Aerobiologia (2012) 28:337–346
DOI 10.1007/s10453-011-9238-7

ORIGINAL PAPER

Estimating extreme daily pollen loads for Szeged, Hungary using previous-day meteorological variables

István Matyasovszky · László Makra

Received: 16 May 2011 / Accepted: 8 November 2011 / Published online: 15 November 2011

Abstract The aim of this paper is to analyse how meteorological elements relate to extreme *Ambrosia* pollen load on the one hand and to extreme total pollen load excluding *Ambrosia* pollen on the other for Szeged, Southern Hungary. The data set comes from a 9-year period (1999–2007) and includes previous-day means of five meteorological variables and actual-day values of the two pollen variables. Factor analysis with special transformation was performed on the meteorological and pollen load data in order to find out the strength and direction of the association of the meteorological and pollen variables. Then, using selected low and high quantiles corresponding to probability distributions of *Ambrosia* pollen and the remaining pollen loads, the quantile and beyond-quantile averages of pollen loads were compared and evaluated. Finally, a nearest neighbour (NN) technique was applied to discriminate between extreme and non-extreme pollen events using meteorological elements as explaining variables. The observed below or above quantile events are compared with events obtained from NN decisions. The number of events exceeding the quantile of 90% and not exceeding that of 10% is strongly underestimated. However, the procedure works well for quantiles of 20 and 80%, and even better for those of 30 and 70%. Using a nearest neighbour technique, explaining variables in decreasing order of their influence on *Ambrosia* pollen load are temperature, global solar flux, relative humidity, air pressure and wind speed, while on the load of the remaining pollen are temperature, relative humidity, global solar flux, air pressure and wind speed.

Keywords Ambrosia · Extreme daily pollen load · Meteorological elements · Factor analysis including special transformation · *t* test · Nearest neighbour technique

1 Introduction

Connection of meteorological elements with pollen concentrations is widely studied in the literature. Finding a statistically significant association between the daily pollen level and daily meteorological elements is of great practical importance. These kinds of examinations concern all pollen types and include correlation analyses (Celenk et al. 2009; Kasprzyk and Walanus 2010), forecasting characteristics of the pollen season (García-Mozo et al. 2009; Kasprzyk 2009) and pollen concentration using regression models (Makra et al. 2004; Ocana-Peinado et al. 2008),

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

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

neural and neuro-fuzzy models (Aznarte et al. 2007) or multivariate statistical methods (Hart et al. 2007; Makra et al. 2006).

However, the role of the values of meteorological elements in forming extreme daily pollen concentrations has received so far a low attention. Frei (2004, 2006) studied the occurrences of extreme events (storms, floods or droughts) with extreme birch and grass pollen concentrations in the data set of Basel. The heat wave over Europe in summer 2003 with mean temperature exceeding the 1961–1990 mean by about 5°C in June, July and August substantially influenced pollen phenology and pollen production in Switzerland (Gehrig 2006). The grass pollen season was most affected starting 1–2 weeks earlier and ending 7–33 days earlier than in general. Extreme high *Chenopodium*, *Plantago* and Poaceae daily pollen concentrations were measured in this pollen season. Cariñanos et al. (2000) analysed the yearly distribution and severity of *Artemisia* and Chenopodiaceae-Amaranthaceae pollen, indicating the highest and very high pollen levels in a rural area with sub-desert climate and extreme dryness.

Due to the worldwide increasing trend and ever-increasing frequency of extreme high temperatures, the start of flowering occurs several days earlier; furthermore, a trend towards higher annual pollen quantities and an increase in the highest daily mean pollen concentrations can also be observed (Frei 2008; Frei and Gassner 2008). Recent climate change, the global warming may facilitate to extend habitat region of herbaceous and arboreal plants contributing to the increase in pollen levels and exacerbation of their adverse effects, hence to the rise of pollen sensitivity and respiratory admissions due to a pollen allergy.

The purpose of this paper is to analyse how previous-day values of meteorological elements relate to actual-day values of extreme *Ambrosia* pollen load on one the hand and to those of extreme total pollen load excluding *Ambrosia* pollen on the other. In the paper, value of pollen load below its quantiles 10, 20 and 30% and above the quantiles 90, 80 and 70% are considered extreme (A p-quantile ($0 < p < 100\%$) q_p is the value below which the pollen load occurs with relative frequency p). Our aim was to determine the chance of occurrence of extreme pollen load for a day in association with the former-day meteorological variables. In this way, results may help physicians and sensitive people to prepare to adverse effects of extreme high pollen loads. This is why actual meteorological variables were disregarded. In order to get a first insight into the relationship between pollen load variables and meteorological variables, a factor analysis with special transformation was performed first. Then, averages of meteorological variables under quantile and beyond-quantile events of *Ambrosia* pollen and the remaining pollen loads were compared and evaluated. If these averages differ significantly, the possibility of distinguishing between extreme and non-extreme pollen events using meteorological elements as explaining variables can be expected. Therefore, a nearest neighbour (NN) technique was applied to discriminate between extreme and non-extreme pollen events using meteorological elements.

2 Materials and methods

2.1 Location and data

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 AMSL. The city is the centre of the Szeged region with 203,000 inhabitants. The climate of Szeged belongs to Köppen's **Ca** type (warm temperate climate) with relatively mild and short winters and hot summers (Köppen 1931). The pollen content 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 surface (Fig. 1) (Makra et al. 2008).

In order to determine the association between meteorological variables on the one hand and *Ambrosia* pollen load as well as the total pollen load excluding *Ambrosia* pollen on the other, previous-day values of five meteorological variables (mean temperature, mean global solar flux, mean relative humidity, mean sea-level pressure and mean wind speed) and actual-day values of the two pollen variables were considered. In another work we received that out of temperature, relative humidity, wind speed and global solar flux, the most important predictor was the daily mean global solar flux for rainy days (Makra and Matyasovszky 2011). Concerning mean sea-level pressure (MSLP),

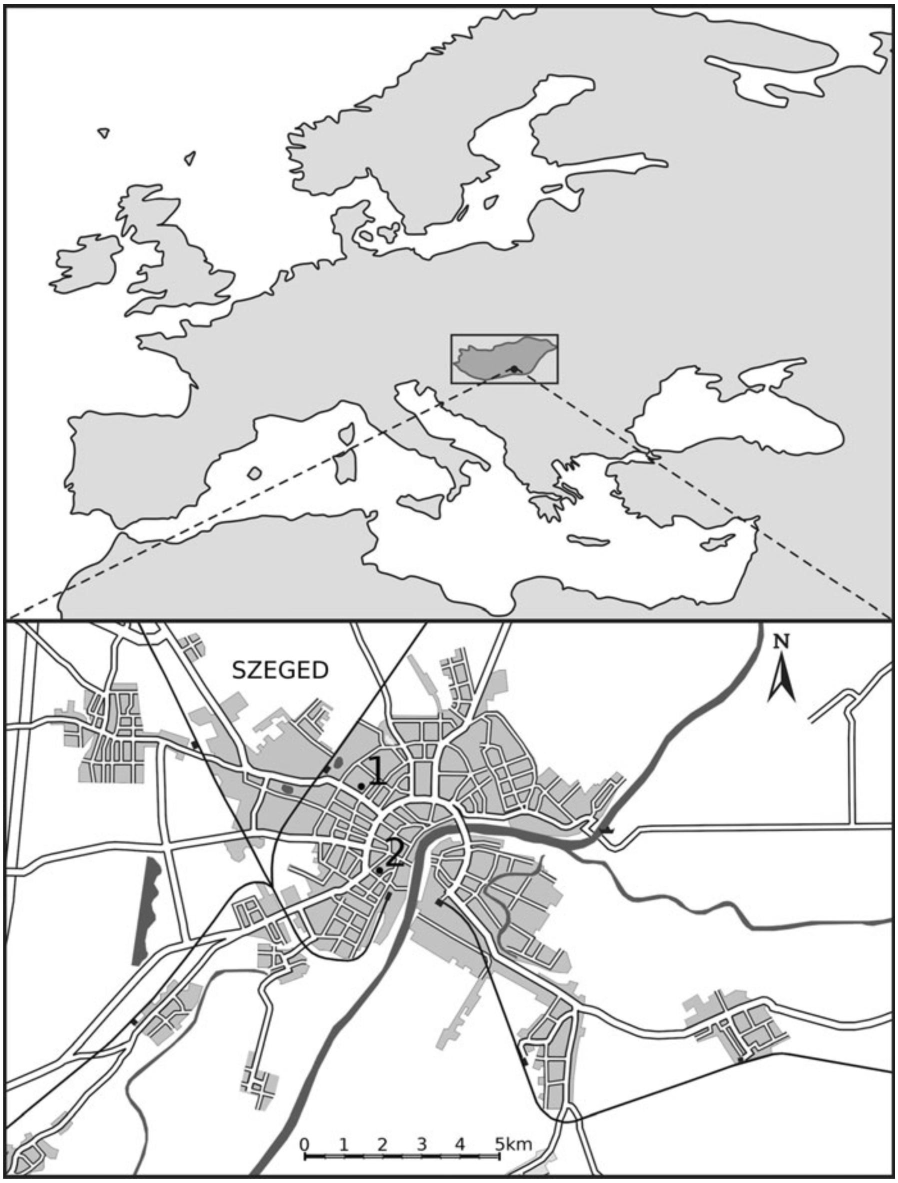

Fig. 1 Location of Europe with Hungary (*upper panel*) and the urban web of Szeged with the positions of the data sources (*lower panel*). *1* meteorological monitoring station, *2* aerobiological station. The distance between the aerobiological and the meteorological station is 2 km

we refer to Schäppi et al. (1998) who identified a strong positive correlation between daily average air temperature and daily grass pollen count. This correlation arises due to anomalies in MSLP, reflecting the association of MSLP with different direction of airflows in the forming of pollen load. A slight refinement of the methodology could involve not only previous-day meteorological variables as influencing variables but their two (or more) days earlier values as well. However, it was found that earlier-days meteorological parameters deliver negligible further information on the actual-day pollen concentration (Makra and Matyasovszky 2011).

Meteorological data were collected in the monitoring station (operating by the Environmental and Natural Protection and Water Conservancy Inspectorate of Lower-Tisza Region, Szeged) located in the downtown of Szeged at a distance of about 10 m from the busiest main road and with 2 km distance of the pollen trap. Though the meteorological station is an urban monitoring station, within such a distance, meteorological data can be considered representative to associate them with pollen characteristics at the pollen station.

Besides pollen of *Ambrosia* (ragweed), pollen of further 23 relevant taxa is taken into account. In Szeged, aerobiological measurements have been performed since 1989 (Makra et al. 2005). Pollen of altogether 24 taxa can be detected in the city (Makra et al. 2006). Ragweed in Szeged discharges the most pollen of all taxa. The ratio of *Ambrosia* pollen amounts to around 60–71% of the total pollen release in the air of Szeged in the late summer period (Juhász and Juhász 2002). For Szeged, the sensitivity of patients to ragweed is 83.7%, while it is 51.8% to mugwort and 56.7% to grass pollen (Kadocsa and Juhász, 2000). The taxa considered are as follows: *Acer* (maple), *Alnus* (alder), *Ambrosia* (ragweed), *Artemisia* (mugwort), *Betula* (birch), *Cannabis* (hemp), *Carpinus* (hornbeam), Chenopodiaceae (goosefoots), *Corylus* (hazel), *Fraxinus* (ash), *Juglans* (walnut), *Morus* (mulberry), *Pinus* (pine), *Plantago* (plantain), *Platanus* (platan), Poaceae (grasses), *Populus* (poplar), *Quercus* (oak), *Rumex* (dock), *Salix* (willow), *Taxus* (yew), *Tilia* (linden), *Ulmus* (elm) and *Urtica* (nettle).

The analysis was performed for the 9-year period 1999–2007 with two data sets according to the pollen season of *Ambrosia* (ragweed) (July 15–October 16) and the pollen season of remaining pollen excluding that of *Ambrosia* (January 14–October 16).

The pollen season is defined by its start and end dates. For the start (end) of the season, we used the first (last) date on which 1 pollen grain m^{-3} of air is recorded and at least 5 consecutive (preceding) days also show 1 or more pollen grains m^{-3} (Galán et al. 2001). The pollen season varies from year to year. The earliest start date and latest end date of individual pollen seasons are accepted to define a uniform pollen season for each year.

It should be noted that we define pollen load as a number indicating to which extent the body is endangered by pollen. When calculating pollen load, allergenic effects of all actually blooming herbaceous and arboreal plants are considered. According to the degree of allergenicity, pollen types can be sorted into four categories: (1) weakly (without any allergic symptoms), (2) moderately (infrequent, it triggers complaints only for a few people), (3) intensely (frequent allergen, it runs with complaints for lots of people) and (4) severely allergenic pollen types (very frequent allergen, it provokes strong reactions for lots of people) (http://www.orvosweb.hu/pdf/pollen_naptar.pdf). For example, allergenicity of *Ambrosia* is severe indicated by the scale value 4, while that of *Juglans* is weak denoted by the value 1. Hence, pollen load is the sum of the pollen concentrations multiplied by their degrees of allergenicity (http://www.pollenindex.hu/) (Fig. 2).

2.2 Methods

2.2.1 *Factor analysis with special transformation*

Factor analysis identifies linear relationships among subsets of examined variables and this helps to reduce the dimensionality of the initial database without substantial loss of information. First, a factor analysis was applied to the initial dataset consisting of 5 meteorological parameters as explaining variables on the one hand and *Ambrosia* pollen load and that of the

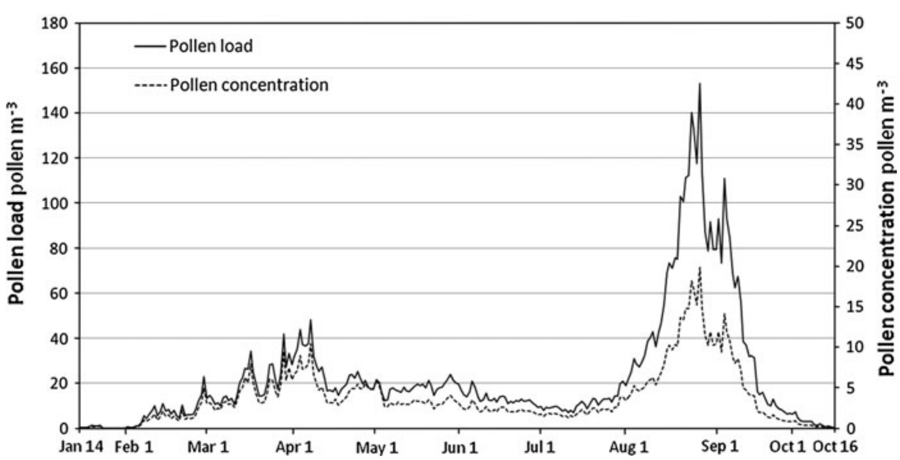

Fig. 2 The mean annual cycle of the daily pollen load and daily pollen concentration, 24 taxa, Szeged, 1999–2007

remaining pollen on the other. The procedure was performed for the two pollen variables as resultant variables separately in order to transform the original variables to fewer variables. These new variables (called factors) can be viewed as latent variables explaining the joint behaviour of weather-pollen variables. The optimum number of retained factors is determined by different statistical criteria (Jolliffe 1993). The most common and widely accepted method is to specify a least percentage (80%) of the total variance in the original variables that has to be achieved (Liu 2009). After performing the factor analysis, a special transformation of the retained factors was made to discover to what degree the above-mentioned explaining variables affect the resultant variable, and to give a rank and sign of their influence (Jahn and Vahle 1968).

2.2.2 *t test*

Quantiles corresponding to probabilities 10, 20 and 30%, furthermore 90, 80 and 70% were determined first. It should be noted that a p-quantile ($0 < p < 100\%$) q_p is the value below which the pollen load occurs with relative frequency p. The pollen loads were then assigned to two categories according to whether the actual pollen load is below or not the actual quantile. Values of daily meteorological variables corresponding to the next-day pollen load below its quantiles 10, 20 and 30% and above the quantiles 90, 80 and 70% were analysed. The Student's t test (Zimmerman 1997) was used to decide whether pollen category related means of each meteorological variable differ significantly under each quantile both for *Ambrosia* pollen and the remaining pollen.

2.2.3 *Nearest neighbour (NN) technique*

An NN technique was developed and applied in order to decide which one of the two categories of the next-day pollen load occurs under actual values of the 5 meteorological variables. A nearest neighbour of the actual daily meteorological variables is identified with the day where the explaining variables are the most similar to the actual explaining variables. Then the decision on the pollen load category for this case is the category being present on the selected day.

The procedure was used for every day available. The similarity is measured with the Euclidean distance defined with the standardised explaining variables. Standardisation is necessary to ensure the same magnitude of each explaining variable and hence to provide the same importance of them. It was performed for every explaining variable separately by dividing the difference between data and their mean by the standard deviation. Due to the annual trends in both the pollen loads and the meteorological variables, a time window h was defined around each actual day t of the year and the nearest neighbours were searched within days from $t - h$ to $t + h$ of the years. Additionally, not only the unique nearest neighbour but the first k nearest neighbours were selected and the final decision on the category was defined as the majority decision of the k number individual decisions. Parameters h and k were determined from the first 8 years (learning set) as to provide a best ratio of good decisions to all decisions, and the procedure was verified using data of the last year.

3 Results and discussion

3.1 Factor analysis with special transformation and *t* test

After performing the two-factor analyses, 4 factors were retained both for the pollen season of *Ambrosia* and the pollen season of remaining pollen excluding that of *Ambrosia*. In order to calculate the rank of importance of the explaining variables (meteorological parameters) for determining the resultant variable (pollen variables), loadings of the retained factors were projected onto Factor 1 with a special transformation (Jahn and Vahle 1968) (Tables 1, 2).

It is found that except for wind speed, the remaining four meteorological variables display significant associations with *Ambrosia* pollen load. Temperature and global solar flux indicate positive proportional, while air pressure and relative humidity inversely proportional associations with *Ambrosia* pollen loads. Explaining variables in decreasing order of their substantial influence on *Ambrosia* pollen load are temperature, air pressure, global solar flux, and relative humidity (Table 1).

The remaining pollen load excluding that of *Ambrosia* indicates notable association with all five meteorological variables (Table 2). The signs of the connections between the meteorological parameters

Table 1 Special transformation: effect of the explanatory variables on *Ambrosia* pollen load and the rank of importance of the explanatory variables on their factor loadings transformed to Factor 1 for determining the resultant variable

Variables	Weight	Rank
Ambrosia (pollen m^{-3} day^{-1})	1.000	–
Temperature (°C)	**0.153**	1
Global solar flux (W m^{-2})	**0.136**	3
Relative humidity (%)	**−0.110**	4
Air pressure (hPa)	**−0.143**	2
Wind speed (m s^{-1})	0.045	5

Rank 1 = highest weight, that is the most important meteorological variable; rank 5 = lowest weight, that is the least important variable

Thresholds of significance: italic: $x_{0.05} = 0.068$; bold: $x_{0.01} = 0.090$)

Table 2 Special transformation: effect of the explanatory variables on total pollen load excluding *Ambrosia* pollen and the rank of importance of the explanatory variables on their factor loadings transformed to Factor 1 for determining the resultant variable

Variables	Weight	Rank
Total pollen excluding *Ambrosia* (pollen m^{-3} day^{-1})	1.000	–
Temperature (°C)	**0.188**	3
Global solar flux (W m^{-2})	**0.237**	2
Relative humidity (%)	**−0.276**	1
Air pressure (hPa)	**−0.068**	5
Wind speed (m s^{-1})	**0.121**	4

Rank 1 = highest weight, that is the most important meteorological variable; rank 5 = lowest weight, that is the least important variable

Thresholds of significance: italic: $x_{0.05} = 0.041$; bold: $x_{0.01} = 0.054$

and the remaining pollen are the same as they are between the meteorological parameters and *Ambrosia* pollen (Table 1, 2). The meteorological variables thus affect the two pollen variables similarly despite their different pollen seasons. Explaining variables in decreasing order of their influence are relative humidity, global solar flux, temperature, wind speed and air pressure. Importance of the individual meteorological parameters based on their factor loadings differ in determining the two pollen variables due to their different characteristics including different length of pollen seasons and different climate requirements (Tables 1, 2).

The *t* test shows rather significant differences between means of meteorological variables corresponding to below and above the quantiles of pollen loads excluding *Ambrosia* (Table 3) potentially due to the annual trends in both the meteorological elements and the pollen load. Here, 14.65% is used instead of 10% as the relative frequency of zero loads is 14.65%. However, similar differences are less significant for the load of *Ambrosia* pollen mainly for wind speed and partially for low quantiles (Table 3). This might partly due to the fact that annual trends are not so characteristic during the relatively short pollen season of *Ambrosia*. One may suspect, therefore, that these highly significant differences are found due to just the annual cycles inherent in both the meteorological variables and pollen concentrations.

Factor analysis gave a first insight into the relationship between pollen load variables and meteorological variables and the *t* test showed, the possibility of distinguishing between extreme and non-extreme pollen events using meteorological elements as explaining variables.

3.2 NN technique

In order to clarify whether the 5 meteorological elements as explaining variables are informative to discriminate between extreme and non-extreme pollen events, an NN technique outlined in Sect. 2.2.3 was applied. The optimal time window h is 3 and 5 days for *Ambrosia* pollen load and total pollen load excluding *Ambrosia* pollen, respectively. The choice of such a small window for *Ambrosia* pollen is reasonable because the pollen load varies in a very wide range (from 0 to 5,540 in the 8 years) during a relatively short pollen season. In contrast, the load of the remaining pollen varies in a narrower range (from 0 to 3,020 in the 8 years) during a three times longer period. The optimal number k of nearest neighbours is 7 for *Ambrosia* pollen and 5 for the remaining of pollen, respectively. The larger value of k for *Ambrosia* seems to balance the narrower time window. Values of h and k were determined as to minimise the number of false decisions only for events exceeding or not exceeding the quantiles corresponding to $p_M = \max\{p, 1-p\}$ or $p_m = \min\{p, 1-p\}$, respectively,

Table 3 Results of *t* test

p	90% $q_{90} = 1{,}016$	80% $q_{80} = 552$	70% $q_{70} = 348$	10% $q_{10} = 4$	20% $q_{20} = 12$	30% $q_{30} = 28$
Ambrosia						
T	xx	xxx	xxxx	xxx	x	xx
G	xxx	xxxx	xxxx	x		
RH		xxx	xxxx		x	xx
P	xxxx	xx	xxx	xxxx	xxxx	xxxx
W				xx		
p	90% $q_{90} = 333$	80% $q_{80} = 245$	70% $q_{70} = 188$	14.65% $q_{14.65} = 1$	20% $q_{20} = 17$	30% $q_{30} = 50$
Total pollen excluding *Ambrosia*						
T		xxxx	xxxx	xxxx	xxxx	xxxx
G	xx	xxxx	xxxx	xxxx	xxxx	xxxx
RH	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
P	x	xxx	xxx	xxx	xxx	xxx
W	xxx	xxxx	xxxx	xxxx	xxxx	xxxx

Significance levels for differences between means of meteorological variables corresponding to below and above the *p* quantiles q_p of pollen loads. Symbols x, xx, xxx and xxxx refer to the 10, 5, 1 and 0.1% probability levels, respectively. *T* temperature, *G* global solar flux, *RH* relative humidity, *P* air pressure, *W* wind speed

as there is a tendency to underestimate these events and overestimate the complementary events.

Tables 4 and 5 compare the observed below or above quantile events to events obtained from NN decisions. Quantiles $p = 10$ and 90% are not included here because the number of events exceeding the quantile of 90% and not exceeding that of 10% is strongly underestimated even with the optimal time window and the number of nearest neighbours. The percentage of correct decisions is slightly over 30% for this case, while the similar percentage for complementary events (not exceeding the quantile of 90% and exceeding that of 10%) is around 97–99%. The procedure, however, works quite well for quantiles of 20 and 80%, and even better for those of 30 and 70%. The question is whether pollen loads corresponding to the quantiles of 20–30% and 70–80% can be labelled extremes. The answer is yes when taking into account the clinical threshold of pollen load. Specifically, Kadocsa et al. (1991) detected *Ambrosia* pollen sensitisation over 10 pollen grains m^{-3} air in Szeged. The quantile of 30% (pollen load of 28 for *Ambrosia*, Table 3) corresponds to 7 pollen grains/m^3 air that approximately fit the limit of 10 pollen grains/m^3 air concentration being a clinical threshold for sensitive people (Kadocsa et al. 1991). In contrast, the quantile of 80% accompanied with pollen load 552 for *Ambrosia* (Table 3) is well above the clinical threshold of pollen load, and hence, this value indicates serious adverse effects for those being sensitive for respiratory ailments.

The relative frequency of the number of decisions for exceeding the quantiles of 80, 70, 20 and 30% is 21.1, 31.1, 82.2 and 72.6% respectively in the learning set, and 20.2, 30.1, 84 and 73.4% respectively in the test set for *Ambrosia* (Table 4). Similar relative frequencies for the remaining pollen are 20.4, 29.3, 79.4 and 71.1% for the learning set and 20.4, 29.6, 80.0 and 70.8% for the test set, respectively (Table 5). These numbers explain that the NN procedure avoids substantial under or overestimation of event frequencies defined by the above quantiles, especially for the pollen load without *Ambrosia*. The relative frequency of good decisions for exceeding/not exceeding the different quantiles show that the five meteorological elements as explaining variables are informative to discriminate between extreme and non-extreme pollen events. It should be noted that the larger the percentages in above–above or below–below rows-columns in Tables 4 and 5, the better is the estimation delivered by the NN technique.

Explaining variables in decreasing order of their influence on *Ambrosia* pollen load are temperature,

Table 4 Percentages of cases when *Ambrosia* pollen load is above and below its quantiles corresponding to probability *p*

Set	*p*							
	80%		70%		20%		30%	
Learning	Above (21.1%)	Below (78.9%)	Above (31.1%)	Below (68.9%)	Above (82.2%)	Below (17.8%)	Above (72.6%)	Below (27.4%)
Above (%)	77.2	22.8	87.1	12.9	94.4	5.6	94.8	5.2
Below (%)	7.2	92.8	7.4	92.6	27.8	72.2	16.8	83.2
Verification	Above (20.2%)	Below (79.8%)	Above (30.1%)	Below (69.9%)	Above (84%)	Below (16%)	Above (73.4%)	Below (26.6%)
Above (%)	73.7	26.3	85.7	4.3	94.9	5.1	95.5	4.5
Below (%)	6.7	93.3	7.6	92.4	31.2	68.8	18.5	81.5

(A quantile assigned to probability *p* is the value below which the pollen load occurs with relative frequency *p*). Values in columns include percentages of estimated cases against observed cases (rows). Percentages in parentheses show relative frequencies of all decisions corresponding to Above/Below events under *p*-quantiles. Estimation is obtained with the nearest neighbour technique

Table 5 Percentages of cases when total pollen load except for *Ambrosia* pollen is above and below its quantiles corresponding to probability *p*

Set	*p*							
	80%		70%		20%		30%	
Learning	Above (20.4%)	Below (79.6%)	Above (29.3%)	Below (70.7%)	Above (79.4%)	Below (30.6%)	Above (71.1%)	Below (28.9%)
Above (%)	54.4	45.6	68.2	31.8	95.3	4.7	96	4
Below (%)	12	88	12.7	87.3	15.3	84.7	13	87
Verification	Above (20.4%)	Below (79.6%)	Above (29.6%)	Below (70.4%)	Above (80%)	Below (20%)	Above (70.8%)	Below (29.2%)
Above (%)	54.2	45.8	68.1	31.9	95.8	4.2	95.8	4.2
Below (%)	11.4	88.6	13.1	86.9	16.7	83.3	12.5	87.5

(A quantile assigned to probability *p* is the value below which the pollen load occurs with relative frequency *p*). Values in columns include percentages of estimated cases against observed cases (rows). Percentages in parentheses show relative frequencies of all decisions corresponding to Above/Below events under *p*-quantiles. Estimation is obtained with the nearest neighbour technique global solar flux, relative humidity, air pressure and wind speed, while on the load of the remaining pollen are temperature, relative humidity, global solar flux, air pressure and wind speed. These orders were determined with the help of the numbers of good decisions for events of exceeding or not exceeding the quantiles corresponding to p_M or p_m. The NN technique was performed with omitting a specific explaining variable from all of the five variables. This narrower set of variables delivered fewer good decisions than the entire set of variables. The larger the effect of this underlying variable on good decisions, the larger is its influence on the extreme pollen load. Performing this procedure with each variable provided an order of their importance. It should be noted that the rank of importance of the meteorological elements determining the two pollen variables partly differ from the above orders when using factor analysis with special transformation. Its reason is that factor analysis explores linear relationships among variables coming from two sources, namely the relationship between two variables is partly due to the similarity (or dissimilarity) of their annual cycles but partly due to the correlation between variations around these annual cycles. Hence, factor analysis shows an overall picture, while the NN technique reflects the relationship between daily variations of explaining variables and pollen loads excluding the annual cycles when using time windows. Additionally, application of the NN procedure allows a nonlinear relationship between explaining variables and pollen loads.

3.3 Comparison of similar studies and techniques

Finding relationships between pollen level characteristics and meteorological elements has a vast amount of literature, but extreme daily pollen concentration has received so far a relatively low attention. For instance, besides studies mentioned in the Sect. 1, Antepara et al. (1995) found that daily peak pollen values higher than 50 grains m^{-3} coincide with average daily temperatures of $18.7 \pm 3°C$. The total severity of the pollination seems to depend on the rainfall prior to the start of the pollen season. According to their model, during pollination, the days with the above temperature and an absence of rainfall between 4 and 12 h will exceed the above pollen threshold. Stach et al. (2007) found that winds coming from the east and northeast were dominant on the peak Artemisia pollen days in Poznaň, Poland. Other authors in Europe (Wahl and Puls 1989; Spieksma et al. 2000) have also indicated high influence of wind on daily peak values of Artemisia species pollen. Nevertheless, de Morton et al. (2011) developed a model for the short- and long-term prediction of atmospheric grass pollen concentrations. They found that extreme pollen events are associated with anomalous downward velocities over Melbourne. This has been demonstrated for two completely different atmospheric conditions leading to extreme pollen count increase in Melbourne.

These results, however, come from case studies or by-products of studies addressing questions not directed at extreme pollen levels. For instance, as a multiple linear regression model is defined on the entire range of observed data, one might hope that such a model is able to reproduce observed extremes, too. Unfortunately, estimates from a regression model are most accurate (have smallest variances) at moderate values and are weakening (have increasing variances) towards high and low values. In other words, such techniques are not tailored to handling extremes, and thus, a different methodology is needed to estimate extreme pollen load events. Such a method can be the NN technique applied in this paper.

Acknowledgments The authors would like to thank Miklós Juhász for providing pollen data of Szeged and Zoltán Sümeghy for the digital mapping in Fig. 1. The European Union and the European Social Fund provided financial support for the project under the grant agreement no. TAMOP 4.2.1/B-09/1/KMR-2010-0003 and TAMOP-4.2.1/B-09/1/KONV-2010-0005.

References

Antepara, I., Fernandez, J. C., Gamboa, P., Jauregui, I., & Miguel, F. (1995). Pollen allergy in the Bilbao area (European Atlantic Seaboard Climate)–pollination forecasting methods. *Clinical and Experimental Allergy, 25*(2), 133–140.

Aznarte, J. L., Sánchez, J. M. B., Lugilde, D. N., Fernández, C. D. L., de la Guardia, C. D., & Sánchez, F. A. (2007). Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. *Expert Systems with Applications, 32*(4), 1218–1225.

Cariñanos, P., Galán, C., Alcázar, P., & Domínguez, E. (2000). Allergenic pollen in the subdesert areas of the Iberian Peninsula. *Journal of Investigational Allergology and Clinical Immunology, 10*(4), 242–247.

Celenk, S., Canitez, Y., Bicakci, A., Sapan, N., & Malyer, H. (2009). An aerobiological study on pollen grains in the atmosphere of North-West Turkey. *Environmental Monitoring and Assessment, 158*(1–4), 365–380.

de Morton, J., Bye, J., Pezza, A., & Newbigin, E. (2011). On the causes of variability in amounts of airborne grass pollen in Melbourne, Australia. *International Journal of Biometeorology, 55*, 613–622.

Frei, T. (2004). Frequency of extreme events with airborne pollen. *Allergologie, 27*(1), 1–10.

Frei, T. (2006). Extreme events concerning airborne pollen. *Allergologie, 29*(9), 349–353.

Frei, T. (2008). Climate change and its impact on airborne pollen in Basel, Switzerland 1969–2007. *Allergologie, 31*(5), 165–169.

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. *International Journal of Biometeorology, 52*(7), 667–674.

Galán, C., Cariñanos, P., García-Mozo, H., Alcázar, P., & Domínguez-Vilches, E. (2001). Model for forecasting Olea europaea L. airborne pollen in South-West Andalusia, Spain. *International Journal of Biometeorology, 45*(2), 59–63.

García-Mozo, H., Galán, C., Belmonte, J., Bermejo, D., Candau, P., de la Guardia, C. D., et al. (2009). Predicting the start and peak dates of the Poaceae pollen season in Spain using process-based models. *Agricultural and Forest Meteorology, 149*(2), 256–262.

Gehrig, R. (2006). The influence of the hot and dry summer 2003 on the pollen season in Switzerland. *Aerobiologia, 22*(1), 27–34.

Hart, M. A., de Dear, R., & Beggs, P. J. (2007). A synoptic climatology of pollen concentrations during the six warmest months in Sydney, Australia. *International Journal of Biometeorology, 51*(3), 209–220.

Hirst, J. M. (1952). An automatic volumetric spore trap. *Annals of Applied Biology, 39*, 257–265.

Jahn, W., & Vahle, H. (1968). *Die Faktoranalyse und ihre Anwendung* (p. 231). Berlin: Verlag die Wirtschaft. (in German).

Jolliffe, I. T. (1993). Principal component analysis: A beginner's guide–II. Pitfalls, myths and extensions. *Weather, 48*, 246–253.

Juhász, M., & Juhász, I. E. (2002). A hazai gyomnövények aeropalinológiai jelentősége. (Aeropalynological importance of domestic weeds). *Környezeti ártalmak és a légzőrendszer (Environmental hazards and the respiratory system), 12*, 149–160. (in Hungarian).

Kadocsa, E., Bittera, L., & Juhász, M. (1991). Pollenszámlálás alapján végzett bőrtesztek eredményei nyárvégi szezonalitású rhinitis allergicás betegeken. (Results of pollen count related skin prick tests on patients suffering in rhinitis allergy in the post summer season). *Orvosi Hetilap, 32*, 1589–1591. (in Hungarian).

Kadocsa, E., & Juhász, M. (2000). A szénanáthás betegek allergénspektrumának változása a Dél-Alföldön (1990–1998) (Change of allergen spectrum of hay fever patients in Southern Great Plain (1990–1998)). *Orvosi Hetilap, 141*(29), 12617–12620 (in Hungarian).

Kasprzyk, I. (2009). Forecasting the start of Quercus pollen season using several methods—the evaluation of their efficiency. *International Journal of Biometeorology, 53*(4), 345–353.

Kasprzyk, I., & Walanus, A. (2010). Description of the main Poaceae pollen season using bi-Gaussian curves, and forecasting methods for the start and peak dates for this type of season in Rzeszow and Ostrowiec Sw. (SE Poland). *Journal of Environmental Monitoring, 12*(4), 906–916.

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

Liu, P. W. G. (2009). Simulation of the daily average PM10 concentrations at Ta-Liao with Box-Jenkins time series models and multivariate analysis. *Atmospheric Environment, 43*, 2104–2113.

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., Juhász, M., Borsos, E., & Béczi, R. (2004). Meteorological variables connected with airborne ragweed pollen in Southern Hungary. *International Journal of Biometeorology, 49*(1), 37–47.

Makra, L., Juhász, M., Mika, J., Bartzokas, A., Béczi, R., & Sümeghy, Z. (2006). An objective classification system of air mass types for Szeged, Hungary with special attention to plant pollen levels. *International Journal of Biometeorology, 50*(6), 403–421.

Makra, L., & Matyasovszky, I. (2011). Assessment of the daily ragweed pollen concentration with previous-day meteorological variables using regression and quantile regression analysis for Szeged, Hungary. *Aerobiologia, 27*, 247–259.

Makra, L., Tombácz, Sz., 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 Research, 37*(1), 99–119.

Ocana-Peinado, F., Valderrama, M. J., & Aguilera, A. M. (2008). A dynamic regression model for air pollen concentration. *Stochastic Environmental Research and Risk Assessment, 22,* S59–S63 (Supplement: Suppl. 1).

Schäppi, G., Taylor, P., Kenrick, J., Staff, I., & Suphioglu, C. (1998). Predicting the grass pollen count from meteorological data with regard to estimating the severity of hayfever symptoms in Melbourne (Australia). *Aerobiologia, 14*, 29–37.

Spieksma, F. Th. M., Van Noort, P., & Nikkels, H. (2000). Influence of nearby stands of Artemisia on street-level versus roof-top-level ratio's of airborne pollen quantities. *Aerobiologia, 16*, 21–24.

Stach, A., García-Mozo, H., Prieto-Baena, J. C., Czarnecka-Operacz, M., Jenerowicz, D., Silny, W., et al. (2007). Prevalence of Artemisia species pollinosis in Western Poland: Impact of climate change on aerobiological trends, 1995–2004. *Journal of Investigational Allergology and Clinical Immunology, 17*(1), 39–47.

Wahl, P. G., & Puls, K. E. (1989). The emission of mugwort pollen (Artemisia vulgaris L.) and its flight in the air. *Aerobiologia, 5*, 55–63.

Zimmerman, D. W. (1997). A note on interpretation of the paired-samples t test. *Journal of Educational and Behavioral Statistics, 22*(3), 349–360.