Theor Appl Climatol (2011) 104:277–283
DOI 10.1007/s00704-011-0431-y

ORIGINAL PAPER

Autoregressive modelling of daily ragweed pollen concentrations for Szeged in Hungary

István Matyasovszky · László Makra

Received: 18 May 2010 / Accepted: 7 March 2011 / Published online: 2 April 2011

Abstract Studying airborne pollen concentrations is an essential part of aerobiology owing to its important applications in allergology. A time-varying first order autoregressive (AR(1)) model able to describe the annual cycle of both the expectation and variance as well as the highly skewed probability distribution of daily ragweed pollen concentrations conditioned on previous-day pollen concentration values is developed. Confidence bands for forecasts obtained with these conditional lognormal distributions are analysed. The probability of exceeding specific pollen concentration thresholds is also addressed with the model based on a refinement of the AutoRegressive To Anything process. In order to have more accurate forecasts for the next-day pollen concentration level, eight meteorological variables influencing pollen concentrations are considered. Based on a procedure similar to the stepwise regression method, only one predictor has been retained, namely the daily mean temperature. Using root mean square error, the percentage variance of the ragweed pollen concentration level accounted for by this extended AR(1) model is 53.5%, while the mean absolute error produced by the model is 32.2 pollen grains m^{-3}. The probability of exceeding pollen concentration thresholds obtained from the conditional lognormal distributions under the extended AR(1) model fits well the observed exceedance events.

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

L. Makra
Department of Climatology and Landscape Ecology,
University of Szeged,
PO Box 653, Szeged 6701, Hungary
e-mail: makra@geo.u-szeged.hu

1 Introduction

In Southern Europe, the main plants that cause pollen allergies in Europe are grasses (Poaceae), birch (Betulaceae), mugwort (*Artemisia* spp.) and the olive tree (Oleaceae) (D'Amato et al. 2007). In the 1980s, a new species started to spread extremely aggressively. It is appearing in more and more countries, its blooming lasts for a long time (in some regions even for 3 months) and it produces a lot of pollen. This is ragweed (*Ambrosia* spp.), which is an annual, wind-pollinated plant.

Studying airborne pollen concentrations is essential in aerobiology due to its important applications in allergology. One of the common tools used for this task is integrated autoregressive–moving average (ARIMA) modelling. Autoregressive (AR) procedures are often used to model time series in order to gain an insight into the statistical nature of the underlying data set. Frequently, such models are also used for forecasting purposes. But the autoregressive modelling of daily ragweed pollen concentration levels has several problems. For instance, pollen concentrations exhibit strong annual cycles that have to be removed. This is traditionally done by using a linear filter or by specifying an analytical form for the annual trend (e.g. Belmonte and Canela 2002). Another possibility is differencing. Here, as the trend is smooth enough, differences between consecutive pollen concentrations approximately filter out the annual cycle. Hence, an ARIMA model should be applied (Arca et al. 2002; Rodríguez-Rajo et al. 2006). The annual cycle is present not only in the expectation, but in the variance as well. Therefore, a logarithmic transformation is applied to the data before the trend removal (e.g. Aznarte et al. 2007) in order to stabilise the variance. Unfortunately, the procedure reduces, but does not remove variations in the variance.

A further problem is that meteorological factors affecting pollen production can produce different pollen concentrations at different times of the pollination season. Just taking the daily temperature for reasons of simplicity, a given temperature (or temperature minus its annual cycle) may have a very different effect on pollen production (or pollen minus its annual cycle) in late September than it will in mid-August. Thus, it is reasonable to assume that an appropriate AR model will have time-dependent coefficients.

Next, the standard AR model is based on the assumption that the residual term of the autoregression comes from identically and independently distributed Gaussian variables that result in Gaussian AR processes. Therefore, confidence bands can be easily constructed for forecasts obtained from such models. However, the Gaussian behaviour assumption is evidently violated for pollen concentration data because their probability distributions are highly skewed.

The aim of this paper is to develop a time-varying AR model that is able to describe the annual cycle of both the expectation and variance as well as the probability distribution of daily ragweed pollen concentrations conditioned on previous-day pollen concentration values. Confidence bands for forecasts obtained with this model are analysed. The probability of exceeding specific pollen concentration thresholds is also discussed using the model. The methodology will be applied to daily ragweed pollen concentrations of Szeged in Hungary, as the area of Szeged and its surroundings is a representative area in the Carpathian basin, and it has the highest peak ragweed pollen concentrations recorded in Europe (Makra et al. 2005).

2 Methodology

Let us take for simplicity a stationary first order autoregressive (AR(1)) process defined by

$$Y_i = a + bY_{i-1} + e_i\,,\ i = 1, 2, ...,$$

where e_i are identically and independently distributed Gaussian residuals with an expectation of zero and a variance of δ^2. Hence the process Y_i is Gaussian. The probability distribution of Y_i conditioned on Y_{i-1}, Y_{i-2},$\cdots$ is also Gaussian with conditional expectation $a+bY_{i-1}$ and variance $\delta^2=(1-b^2)\sigma^2$, where σ^2 is the variance of Y_i. When the underlying process is not Gaussian, both the conditional density $f(y|x)$ of Y_i at $Y_{i-1}=x$ and density of the residuals are hard to determine.

Lawrance (1982) developed a gamma distributed AR(1) process based on residuals. It was found that there is a positive probability for residuals to have zero values and hence for having two consecutive values of the process to be identical. In order to remove this unrealistic property, Gourieroux and Jasiak (2006) defined a gamma autoregressive process based on the conditional density $f(y|x)$. Grunwald et al. (2002) put the problem in a more general framework for a wide class of conditional densities. Our original intention was to use this advanced technique, but the AutoRegressive To Anything (ARTA) process introduced for simulation purposes by Cario and Nelson (1996) was applied to the present problem because it was easier to implement. The results produced by a refined version of the original ARTA model are usually just as good as those produced by the approach of Grunwald et al. (2002).

The ARTA model seeks to describe the correlation structure of a stationary stochastic process Y_1, Y_2,$\cdots$ reproducing their stationary probability distribution function $F(y)$. A Gaussian AR(1) model is related to Y_1, Y_2,$\cdots$ by

$$Z_i = bZ_{i-1} + e_i,$$

where the identically and independently distributed Gaussian residuals e_i have an expectation of zero and a variance of $1-b^2$. Adopting a time series y_1, y_2,$\cdots$, y_n, the correlation structure of $z_i = \Phi^{-1}(F(y_i))$ and thus the autoregressive coefficient b can be determined from the estimated autocorrelations of y_1, y_2,$\cdots$, y_n, where $\Phi(u)$ denotes the standard normal probability distribution function evaluated at u (Cario and Nelson 1996). Here, the conditional expectation and variance of z_i are bz_{i-1} and b^2, respectively. As $y_i = F^{-1}(\Phi(z_i))$, the conditional expectation of y_i seems to be $F^{-1}(\Phi(bz_{i-1}))$. Unfortunately, this is not true because the last expression gives the conditional median. Knowing, however, the relationship between the median and expectation for $F(y)$, the conditional expectation can be obtained. The procedure can easily be extended to AR processes of orders $p>1$.

Next, the probability distribution function $F(y)$ has to be specified. Comtois (2000) suggested using gamma distributions for pollen concentration data. Seeing how well this (or any other) distribution type fits ragweed data is not easy as the distribution is time-varying due to the strong annual cycle. Therefore, a window is defined around any specific day of the pollination season, and every data value in the window width is considered. Then, a standard technique such as the chi-square test or the Kolmogorov–Smirnov test is applied. The window should be wide enough to have an appropriate number of data items, but it has to be quite narrow to neglect the seasonal trend within the window width. Here, the window is chosen as to provide 90 data for every particular day of the year during the entire 10-year data set. Unfortunately, the gamma distribution does not fit daily ragweed data very well in our case. The lognormal distribution suggested for instance by Limpert et al. (2008) was chosen instead as this distribution type is much more suitable for our purposes. Both tests mentioned above show that this distribution type fits the data at an 80–95%

significance level except at the beginning and end of the pollination season. For these latter periods, the significance level is only around 99–99.9%. These unconvincing levels are probably due to the choice of window widths. Specifically, windows providing the above mentioned 90 data cover only 4-day intervals around days in the interior of the pollination season. However, this same amount of data for the first or last day of the pollination season can be provided by 9-day window widths that are too wide to neglect the seasonal trend within these widths. Hence, our decision is to apply lognormal distributions for the entire pollination season. As the transformations $y_i = \exp(z_i)$, $z_i = \ln(y_i)$ between ragweed data and the associated Gaussian data are simple, a slight modification of the ARTA model is introduced by estimating the autoregressive parameter b from the data set $z_1, z_2, \cdots, z_n$.

The remaining task is to extend the methodology to the time-varying case. The proposed method is a nonparametric approach similar to linear filters, but here, both the width and coefficients of the filter are selected in an optimal way (Fan 1992). Let daily ragweed pollen concentrations from July 15 to October 15 be denoted by y_i, $i=1,\cdots,n$ at times $t_1,\cdots,t_n$. These latter values are scaled from July 15 in each particular year, i.e. $t_{u+vM} = t_u$, $u = 1, \ldots, M-1, v = 1, \ldots, N-1$, where $N=10$ is the number of years and $M=93$ is the length of the period examined for each year (see Section 3.1). Using the chosen $F(y)$, the conditional density $f(y|y_{i-1})$ is lognormal with parameters $\mu(t_i) = a(t_i) + b(t_i)z_{i-1}$ and $\sigma(t_i)$, i.e.

$$f(y|y_{i-1}) = \frac{1}{\sqrt{2\pi}\sigma(t_i)y} \exp\left(-\frac{1}{2\sigma^2(t_i)}(\ln(y) - a(t_i) - b(t_i)\ln(y_{i-1}))^2\right).$$

If we suppose that coefficients in the associated Gaussian AR(1) process "smoothly" vary in time, then they can be approximated locally linearly. According to the idea of Cai (2007), the quantity

$$\sum_{k=2}^{n} (z_k - [\alpha_0 + \beta_0 z_{k-1} + (\alpha_1 + \beta_1 z_{k-1})(t_k - t_i)])^2 K\left(\frac{t_k - t_i}{h}\right) \tag{1}$$

has to be minimised with respect to $\alpha_j, \beta_j, j=0,1$ for $2\le i\le n$, and $\widehat{a}(t_i) = \widehat{\alpha}_0$, $\widehat{b}(t_i) = \widehat{\beta}_0$. The function K is called the Epanechnikov kernel and is defined as $K(u) = 3/4(1-u^2)$ in [−1, 1] and zero otherwise (Fan 1992). The so-called 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 recognises the trade-off between the bias and variance has to be found as, for instance, in the paper by Cai (2007). The $\sigma(t_i)$ is estimated as $\widehat{\sigma}^2(t_i) = \widehat{\lambda}_0$ with $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ minimising the quantity

$$\sum_{k=2}^{n} (\widehat{z}_k - [\lambda_0 + \lambda_1(t_k - t_i)])^2 K\left(\frac{t_k - t_i}{h}\right),$$

for $2\le i\le n$, where $\widehat{z}_k = z_k - \widehat{a}(t_k) - \widehat{b}(t_k)z_{k-1}$. Utilising the relationship between the median and expectation of a lognormal distribution, the expectation of y_i conditioned on y_{i-1} is

$$\exp\left(\widehat{a}(t_i) + \widehat{b}(t_i)\ln(y_{i-1})\right)\exp\left(\widehat{\sigma}^2(t_i)/2\right).$$

3 An application of the method

3.1 Geographical location and datasets

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. The area is characterised by an extensive flat landscape of the Great Hungarian Plain with an elevation of 79 m AMSL. The city is the centre of the Szeged region with 203,000 inhabitants. The climate of Szeged can be characterised by Köppen's Ca type (warm temperate climate), with relatively mild and short winters and hot summers (Köppen 1931). The pollen concentration of the air was measured using a 7-day recording "Hirst-type" volumetric trap (Hirst 1952). The air sampler is located on top of the building of the Faculty of Arts at the University of Szeged some 20 m above the ground (Makra et al. 2008).

The period July 15–October 15 for the years 1997–2006 available was selected. This interval covers most of the ragweed pollination period in Szeged using the criterion of Galán et al. (2001). Namely, the start (end) of the pollen season is the earliest (latest) date on which at least 1 pollen grain m^{-3} is recorded and at least five consecutive (preceding) days also have 1 or more pollen grains m^{-3}. The mean of this yearly varying period is selected for the 10-year (1997–2006) period examined.

3.2 Results

The methodology was applied to the data sets mentioned above. An evaluation of the performance of our AR modelling approach requires a comparison with a base estimation. This base estimate is the annual trend of the ragweed pollen concentration obtained with Eq. (1), but omitting terms containing z_{k-1}. As the conditional expectation and the conditional median minimise the root mean square error (RMSE) and the mean absolute error (MAE), respectively, these two quantities are compared to the same quantities obtained from the annual trend. Using RMSE, the percentage variance of the ragweed pollen concentration accounted for by the AR(1) model is 51.7%. The MAE value produced by the model is 32.7 pollen grains m^{-3}, which is well below the value of 53.9 pollen grains m^{-3} provided by the annual trend.

In order to check the accuracy of the conditional lognormal distributions, time-varying conditional quantiles y_ε and $y_{1-\varepsilon}$ are calculated from these distributions. As the εth quantile ($0<\varepsilon<1$) of a random variable denotes that value below which the random variable takes values with probability ε, the relative frequency of pollen concentration values below y_ε and above $y_{1-\varepsilon}$ should be approximately ε. The question of whether this relative frequency differs significantly from ε has to be examined. Taking for instance ε=0.1, the corresponding conditional relative frequency is 0.100 and 0.098 for $y_{0.1}$ and $y_{0.9}$, respectively. Similar values for $y_{0.05}$ and $y_{0.95}$ are 0.055 and 0.038 (97.6%), while for $y_{0.01}$ and $y_{0.99}$ are 0.008 and 0.006 (94.4%). The percentage values in parentheses show the significance levels, where the relative frequency values do not differ significantly from the corresponding probability values. No percentages added means that the relative frequency does not differ from the corresponding probability at any reasonable significance level. The question of whether the observed relative frequencies have annual cycles is also dealt with. Both the chi-square and Kolmogorov–Smirnov tests show at any reasonable significance level that the pollen concentration values beyond the interval [y_ε, $y_{1-\varepsilon}$) are uniformly distributed in time. It may be concluded here that conditional lognormal distributions defined above are suitable for describing the statistical nature of consecutive daily ragweed pollen concentrations even though applying (unconditional) lognormal distributions seems inappropriate at the beginning and end of the pollination season (see Section 2).

Note that the AR model introduced above provides forecasts for the next-day pollen concentration level. To have even more accurate forecasts, meteorological variables that influence pollen concentrations can be involved in the model building process (Grunwald et al. 2002). Eight meteorological variables (mean temperature, mean wind speed, mean relative humidity, mean global solar flux, mean atmospheric sea-level pressure, minimum temperature, maximum temperature and precipitation amount) as candidate predictors are analysed. The selection of optimal predictors is based on a procedure similar to the well-known stepwise regression method (Draper and Smith 1981). Namely, suppose we have a set of chosen predictors with their bandwidth. When including additional predictors in the estimation procedure, there is an obvious chance of getting higher optimal bandwidths because higher dimensional forecasting surfaces can be represented by larger amounts of data produced with larger bandwidths. However, larger bandwidths provide biases larger than those obtained with predictors chosen earlier. 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 value. Based on the procedure summarised above, only one predictor was retained, namely the daily mean temperature. The conditional density $f(y|y_{i-1})$ is again lognormal, but with parameters $\mu(t_i) = a(t_i) + b(t_i)z_{i-1} + c(t_i)x_{i-1}$ and $\sigma(t_i)$, where $x_1, \cdots, x_n$ denote daily mean temperatures. Eq. (1) becomes

$$\sum_{k=2}^{n} (z_k - [\alpha_0 + \beta_0 z_{k-1} + \gamma_0 x_{k-1} + (\alpha_1 + \beta_1 z_{k-1} + \gamma_1 x_{k-1})(t_k - t_i)])^2 K\left(\frac{t_k - t_i}{h}\right),$$

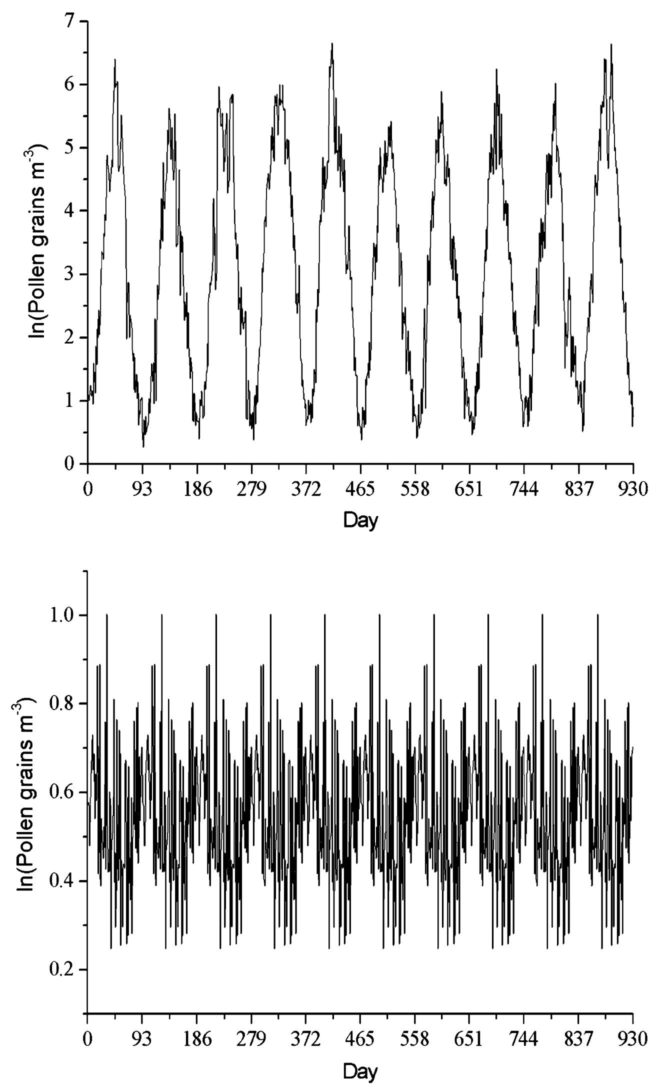

Fig. 1 Parameters μ (*top*) and σ (*bottom*) of the lognormal conditional distribution. Here, the 93 value (and its multiples) on the horizontal axis refers to the length of the period examined for each year

and $\widehat{a}(t_i) = \widehat{\alpha}_0$, $\widehat{b}(t_i) = \widehat{\beta}_0$, $\widehat{c}(t_i) = \widehat{\gamma}_0$. Lastly, $\sigma(t_i)$ is estimated as $\widehat{\sigma}^2(t_i) = \widehat{\lambda}_0$ with $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ minimising the quantity

$$\sum_{k=2}^{n} (\widehat{z}_k - [\lambda_0 + \lambda_1(t_k - t_i)])^2 K\left(\frac{t_k - t_i}{h}\right),$$

where $\widehat{z}_k = z_k - \widehat{a}(t_k) - \widehat{b}(t_k)z_{k-1} - \widehat{c}(t_k)x_{t-k}$. Figure 1 shows how the lognormal parameters μ and σ vary in time.

Using RMSE, the percentage variance of the ragweed pollen concentration accounted for by this extended AR(1) model is 53.5%, while the MAE value produced by the model is 32.2 pollen grains m^{-3}. Figure 2 demonstrates the good fit of the expected daily pollen concentrations conditioned on previous-day values with observed pollen concentrations. Comparing these values with those obtained from the AR(1) model shows just moderate improvement. In order to check the accuracy of the conditional lognormal distributions, time-varying conditional quantiles y_ε and $y_{1-\varepsilon}$ are also calculated from these new distributions. Taking again ε=0.1, the conditional relative frequency of pollen concentration values below y_ε and above $y_{1-\varepsilon}$ is 0.100 and 0.102 for $y_{0.1}$ and $y_{0.9}$, respectively. Similar values for $y_{0.05}$ and $y_{0.95}$ are 0.050 and 0.040 (95.9%), while for $y_{0.01}$ and $y_{0.99}$ are 0.010 and 0.008, respectively. Note that these relative frequencies are even closer to the expected probabilities than in the case of the AR(1) model. Taking a substantially larger probability value, say ε=0.25, the conditional relative frequency of pollen concentration values below $y_{0.25}$ and above $y_{0.75}$ is 0.252 and 0.277 (98.5%), respectively. These values again appear satisfactory.

In order to evaluate the performance of the extended AR (1) model, we relate the probability of exceeding specific pollen concentration thresholds to observed exceedance events. Namely, these probability values obtained from the conditional distributions are compared to an indicator

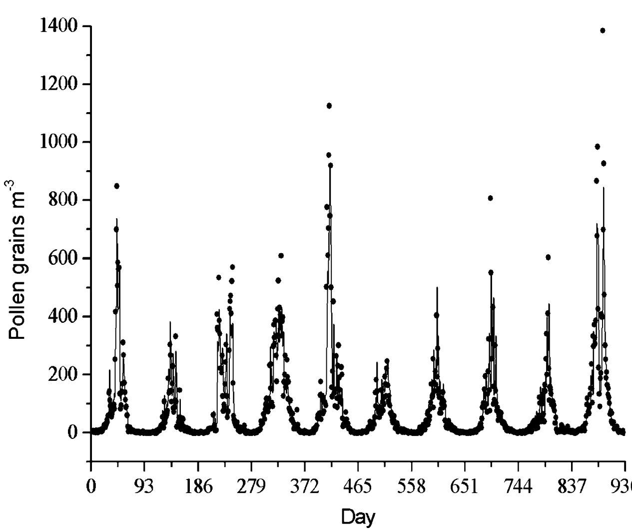

Fig. 2 Expected daily pollen concentrations conditioned on previous-day values (*solid*) and observed pollen concentrations. Here, the 93 value (and its multiples) on the horizontal axis refers to the length of the period examined for each year

Table 1 RMSE and MAE for the daily threshold exceedance probability value got with the extended AR(1) model and annual trend

Threshold pollen grains m^{-3}	RMSE		MAE	
	AR(1)	Trend	AR(1)	Trend
20	0.213	0.261	0.092	0.149
50	0.196	0.254	0.081	0.142
100	0.241	0.295	0.113	0.187
200	0.213	0.270	0.090	0.152

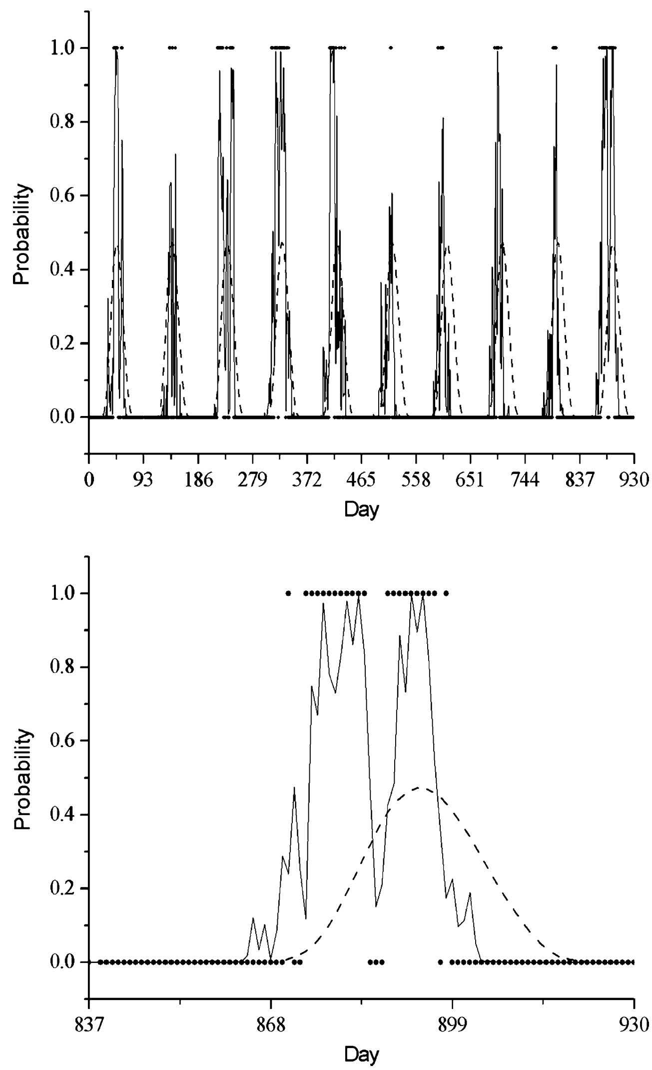

Fig. 3 Probability value of exceeding a 200-pollen grain m^{-3} pollen concentration threshold got from conditional lognormal distributions (*solid line*) and from the annual trend of an indicator variable taking values of one or zero (*dashed line*) with values of the indicator variable for the entire period (*top*) and for the previous year (*bottom*). Here, the 93 value (and its multiples) on the horizontal axis refers to the length of the period examined for each year

variable that equals one when the actual pollen concentration exceeds the threshold and zero otherwise. The difference between estimated probabilities and observed exceedances is quantified by RMSE and MAE. Similar calculations are also performed with the annual trend of the indicator variable data. The threshold of 20 grains m^{-3} corresponding to the critical level for health risks (Jäger 1998) is examined, but higher thresholds 50, 100 and 200 grains m^{-3} are also analysed. Table 1 and Fig. 3, especially the enlarged last year, demonstrate a good fit of the exceedance probabilities to the indicator variable values as compared to the annual trend of this indicator variable.

4 Conclusions

Here, a time-varying AR(1) model able to describe the annual cycle of both the expectation and variance as well as the probability distribution of daily ragweed pollen concentrations conditioned on previous-day pollen concentration values was developed. The methodology could be used to fit higher order AR models, but the AR(1) case was found optimal in terms of minimising the RMSE value. Note that searching for higher orders is a statistical problem similar to examining meteorological variables that could be included in the extended AR model.

It was found by using the chi-square and Kolmogorov–Smirnov tests that the lognormal probability distribution fits pollen concentration data at only a 99–99.9% significance level at the beginning and end of the pollination season, but the significance level was much more favourable in the middle of the pollination season. The unconvincing 99–99.9% levels represent just an artefact produced by the annual cycle. Similar tests cannot be performed for the lognormal conditional densities $f(y|y_{i-1})$ generated by the AR model, because no conditional samples are available for any fixed y_{i-1} as every y_{i-1} is followed by only one unique value. Therefore, time-varying conditional quantiles corresponding to several given probability values were calculated from the conditional distributions, and the observed relative frequency values that did not exceed these quantiles were compared to the given probability values. The correspondence between these probability values and relative frequency values is quite satisfactory, confirming that the AR model is sound.

Using RMSE, the percentage variance of the ragweed pollen concentration level accounted for by the AR(1) model is 51.7%. The MAE value produced by the model is 32.7 pollen grains m^{-3}, which is well below 53.9 pollen grains m^{-3} produced by the annual trend. As the AR model provides forecasts for the next-day pollen concentration level, more accurate forecasts should be possible if those meteorological variables that influence pollen concentration levels are taken into account. However, only one of eight candidate predictors, the daily mean temperature was retained. The percentage variance of the ragweed pollen concentration accounted for by this extended AR(1) model is 53.5%, while the MAE produced by the model is 32.2 pollen grains m^{-3}, indicating a moderate improvement. Also, the probability of exceeding specific pollen concentration thresholds obtained from the conditional lognormal distributions corresponding to the extended AR (1) model fits nicely the observed exceedance events (Fig. 3, Table 1).

References

Arca B, Pellizzaro G, Canu A, Vargiu G (2002) Airborne pollen forecasting: evaluation of ARIMA and neural network models. 15th Conference on Biometeorology and Aerobiology, Kansas City, MO, U.S.A., P 3.7

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 32:1218–1225

Belmonte J, Canela M (2002) Modelling aerobiological time series: application to Urticaceae. Aerobiologia 18:287–295

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

Cario MC, Nelson BL (1996) Autoregressive to anything: time-series input processes formulation. Oper Res Lett 19:51–58

Comtois P (2000) The gamma distribution as the true aerobiological probability density function (PDF). Aerobiologia 16:171–176

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 62:976–990

Draper N, Smith H (1981) Applied regression analysis. Wiley, New York

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

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

Gourieroux C, Jasiak J (2006) Autoregressive gamma processes. J Forecast 25:129–152

Grunwald GK, Hyndman RJ, Tedesco L, Tweedie RL (2002) Non-gaussian conditional linear AR(1) models. Aust NZ J Stat 42:479–495

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

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

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

Lawrance AJ (1982) The innovation distribution of a gamma distributed autoregressive process. Scand J Statist 9:234–236
Limpert E, Burke J, Galán C, del Mar TM, West JS, Stahel WA (2008) Data, not only in aerobiology: how normal is the normal distribution? Aerobiologia 24:121–124
Makra L, Juhász M, Béczi R, Borsos E (2005) The history and impacts of airborne *Ambrosia* (Asteraceae) pollen in Hungary. Grana 44:57–64
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. Clim Res 37:99–119
Rodríguez-Rajo FJ, Valencia-Barrera RM, Vega-Maray AM, Suarez FJ, Fernández-González D, Jato V (2006) Prediction of airborne alnus pollen concentration by using Arima models. Ann Agr Env Med 13:25–32