Grana, 2011; 50: 208–216

Taylor & Francis
Taylor & Francis Group

Estimating the daily Poaceae pollen concentration in Hungary by linear regression conditioning on weather types

ISTVÁN MATYASOVSZKY[1], LÁSZLÓ MAKRA[2], ZOLTÁN GUBA[2], ZSOLT PÁTKAI[3], ANNA PÁLDY[4] & ZOLTÁN SÜMEGHY[2]

[1]*Department of Meteorology, Eötvös Loránd University, Budapest, Hungary,* [2]*Department of Climatology and Landscape Ecology, University of Szeged, Szeged, Hungary,* [3]*Hungarian Meteorological Service, Budapest, Hungary,* [4]*Fodor József National Institute of Environmental Health, Budapest, Hungary*

Abstract
Objectively defined clusters of meteorological elements and weather types described by weather fronts and precipitation occurrences are produced in order to classify Poaceae pollen levels. The Poaceae pollen concentration was then estimated one day ahead for each of the above categories of days at the villages of Szeged and Győr in Hungary. The database describes an 11-year period from 1997 to 2007. For weather-front recognition purposes, the ECMWF ERA-INTERIM database was used. In order to estimate the actual daily Poaceae pollen concentrations, previous-day values of five meteorological variables and previous-day Poaceae pollen concentrations were applied. We find that both for Szeged and Győr, as well as both for the subjective and objective classifications, high daily mean Poaceae pollen levels are favoured by anticyclone ridge weather situations, as we might expect. When estimating the Poaceae pollen level, the previous-day pollen concentration, previous-day mean temperature, and previous-day mean global solar flux for Győr were statistically significant, but for Szeged it was only the previous-day pollen concentration. Taking into account the clusters, the objective classification based on original data proved the most effective. For the subjective classification, the best estimates were obtained for days with a warm front and precipitation.

Keywords: *Poaceae, grass pollen, pollination period, weather front, factor analysis, cluster analysis, ANOVA, linear regression*

Pollen allergy became a widespread medical problem by the end of the twentieth century. Nowadays, some 20% of the inhabitants on average suffer from this immune system problem in Europe (D'Amato et al., 2007). In Hungary, about 30% of the population have some type of allergy and 65% of them have pollen-sensitivity (Járai-Komlódi & Juhász, 1993; Makra et al., 2004). The pollen from grasses (Poaceae) is one of the most important airborne allergen sources worldwide (Mohapatra et al., 2005). For sensitive people, the threshold value is 30 grains/m^3, above which the symptoms of pollinosis occur (Puc & Puc, 2004). Grass pollen is the main cause of pollinosis in many countries (Subiza et al., 1995). For example, 56.7% of the people in Szeged, Hungary, are allergic to grasses (Kadocsa & Juhász, 2000), 40.4% in Thessaloniki, Greece (Gioulekas et al., 2004) and 24.0% in Hamburg, Germany (Nowak et al., 1996). The Poaceae pollen season usually begins when the average daily temperature exceeds 13.5 °C (Peternel et al., 2006). In Hungary, Poaceae have the longest blooming period of all species: it lasts from the middle of April until the middle of October (Makra et al., 2006) and, after *Ambrosia*, discharges the second biggest amount of pollen of all taxa. The ratio of its pollen release to the total pollen release in Hungary is around 17% (Juhász, 1996).

Poaceae pollen is considered to be the result of medium-range transport involving local pollen dispersion (Makra et al., 2010). The main grass

Correspondence: László Makra, Department of Climatology and Landscape Ecology, University of Szeged, PO Box 653, HU-6701 Szeged, Hungary. E-mail: *makra@geo.u-szeged.hu*

(Received 8 February 2011; accepted 27 June 2011)

ISSN 0017-3134 print / ISSN 1651-2049 online
http://dx.doi.org/10.1080/00173134.2011.602984

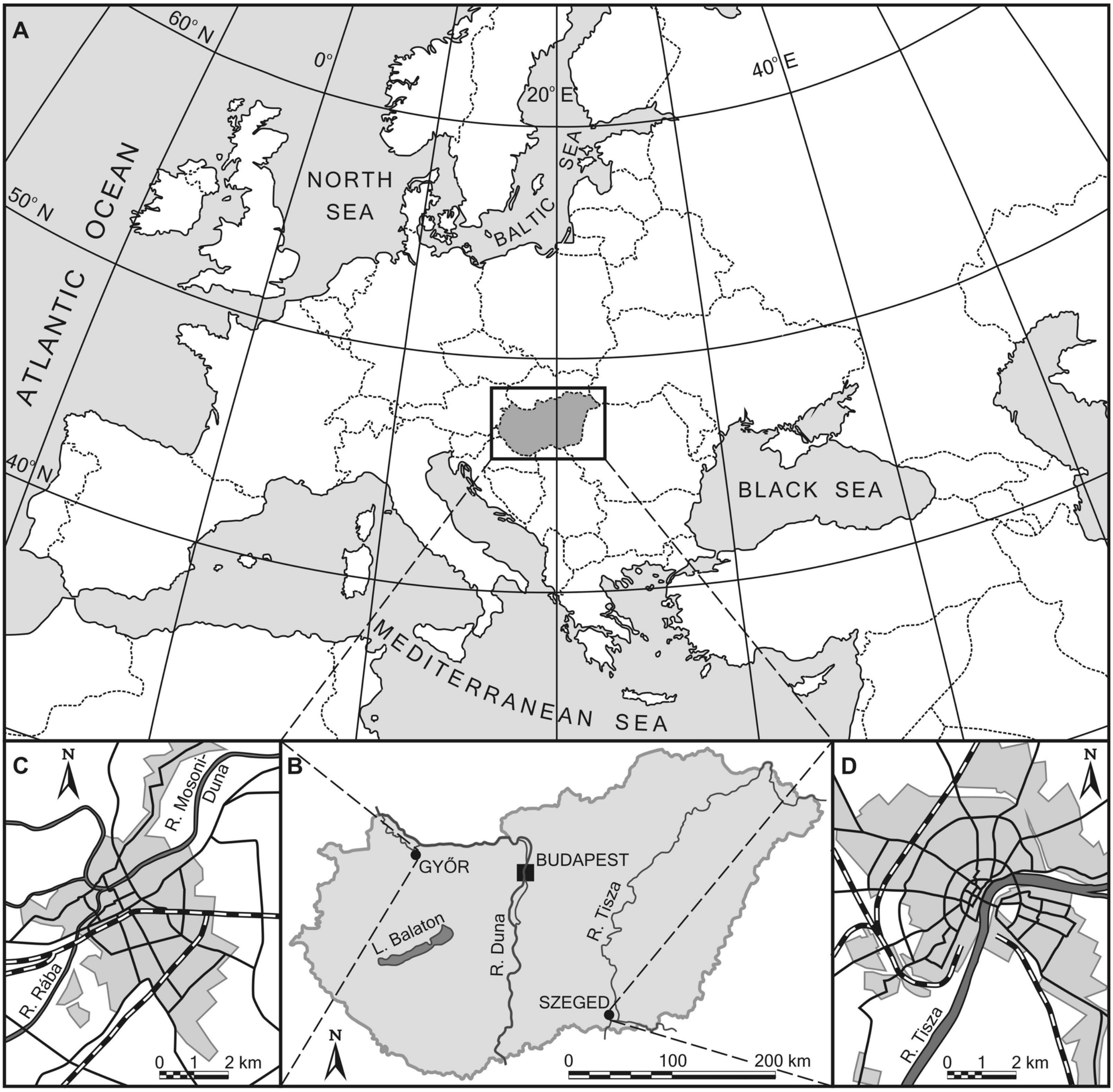

Figure 1. Geographical location of the study area. **A.** Hungary within the Carpathian Basin in Europe. **B.** The locations of Szeged and Győr. **C.** The urban area of Győr. **D.** The urban area of Szeged.

pollen season is generally double-peaked. The pattern of successive flowering in grass species and meadow cutting dates appear to be the factors, which cause the characteristic bimodal behaviour of the grass pollen season (Kasprzyk & Walanus, 2010). In Hungary, the first peak occurs from February to April, while the second peak is from May to July (Juhász, 1996).

Finding a connection between the daily Poaceae pollen concentration and daily meteorological elements is of great practical importance. Applying simple statistical analyses, several studies found significant positive correlations between daily Poaceae pollen concentration and daily maximum temperature (Valencia-Barrera et al., 2001; Green et al., 2004; Kasprzyk & Walanus, 2010), daily minimum temperature (Green et al., 2004), daily mean temperature (Puc & Puc, 2004; Peternel et al., 2006; Kasprzyk & Walanus, 2010), and daily global solar flux (Valencia-Barrera et al., 2001; Kasprzyk & Walanus, 2010). The relative humidity (Valencia-Barrera et al., 2001; Puc & Puc, 2004; Peternel et al., 2006; Kasprzyk & Walanus, 2010) and rainfall (Fehér & Járai-Komlódi, 1997; Valencia-Barrera et al., 2001; Green et al., 2004; Puc & Puc, 2004; Peternel et al., 2006; Kasprzyk & Walanus, 2010), however, had a negative effect. Wind speed plays an ambivalent role, partly having a positive impact by increasing pollen shed from the anthers, partly a negative association by diluting pollen from the air (Valencia-Barrera et al., 2001).

Meteorological elements affect pollen concentration not by means of their individual effects, but through their interrelationships and so it is useful to study the connection between daily Poaceae pollen concentrations and the daily values of meteorological elements as a whole. Only relatively few papers have reported results of approaches like these using multivariate statistical analysis techniques. Makra et al. (2006) objectively defined weather types with factor and cluster analyses in order to associate given daily pollen concentrations with their representative meteorological parameters. Hart et al. (2007) analysed the influence of weather elements on pollen concentrations for Sydney, Australia, producing a synoptic classification of pollen concentrations using principal component analysis and cluster analysis. Tonello and Prieto (2008) classified pollen data of 17 taxa and climate parameters using principal component analysis and cluster analysis to identify relationships between potential natural vegetation, pollen, and climate.

The purpose of this paper is to determine the most homogeneous groups of meteorological elements by cluster analysis, and subsequently, Poaceae pollen levels are examined by conditioning on these clusters. Another aim is to utilise the possible information of meteorological elements on Poaceae pollen concentration when cold and warm weather fronts pass through a city. Furthermore, those days with a front but no rain and those days with rain but no front will also be considered. Lastly, we attempt to estimate the Poaceae pollen concentration one day ahead for each of the above day categories at Szeged and Győr. These medium-sized cities are located in south-eastern and north-western Hungary, respectively, about 340 km apart (Figure 1).

Material and methods

Location and data

Szeged (46° 15′ 14″ N, 20° 08′ 46″ E), the largest settlement in south-eastern Hungary, is located at the confluence of the rivers Tisza and Maros (Figure 1). The area is characterised by an extensive flat landscape of the Great Hungarian Plain with an average elevation of 79 m above sea level (a.s.l.). 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 seven-day recording 'Hirst-type' volumetric trap (Hirst, 1952). The air sampler was located on top of the building of the Faculty of Arts at the University of Szeged some 20 m above the ground surface (Makra et al., 2008).

Győr (47° 41′ 02″ N, 17° 38′ 06″ E) lies in northern Transdanubia of Hungary. The city is located in the Rába valley with an average elevation of 123 m a.s.l. at the confluence of the Rába Rápca and Mosoni-Duna rivers (Figure 1). Győr is the sixth largest city in Hungary with a population of 132 000. In the Köppen system, its climate is of the Cbfx type. It has a temperate oceanic climate with mild winters and cool-to-warm summers as well as a near-uniform annual precipitation distribution with early summer maximum (Köppen, 1931). The air sampler was situated on the roof of the Oncology Department building, Petz Aladár County Hospital, south of the centre of Győr, approximately 20 m above ground level.

Meteorological data were collected in monitoring stations (operated by the Environmental and Natural Protection and Water Conservancy Inspectorates of Lower-Tisza Region, Szeged, and Northern Transdanubia, Győr, respectively) located in the downtown area of Szeged and Győr at a distance of about 10 m from the busiest main roads.

For weather front recognition purposes, the ECMWF (European Centre for Medium-Range Weather Forecasts) ERA-INTERIM database was used on the grid network with a density of 1° × 1°. Only those grid points (for Szeged: 46° N, 20° E; for Győr: 48° N, 18° E) were selected that were nearest to the geographical coordinates of both Szeged and Győr. Meteorological parameters considered are as follows: RT 500/850 hPa relative topographies (for calculating the thickness of the air-layer), 700 hPa temperature fields (for calculating the temperature advection in order to determine the sign of the front) and 700 hPa wind fields (to include advection when calculating the thermal front parameter, TFP).

In order to estimate actual daily Poaceae pollen concentrations, previous-day values of five meteorological variables (mean temperature, mean global solar flux, mean relative humidity, mean sea-level pressure, mean wind speed) and previous-day Poaceae pollen concentrations as candidate predictors were applied. The statistical relationship between the predictors and the pollen concentration was conditioned on the clusters as well as frontal and precipitation information. The rationale behind this approach is that the existence of fronts and precipitation is forecasted at a high accuracy, but the forecast of the above meteorological variables is much less reliable. Hence, previous-day values were considered instead of their values forecasted for the actual days in question.

The longest pollination period of the two cities (4 April–16 October, 1997–2007) was used for this study. This interval covers most of the Poaceae pollination period both in Szeged and Győr 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 one pollen grain/m^3 is recorded and at least five consecutive (preceding) days also have one or more pollen grains/m^3. The mean of this annually varying period was selected for the 11-year period examined.

Objective identification of weather fronts

The objective identification of weather fronts is a difficult task due to the lack of a unique definition described in mathematical terms. In addition, different fronts can be characterised by different sets of meteorological parameters. The contribution of Renard and Clarke (1965) is a classic advance in this area. The authors analysed the horizontal gradient of magnitude of the horizontal potential temperature gradient at an air pressure of 850 hPa. Later, Hewson (1998) improved the procedure, resulting in the quantity TFP. Using this methodology, Yan et al. (2008) produced an automatic weather system identification method that can identify weather systems with 80 to 100% accuracy and provide objective information on identifying and positioning weather systems. There is a clear relationship between the TFP as frontal analysis parameter and the well-known basic front definition that fixes a cold front where the temperature begins to fall and a warm front where the rise in temperature ends. This definition corresponds to the maximum of the TFP.

Frequently, the relative topography is used for simplicity instead of the potential temperature. This is because there is a close connection between the thickness and the average potential temperature of an air-layer. Therefore, the relative topography (RT 500/850 hPa) was used in our calculations. The algorithm was run for six-hourly datasets within the period examined for both cities. Dates when fronts passed through the cities were thus produced with a six-hour resolution.

Statistical methods

Factor analysis (FA) explains linear relationships among the variables examined and reduces the dimensionality of the initial database without a substantial loss of information. In particular, the six variables (five climate variables and the previous-day pollen concentration) were transformed into *m* number of factors. These factors can be viewed as main latent variables potentially influencing daily pollen concentration. The optimum number *m* of the retained factors is defined such that the total variance of the *m* factors reaches a pre-specified portion (80% in our case) of the total variance of the original variables (Jolliffe, 1993).

A cluster analysis was applied to the original data sets that objectively classifies the days of the given groups with similar climate conditions. We applied hierarchical cluster analysis using Ward's method on the climatic variables of the period 4 April–16 October over the 11-year period examined. Ward's method attempts to minimise the sum of squares of elements within clusters forming at each step during the procedure (Ward, 1963). The procedure works with the Mahalanobis metric, which is deemed better than the Euclidean metric (Mahalanobis, 1936). The Mahalanobis metric takes into account the different standard deviations of the components of the vectors to be clustered as well as the correlations among the components. We select the number of clusters under possible cluster numbers from 3 to 30 so as to ensure nearly uniform occurrence frequencies of the clusters. Intuitively, the final system of clusters produces a small variation of occurrence frequencies of the clusters constrained on forming these clusters by Ward's method (Anderberg, 1973; Hair et al., 1998).

Another classification is based on frontal information. In particular, the following six weather types were defined: (1) warm front with rain, (2) warm front with no rain, (3) cold front with rain, (4) cold front with no rain, (5) no front with rain, and (6) no front with no rain.

The one-way analysis of variance (ANOVA) was then used to decide whether the inter-cluster variance was significantly higher than the intra-cluster variance of daily Poaceae concentrations. A *post hoc* Tukey test was applied to find the clusters that differ significantly from others (Tukey, 1953) from the viewpoint of the cluster-dependent mean pollen levels.

A further task was to establish a relationship between predictors and the pollen concentration. As both kinds of variables exhibit annual trends, standardised data sets were used. Denoting an underlying data set by $x_t, t = 1, \ldots, n$ the expected value function $m(t)$ of x_t is approximated by a linear combination of cosine and sine functions with periods of one year and one half year (note that the latter cycle was introduced to describe the asymmetries of the annual courses). Namely, $m(t) = a_0 + a_1 \cos(w_1 t) + a_2 \cos(w_2 t) + b_1 \sin(w_1 t) + b_2 \sin(w_2 t)$ with $w_1 = 2\pi/365.25$ and $w_2 = 2w_1$. Unknown coefficients in this linear combination were estimated via the least squares technique. Then the standardised and, thus, annual course-free data set is $y_t = (x_t - m(t))/d(t), t = 1, \ldots, n$, where the unknown coefficients in $d^2(t) = a_0 + a_1 \cos(w_1 t) + a_2 \cos(w_2 t) + b_1 \sin(w_1 t) + b_2 \sin(w_2 t)$ were estimated like those in $m(t)$, except that x_t was replaced by $x_t^* = (x_t - m(t))^2, t = 1, \ldots, n$ when applying the least squares technique. Linear regressions for both the entire data set and data sets corresponding to both systems of weather types were performed. The order of importance of predictors in the formation of pollen concentration was determined by the well-known stepwise regression method (Draper & Smith, 1981). The mean squared error obtained for the entire data set was compared to the weighted sum of mean squared errors obtained for the weather types. The weights were the relative frequencies of these types.

Results

Cluster analysis and ANOVA

The days of the 11-year period for both Szeged and Győr (Tables I and II) were classified into the earlier-mentioned six categories. Afterwards, cluster analyses were carried out for the two cities using the original and standardised data sets. Hence, altogether, four cluster analyses were performed (Tables I and II).

The analysis of variance revealed a significant difference at least at a 95% probability level in the mean values of Poaceae pollen levels among the individual clusters. For Szeged, using frontal categories, pairwise comparisons of the cluster averages found two significant differences among the possible 15 cluster pairs of six clusters (13.3%). In this case, only mean values of clusters 4 and 5, as well as 4 and 6, differed significantly from each other. The clustering of original variables resulted in 14 significant pairwise differences among the mean pollen levels of all 21 cluster pairs of seven clusters (66.7%). Clusters defined with standardised variables revealed seven significant differences among the possible 21 cluster pairs of seven clusters (33.3%). The role of cluster 2 is important because its average differed significantly from those of the remaining clusters.

For Győr, using frontal categories, only three significant differences were found among the possible 15 cluster pairs of six clusters (20%). An objective classification of original variables resulted in seven clusters. Here, with the exception of clusters 2 and 5, 2 and 6, and 4 and 7, the means of all remaining cluster pairs notably differed from each other (85.7%). For standardised variables, six objective clusters and 11 significant pairwise differences among the mean pollen levels of all 15 cluster pairs (73.3%) were obtained.

Only clusters of the significantly different cluster averaged pollen levels will be considered later and analysed in detail, principally, the clusters with extreme pollen levels (Tables I and II). Comparing the results obtained by frontal categories and objective clusters (Table I) defined on the original data for Szeged, the following key conclusions can be drawn. The category (cluster) displaying the highest mean pollen level includes a medium (fewest) number of days. The frontal category (cluster) involving the highest mean pollen level includes (does not include) extreme values of the influencing variables. For the subjective classification, the category of the highest mean pollen level (cluster 4: cold front with no rain) is influenced by high previous-day mean Poaceae pollen concentration, highest mean values of temperature and air pressure as well as low relative humidity and the lowest wind speed. These values of the meteorological parameters assume an ante-cold front weather situation that is a possible anticyclone ridge weather situation, which facilitates high pollen levels (Table I).

For the objective classification, the cluster of the highest mean Poaceae pollen concentration (cluster 5) is affected by the highest previous-day pollen level, low temperature as well as high global solar flux, relative humidity, air pressure, and wind speed. These values suppose a weak anticyclone ridge weather situation promoting the enrichment of Poaceae pollen (Table I).

A similar comparison for Győr (Table II) reveals the following main points: the subjective category involving the highest mean pollen concentration occurs frequently, while the frequency of the objective cluster having the highest mean pollen level is the smallest. The subjective category having the highest mean pollen level is associated with the second highest previous-day mean Poaceae pollen concentration, while the objective cluster displaying the highest value of the resultant variable involves the highest mean previous-day pollen level as its apparently most important influencing variable. For the subjective classification, the category of the highest mean pollen level (cluster 2: warm front with no rain) is influenced by the highest temperature and wind speed, high global solar flux and air pressure as well as low relative humidity. These values assume an ante-warm front weather situation that is possibly again an anticyclone ridge weather situation that aids high pollen levels (Table II). For the objective classification, the cluster of the highest mean Poaceae pollen concentration (cluster 3) is probably mostly affected by the highest previous-day pollen level, since no meteorological variables have extreme values. Nevertheless, the temperature and global solar flux are high, while the relative humidity is low; furthermore, air pressure

Table I. Mean values (**bold**, maximum; *italic*, minimum) of the meteorological parameters and pollen concentrations for Szeged.

	Cluster[a]						
	1	2	3	4	5	6	
Subjective categories, original data							
Total number of days	123	320	*46*	203	211	**484**	
Frequency (%)	8.9	23.1	*3.3*	14.6	15.2	**34.9**	
Temperature (°C)	19.4	20.8	19.0	**20.9**	*17.8*	19.5	
Global solar flux (W/m²)	259.3	**361.6**	270.1	340.5	*250.1*	357.5	
Relative humidity (%)	75.2	*63.7*	77.6	64.7	**80.7**	65.8	
Air pressure (hPa)	1000.6	**1006.1**	1002.1	**1006.1**	*1001.1*	1004.5	
Wind speed (m/s)	**1.1**	1.0	**1.1**	*0.9*	1.0	1.0	
Poaceae pollen, previous day (pollen/m³/day)	16.7	14.7	**18.2**	17.7	14.5	*14.4*	
Poaceae pollen, same day (pollen/m³/day)	15.8	15.9	17.2	**17.7**	*13.2*	14.6	
Objective clusters, original data, whole database							
Total number of days	174	249	232	137	*82*	**343**	170
Frequency (%)	12.5	18.0	16.7	9.9	*5.9*	**24.7**	12.3
Temperature (°C)	**22.4**	16.2	*15.8*	22.3	19.3	23.3	18.2
Global solar flux (W/m²)	285.7	*165.3*	386.3	349.1	350.5	**413.5**	329.9
Relative humidity (%)	54.8	**89.5**	64.4	*49.5*	68.3	69.3	72.3
Air pressure (hPa)	**1012.7**	1000.3	1007.5	*997.1*	1005.0	1003.2	1003.9
Wind speed (m/s)	0.9	0.9	*0.8*	1.0	1.0	0.9	**1.7**
Poaceae pollen, previous day (pollen/m³/day)	14.7	*8.7*	9.1	12.1	**66.6**	14.7	13.2
Poaceae pollen, same day (pollen/m³/day)	14.9	*8.4*	10.8	12.9	**55.3**	15.4	14.9

[a] 1, warm front with rain; 2, warm front with no rain; 3, cold front with rain; 4, cold front with no rain; 5, no front with rain; 6, no front with no rain.

and wind speed assume medium values. Accordingly, this cluster is supposed to be prevailed by a weak anticyclone ridge weather situation that assists the enrichment of Poaceae pollen (Table II).

Linear regression

Predictors significant at least at a 90% level were retained and evaluated. These include previous-day pollen concentration, previous-day mean temperature, and previous-day mean global solar flux for Győr, but only previous-day pollen concentration for Szeged when the entire data set was used. The previous-day concentration is apparently the best predictive parameter for both locations and for each weather type defined either by clustering or fronts with rain occurrences. The rank of importance of temperature and global solar flux for Győr is variable under both sets of clusters and even the wind speed has a slight (although statistically significant) role in types with fronts.

The ratio of the variance explained by these variables to the variance of the pollen concentration is 12.8% and 29.9% for Győr and Szeged, respectively, using all data sets. With cluster-dependent regressions, Győr provides an explained variance of 27.9%, 17.0%, and 16.3% using clusters with original data, clusters with standardised data, and subjective weather types, respectively. The most effective classification is, therefore, the clustering with original data. A subjective classification and clustering with standardised data do not make any substantial difference. The corresponding values for Szeged under cluster-dependent regressions are 31.0%, 29.6%, and 29.7%. Here, the different classifications perform quite similarly, especially the subjective classification and clustering with standardised data. Note that the estimation of daily Poaceae pollen concentration is more reliable for Szeged than for Győr because it has a warmer and dryer climate. However, the cluster-dependent regression for Szeged, in contrast to Győr, yields only a slight gain under clustering with original data and yields no gain under subjective classification and clustering with standardised data.

For the subjective classification, the best estimates can be obtained for days of warm front with precipitation (39.0% and 70.4% explained variance for Győr and Szeged, respectively), while the poorest estimates occur on days of cold front with no precipitation. Under this type, the mean squared error is larger than this quantity defined for the entire data set (by 20% and 13% for Győr and Szeged, respectively). For clustering with original data, the best estimates can be obtained for cluster 3 and cluster 4 for Szeged (74.6% explained variance) and

Table II. Mean values (**bold**, maximum; *italic*, minimum) of the meteorological parameters and pollen concentrations for Győr.

	Cluster[a]						
	1	2	3	4	5	6	
Subjective categories, original data							
Total number of days	90	264	*53*	145	168	**416**	
Frequency (%)	7.9	23.2	*4.7*	12.8	14.8	**36.6**	
Temperature (°C)	17.8	**19.9**	18.1	19.8	*16.6*	18.7	
Global solar flux (W/m^2)	*145.1*	196.5	154.7	193.8	149.0	**201.7**	
Relative humidity (%)	72.8	*60.7*	69.8	60.4	**74.8**	61.2	
Air pressure (hPa)	1002.2	1004.8	1003.5	1005.1	*1001.7*	**1006.1**	
Wind speed (m/s)	**1.2**	**1.2**	**1.2**	*1.1*	**1.2**	*1.1*	
Poaceae pollen, previous day (pollen/m^3/day)	7.1	7.7	**9.3**	7.7	*5.8*	6.1	
Poaceae pollen, same day (pollen/m^3/day)	6.6	**7.9**	6.2	7.4	*5.2*	7.0	
Objective clusters, original data, whole database							
Total number of days	214	102	*80*	139	195	**217**	189
Frequency (%)	18.8	9.0	*7.0*	12.2	17.2	**19.1**	16.6
Temperature (°C)	20.1	17.3	20.6	*14.1*	18.2	**23.6**	15.3
Global solar flux (W/m^2)	**295.8**	168.6	224.2	118.3	178.7	187.0	*105.2*
Relative humidity (%)	59.8	69.2	61.9	75.6	*49.9*	60.2	**78.9**
Air pressure (hPa)	1008.2	999.5	1004.3	**1012.8**	1003.7	1003.9	*999.1*
Wind speed (m/s)	*1.0*	**1.9**	1.2	1.1	1.1	1.1	1.1
Poaceae pollen, previous day (pollen/m^3/day)	7.1	3.9	**37.2**	*2.3*	5.4	4.1	3.5
Poaceae pollen, same day (pollen/m^3/day)	10.0	5.0	**24.6**	*2.2*	6.9	5.0	2.7

[a] 1, warm front with rain; 2, warm front with no rain; 3, cold front with rain; 4, cold front with no rain; 5, no front with rain; 6, no front with no rain.

Győr (42.8% explained variance), respectively. The poorest estimates are associated with cluster 3 for Győr and cluster 7 for Szeged. The corresponding mean squared error increases by 106.5% (Győr) and 55.4% (Szeged) relative to the quantity defined for the entire data set.

The variability of mean pollen concentrations among clusters is wider than the similar range under subjective weather types. This is due to the fact that a subjective grouping might consider fewer influencing variables or selects the types arbitrarily, while an objective classification makes it possible to select an optimum number of groups of weather types, providing a higher explained variance of the pollen level (Makra et al., 2009). The category of warm front with precipitation involves very similar days of a relatively stable weather situation. Though the wind speed is high, the influencing variables display a small variability, producing a higher explained variance of the Poaceae pollen concentration. On the contrary, on days of cold front with no rain in spite of low winds the higher variability of the influencing variables involves a lower explained variance of the pollen level (Tables I and II). The estimation of the pollen level for Szeged is more reliable than for Győr. The warm and temperate climate (Köppen's Ca type) of Szeged fits the climate optimum of Poaceae better than temperate oceanic climate (Köppen's Cbfx type) of Győr. This is why the explained variance of the six influencing variables to the variance of the pollen level is higher for Szeged than for Győr.

Conclusions

Both for Szeged and Győr, as well as for the subjective and objective classifications, high daily mean Poaceae pollen levels are facilitated by anticyclone ridge weather situations, as we expected. When estimating Poaceae pollen level, the previous-day pollen concentration, mean temperature, and mean global solar flux for Győr, but only previous-day pollen concentration for Szeged were significant at least at a 90% level using the entire data set. Regarding clusters with original data, clusters with standardised data, and subjective weather types, the objective classification with original data proved the most effective.

For the subjective classification, the best estimates can be gained for days of warm front with precipitation for Győr and Szeged. The poorest estimates are obtained on days of cold front with no precipitation. The ratio of the variance explained by the six influencing variables to the variance of the pollen concentration was higher for the objective than for the subjective classification of the weather types, which confirms the expectations. The difference between explained variances of the Poaceae pollen concentration under the category of warm front with precipitation and under cold front with no precipitation can be explained by the different variability of the influencing variables. The estimates of the pollen level for Szeged is much more reliable than those for Győr, since the warm and temperate climate of Szeged fits the climate optimum of Poaceae better than the temperate oceanic climate of Győr.

The earlier-mentioned relationship between the pollen level and the six influencing variables allow a preliminary, cluster related analysis of variable dependencies for the pollen level to be performed. In order to get a reliable forecast of the Poaceae pollen concentration, a more advanced methodology will be needed. However, both the objectively and subjectively defined weather types produce useful information on the accuracy of the forecast. For instance, the Poaceae pollen concentration following a warm front with rain can be accurately forecasted, while a cold front with no rain involves highly uncertain factors.

Acknowledgements

The authors would like to thank Gábor Motika (Environmental and Natural Protection and Water Conservancy Inspectorate of the Lower-Tisza Region, Szeged, Hungary), for providing meteorological data of Szeged, and Péter Lautner (Environmental and Natural Protection and Water Conservancy Inspectorate of Northern Transdanubia, Győr, Hungary), for their part in collecting meteorological data for Győr, and Miklós Juhász (University of Szeged, Department of Botany) for providing pollen data of Szeged. The European Union and the European Social Fund provided financial support for the project under the grant agreement numbers TÁMOP 4.2.1/B-09/1/KMR-2010-0003 and TÁMOP-4.2.1/B-09/1/KONV-2010-0005.

References

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

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*. New York: John Wiley & Sons.

Fehér, Z. & Járai-Komlódi, M. (1997). An examination of the main characteristics of the pollen seasons in Budapest, Hungary (1991–1996). *Grana*, *36*, 169–174.

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*, 59–63.

Gioulekas, D., Papakosta, D., Damialis, A., Spieksma, F., Giouleka, P. & Patakas, D. (2004). Allergenic pollen records (15 years) and sensitisation in patients with respiratory allergy in Thessaloniki, Greece. *Allergy*, *59*, 174–184.

Green, B. J., Dettmann, M., Yli-Panula, E., Rutherford, S. & Simpson, R. (2004). Atmospheric Poaceae pollen frequencies and associations with meteorological parameters in Brisbane, Australia: a 5-year record, 1994–1999. *International Journal of Biometeorology*, *48*, 172–178.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998). *Multivariate data analysis.* Englewood Cliffs, NJ: Prentice Hall.

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*, 209–220.

Hewson, T. D. (1998). Objective fronts. *Meteorological Applications*, *5*, 37–65.

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

Járai-Komlódi, M. & Juhász, M. (1993). *Ambrosia elatior* (L.) in Hungary (1989–1990). *Aerobiologia*, *9*, 75–78.

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

Juhász, M. (1996). Pollen as an indicator of environmental pollution (A pollen, mint környezetszennyezést jelző indikátor). *2nd Symposium on Analytical and Environmental Problems*, SZAB, Szeged, Hungary, 4 November 1996, book of abstracts, 152–159 (in Hungarian).

Kadocsa, E. & Juhász, M. (2000). A szénanáthás betegek allergénspektrumának változása a Dél-Alföldön (1990–1998). *Orvosi Hetilap*, *141*, 12617–12620 (in Hungarian).

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*, 906–916.

Köppen, W. (1931). *Grundriss der Klimakunde*. Berlin: De Gruyter & Co. (in German).

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, *12*, 49–55.

Makra, L., Juhász, M., Borsos, E. & Béczi, R. (2004). Meteorological variables connected with airborne ragweed pollen in Southern Hungary. *International Journal of Biometeorology*, *49*, 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*, 403–421.

Makra, L., Mika, J., Bartzokas, A., Béczi, R. & Sümeghy, Z. (2009). Comparison of objective air-mass types and the Péczely weather types and their ability to classify levels of air pollutants in Szeged, Hungary. *International Journal of Environment and Pollution, "Air Pollution" Special Issue*, *36*, 81–98.

Makra, L., Sánta, T., Matyasovszky, I., Damialis, A., Karatzas, K., Bergmann, K. C. & Vokou, D. (2010). Airborne pollen in three European cities: detection of atmospheric circulation pathways by applying three-dimensional clustering of backward trajectories. *Journal of Geophysical Research-Atmospheres*, *115*, D24220. DOI: 10.1029/2010JD 014743

Makra, L., Tombácz, S., 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*, 99–119.

Mohapatra, S. S., Lockey, R. F. & Shirley, S. (2005). Immunobiology of grass pollen allergens. *Current Allergy and Asthma Reports*, *5*, 381–387.

Nowak, D., Heinrich, J., Jorres, R., Wassmer, G., Berger, J., Beck, E., Boczor, S., Claussen, M., Wichmann, H. E. & Magnussen, H. (1996). Prevalence of respiratory symptoms, bronchial hyperresponsiveness and atopy among adults: West and East Germany. *European Respiratory Journal*, *9*, 2541–2552.

Peternel, R., Srnec, L., & Čulig, J., Hrga, I. & Hercog, P. (2006). Poaceae pollen in the atmosphere of Zagreb (Croatia), 2002–2005. *Grana*, *45*, 130–136.

Puc, M. & Puc, M. I. (2004). Allergenic airborne grass pollen in Szczecin, Poland. *Annals of Agricultural and Environmental Medicine*, *11*, 237–244.

Renard, R. J. & Clarke, L. C. (1965). Experiment in numerical objective frontal analysis. *Monthly Weather Review*, *93*, 547–556.

Subiza, J., Jerez, M., Jiménez, J. A., Narganes, M. J., Cabrera, M., Varela, S. & Subiza, E. (1995). Allergenic pollen and pollinosis in Madrid. *Journal of Allergy and Clinical Immunology*, *96*, 15–23.

Tonello, M. S. & Prieto, A. R. (2008). Modern vegetation–pollen–climate relationships for the Pampa grasslands of Argentina. *Journal of Biogeography*, *5*, 926–938.

Tukey, J. W. (1953). The problem of multiple comparisons. In H. I. Braun (Ed., 1984), *The collected works of John W. Tukey VIII. Multiple comparisons*: 1948–1983 (pp. 1–300). New York: Chapman & Hall/CRC.

Valencia-Barrera, R. M., Comtois, P. &Fernández-González, D. (2001). Biogeography and bioclimatology in pollen forecasting – an example of grass in León (Spain) and Montreal (Canada). *Grana*, *40*, 223–229.

Ward, J. H. (1963). Hierachical grouping to optimise an objective function. *Journal of American Statistical Association*, *58*, 236–244.

Yan, W. K., Lap, Y. C. & Wah, L. P. (2008). Automatic identification of weather systems from numerical weather prediction data using genetic algorithm. *Expert Systems with Applications*, *35*, 542–555.