań, Poland, constructed using regression analysis. Environ Exp Bot 62(3):323–332

Štefanič E, Kovačevič V, Lazanin Ž (2005) Airborne ragweed pollen concentration in north-eastern Croatia and its relationship with meteorological parameters. Ann Agric Environ Med 12:75–79

Stepalska D, Myszkowska D, Wolek J, Piotrowicz K, Obtulowicz K (2008) The influence of meteorological factors on Ambrosia pollen loads in Cracow, Poland, 1995–2006. Grana 47(4):297–304

Teran L, Haselbarth-Lopez MMM, Quiroz-Garcia DL (2009) Allergy, pollen and the environment. Gac Méd Méx 145(3):215–222

Traidl-Hoffmann C, Kasche A, Menzel A, Jakob T, Thiel M, Ring J, Behrendt H (2003) Impact of pollen on human health: More than allergen carriers? Int Arch Allergy Imm 131:1–13

Turos OI, Kovtunenko IN, Markevych YP, Drannik GN, DuBuske LM (2009) Aeroallergen monitoring in Ukraine reveals the presence of a significant ragweed pollen season. J Allergy Clin Immun 123(2):S95-S95, Suppl. S358

Verma KS, Pathak AK (2009) A comparative analysis of forecasting methods for aerobiological studies. Asian J Exp Sci 23:193–198

Wan SQ, Yuan T, Bowdish S, Wallace L, Russell SD, Luo YQ (2002) Response of an allergenic species *Ambrosia psilostachya* (Asteraceae), to experimental warming and clipping: Implications for public health. Am J Bot 89(11):1843–1846

Wopfner N, Gadermaier G, Egger M, Asero R, Ebner C, Jahn-Schmid B, Ferreira F (2005) The spectrum of allergens in ragweed and mugwort pollen. Int Arch Allergy Imm 138:337–346