

L. Makra · M. Juhász · E. Borsos · R. Bécsi

Meteorological variables connected with airborne ragweed pollen in Southern Hungary

Received: 24 July 2003 / Accepted: 26 February 2004 / Published online: 22 April 2004
© ISB 2004

Abstract About 30% of the Hungarian population has some type of allergy, 65% of them have pollen sensitivity, and at least 60% of this pollen sensitivity is caused by ragweed. The short (or common) ragweed (*Ambrosia artemisiifolia* = *Ambrosia elatior*) has the most aggressive pollen of all. Clinical investigations prove that its allergenic pollen is the main reason for the most massive, most serious and most long-lasting pollinosis. The air in the Carpathian Basin is the most polluted with ragweed pollen in Europe. The aim of the study is to analyse how ragweed pollen concentration is influenced by meteorological elements in a medium-sized city, Szeged, Southern Hungary. The data basis consists of daily ragweed pollen counts and averages of 11 meteorological parameters for the 5-year daily data set, between 1997 and 2001. The study considers some of the ragweed pollen characteristics for Szeged. Application of the Makra test indicates the same period for the highest pollen concentration as that established by the main pollination period. After performing factor analysis for the daily ragweed pollen counts and the 11 meteorological variables examined, four factors were retained that explain 84.4% of the total variance of the original 12 variables. Assessment of the daily pollen number was performed by multiple regression analysis and results based on deseasonalised and original data were compared.

Keywords Pollen allergy · Ragweed pollen concentration · Makra test · Factor analysis · Regression analysis

Introduction

Pollen allergy

Pollen allergy had become a widespread disease by the end of the 20th century. Nowadays, every 5th or 6th person, on average, suffers from this disease of the immune system in Europe. Pollinosis involves unpleasant symptoms and can become asthma. It has been proved that those who fall ill with pollen allergy can not concentrate on their work, feel unwell and are frequently on sick leave.

About 30% of the Hungarian population has some type of allergy, 65% of them have pollen sensitivity, and at least 60% of this pollen sensitivity is caused by ragweed. It is a shocking fact that the number of patients with registered allergic illnesses has doubled and the number of cases of allergic asthma has increased by a factor of four in Southern Hungary over the last 40 years. However, we have to remember that the diagnosis of asthma has also developed significantly during this period.

The aim of the study is to analyse the relation of ragweed pollen concentration to meteorological elements in a medium-sized city: Szeged, Southern Hungary.

The main plants causing pollen allergy in Europe are grasses (Poaceae), birch (*Betula*), mugwort (*Artemisia*) and, in Southern Europe, the olive-tree (Oleaceae). In the 1980s a new plant that spreads extremely aggressively joined the list. It appears in more and more countries, blooms for a long time (in some cases for 3 months) and it produces a great deal of pollen, which when breathed-in, rapidly produces characteristic symptoms of pollinosis (coughing, sneezing, nasal discharge and inflammation of the mucous membranes of the eyes and nose). The plant is the short (or common) ragweed (*Ambrosia artemisiifolia* = *Ambrosia elatior*) and clinical investigations have demonstrated that its very allergenic pollen is the main reason for the most massive, most serious and most long-lasting pollinosis.

Ambrosia was the name given to the delicious food eaten by the mythical Greek gods to make them live

L. Makra (✉) · E. Borsos · R. Bécsi
Department of Climatology and Landscape Ecology,
University of Szeged, H-6701 Szeged, P.O.B. 653, Hungary
e-mail: makra@geo.u-szeged.hu
Tel.: +36-62-544856
Fax: +36-62-544624

M. Juhász
Department of Botany, University of Szeged, H-6701 Szeged,
P.O.B. 657, Hungary

forever. Though the Latin name for ragweed is also *Ambrosia* it bears no resemblance to the heavenly food, the only food of newborns of gods. It is mentioned in the literature either as a food or as a drink and one imagines that it was some kind of honey. It was also believed to be effective for curing wound. Of course, the Greek gods did not eat ragweed.

Ragweed is an important genus of the Asteraceae family, which probably has its origin in Southern North America. This plant has evolved in response to a dry climate and open environment and from an entomophilous to an anemophilous pollination. In Europe, a steppe-prairie dry environment is found in the east-central part of the continent which, like ragweed's region of origin, favours for its growth. The presence of ragweed here is recorded in botanical writings of the 19th century. However, it became part of the human landscape in Europe after the First World War, when contaminated seed shipments to the former Austro-Hungarian Empire allowed the plant to establish itself on this part of the continent. Ragweed can take hold and prosper, if the climatic threshold for its floral and seed development is reached, i.e. up to 1,200 °C heating degree days in total. Under a maritime climate, e.g. in North-Eastern Spain or in the Netherlands, ragweed populations seem to just survive and do not seem able to reach population levels that would make them noxious. Hence, it is the climate that limits ragweed, even if the human environment would allow its establishment (Comtois 1998).

In Western Europe, four American species have already become established. These are ragweed with mugwort leaves (= short ragweed; *Ambrosia artemisiifolia* = *Ambrosia elatior*), giant ragweed (*Ambrosia trifida*), perennial ragweed (*Ambrosia psilostachya*) and silver ragweed (*Ambrosia tenuifolia*). However, short ragweed is the most widely spread of all, not only in Western Europe but in Hungary too (Járai-Komlódi and Juhász 1993).

Considering the annual total pollen counts of various plants measured between 1990 and 1996 in Southern Hungary, ragweed generates about half of the total pollen production (47.3%). Though this ratio depends strongly on meteorological factors year by year (in 1990 this ratio was 35.9%, while it was 66.9% in 1991), ragweed seems to be the main aero-allergen plant (Juhász 1995).

The Carpathian Basin can be considered the region most polluted with airborne ragweed pollen in Europe. Some of the highest counts on peak days are 3,247 in Novi Sad (Serbia-Montenegro 2001), 2,003 in Szeged (Hungary 1991) and 1,394 in Pécs (Hungary 1994). The highest values observed in Novi Sad and Szeged on peak days are about one order of magnitude higher than those in other cities of Europe considered to be rather polluted.

Climatic characteristics of Szeged

Szeged (20°06'E; 46°15'N) lies near the confluence of the Tisza and Maros Rivers. It is the largest city in the south-eastern part of Hungary. The population of the city is

around 155,000 and its built-up area is about 46 km². Szeged and its surroundings are a flat and open region and the city has the lowest elevation in Hungary. In addition, the country lies in the Carpathian Basin; hence, Szeged is situated in a so-called double basin, which strengthens the effects of anticyclonic circulation patterns in accumulating pollutants together with pollen concentrations. The mean annual temperature is 11.2 °C and the mean annual precipitation is 570 mm.

Data and methods

In Szeged, the pollen content of the air has been monitored with the help of a high volume pollen trap (Lanzoni VPPS 2000) since 1989. The air sampler is located in the city, on the roof (20 m above ground level) of the building of the Faculty of Arts, University of Szeged. Daily pollen data were obtained by counting all pollen grains on four longitudinal transects (Käpylä and Penttinen 1981). Meteorological data were obtained from the monitoring station located 2 km from the sampling site in the city center, which is operated by ATIKÖFE (Environmental Protection Inspectorate of the Lower-Tisza Region, Branch of the Ministry of Environment).

There have been many studies analysing the effect of meteorological parameters on the pollen concentration of various plants (Giner et al. 1999; Galán et al. 2000, 2001; Jato et al. 2000). According to these the most important parameters are temperature, humidity and precipitation.

The data basis consists of daily ragweed pollen counts and daily averages of 11 meteorological variables not only for the main pollination periods but for the whole of the period 1997–2001. The criterion of main pollination periods was introduced by Nilsson and Persson (1981), and takes into account 90% of the annual total pollen concentration, eliminating the initial 5% and the final 5%.

The daily averages of the following meteorological variables were considered mean air temperature (T_{mean} , °C), maximum air temperature (T_{max} , °C), minimum air temperature (T_{min} , °C), daily temperature range ($\Delta T = T_{\text{max}} - T_{\text{min}}$, °C), relative humidity (RH, %), irradiance (I , Wm⁻²), wind speed (WS, m/s), vapour pressure (VP, mbar), saturation vapour pressure (E , mbar), potential evapotranspiration (PE, mm) and dew point temperature (T_d , °C).

A new interpretation of the two-sample test

The statistical test applied developed by Makra, is a method that is considered to be a new interpretation of the two-sample test.

The basic question addressed by this test is whether or not a significant difference can be found between the averages of an arbitrary subsample of a given time series and the whole sample (Makra et al. 2000, 2002; Tar et al. 2001).

Let $\xi_1, \xi_2, \dots, \xi_n, \dots, \xi_N$ represent independent random variables of normal distribution, with mean m . Let $E(\xi)$ be the expected value of ξ , with $D(\xi)$ its standard deviation.

Suppose that the standard deviations of ξ_i values are identical and equal to σ . Now, choose an optional subsample of n elements from the given whole time series of N elements ($n < N$).

Let

$$\bar{m} = \frac{\xi_1 + \dots + \xi_n}{n} \quad \text{and} \quad \bar{M} = \frac{\xi_1 + \dots + \xi_N}{N} \quad (1)$$

where $n < N$.

Then, after elementary steps, the random variable

$$\text{Probe statistics(PS)} = \frac{\bar{M} - \bar{m}}{\sqrt{\frac{N-n}{N \cdot n}} \cdot \sigma} \quad (2)$$

can be characterised by a standard normal distribution, $N(0;1)$.

This implies that having fixed the sample mean \bar{M} and the standard deviation σ , a test of the above null hypothesis for a given subsample mean \bar{m} leads to the following comparison of PS and x_p ,

$$p \left(\left| \frac{\bar{M} - \bar{m}}{\sqrt{\frac{N-n}{N \cdot n}} \cdot \sigma} \right| > x_p \right) = p \quad (3)$$

In Eq. 3, x_p is taken with p probability from the distribution function of the standard normal distribution, corresponding to a selected $0 < p \ll 1$ probability threshold.

If the absolute value of PS (Eq. 2) is higher than x_p , then \bar{M} and \bar{m} differ significantly. The null hypothesis, according to which there is no significant difference between \bar{M} and \bar{m} , can be rejected with the significance level p . (The significance tests are carried out at $p = 0.01$ significance level.)

This is a sufficient condition, ensuring the normal distribution of PS in Eq. 2, but it can be softened in the case of very large N and n , in the light of the central limit theorem. Stationarity and independence of the original distribution, however, are unavoidably necessary conditions of the Makra test.

The Makra test performs Eq. 3 for all possible subsamples with $n = 3, 4, \dots, N - 1$ elements of duration, starting from the 1st, 2nd, ..., $(N - n)$ th element of the time series. For example, in the case of 99 pieces of data (years) this means 4,752 repeated comparisons of the subsample average and the overall mean. Detection of significant deviations also includes information on their duration, onset and end (Makra et al. 2000, 2002; Tar et al. 2001).

When performing the Makra test, the significant breaks obtained may or may not be distinct. If they are not distinct, only one break is considered to be significant; namely, that for which the value of the test statistics is the maximum.

Factor analysis

In order to reduce the dimensionality of the above data set and thus to explain the relations between the 12 variables, the multivariate statistical method of factor analysis is used. The main object of factor analysis is to describe the initial variables X_1, X_2, \dots, X_p in terms of m linearly independent indices ($m < p$), the so-called factors, measuring different “dimensions” of the initial data set. Each variable X can be expressed as a linear function of the m factors:

$$X_i = \sum_{j=1}^m \alpha_{ij} F_j \quad (4)$$

where $\alpha_{i,j}$ are constants called factor loadings. The square of $\alpha_{i,j}$ represents the part of the variance of X_i that is accounted for by the factor F_j . It is a common practice for both the initial set of parameters X_i and the resultant factors F_j to be standardised having zero mean and unit variance. The first argument for using standardised variables is to give all variables equal weight, whereas the original variables may have extremely different variances. Another reason for using standardised variables is to overcome the problem of the different units of the various variables used. From the above, it is apparent that $\alpha_{i,j} \leq 1$. If a factor loading $|\alpha_{i,j}| \rightarrow 1$, the variable X_i is highly correlated to the factor F_j . Furthermore, a high correlation of some of the initial variables with the same factor is strong evidence of their covariability. The knowledge of covariability among the variables is a very important tool, since significant conclusions can be drawn for the causes of variation and/or the linkages between the initial variables (Bartzokas and Metaxas 1993, 1995; Sindosi et al. 2003).

One important step in this method is the decision about the number (m) of retained factors. Many criteria have been proposed for this. In this study, the Guttman criterion or rule 1 is used, which keeps the factors with eigenvalues greater than 1 and neglects those that do not account for at least the variance of one standardised variable X_i . Extraction was performed by principal component analysis (the k th eigenvalue is the variance of the k th principal component). There are an infinite number of alternative equations to Eq. 4. In order to select the best or the desirable ones, so-called factor rotation is applied, a process that either maximises or minimises factor loadings for a better interpretation of the results. In this study, “varimax” or orthogonal factor rotation is applied, which keeps the factors uncorrelated (Jolliffe 1990, 1993).

The factors can be considered vectors, components of which are the factor loadings. The complexity of the factors represents a matrix of factor loadings. The number of the factors is generally substantially less than that of the original variables. Since the complexity of the factors reflects that of the original variables, the problem examined becomes much simpler and can be interpreted much more easily.

The analysis was applied to a data set consisting of 12 columns (variables) and 465 rows (days) between 15 July and 15 October (first day: 15 July; last day: 15 October 1997–2001).

The linear regression model

In the linear regression model, the dependent variable is assumed to be a linear function of one or more independent variables plus an error introduced to account for all other factors:

$$y_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i \quad (5)$$

In the above regression equation, y_i is the dependent variable, x_{i1}, \dots, x_{ik} are the independent or explanatory variables, and u_i is the disturbance or error term. The goal of regression analysis is to obtain estimates of the unknown parameters β_1, \dots, β_k , which indicate how a change in one of the independent variables affects the values taken by the dependent variable.

The regression line expresses the best prediction of the dependent variable y_i . However, nature is rarely (if ever) perfectly predictable and usually there is substantial variation of the observed points around the fitted regression line. The deviation of a particular point from the regression line (its predicted value) is the error of the estimate, which is known as the deviation or residual value.

The aim of the method is to determine the values of the parameters that minimize the sum of the squared residual values (the sum of the squared errors of prediction) for the set of observations. This is known as a least-squares regression fit.

The square root of the average squared error of prediction is used as a measure of the accuracy of prediction. This measure is called the standard error of the estimate and is designated as σ_{est} . The formula for the standard error of the estimate is as follows:

$$\sigma_{est} = \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n}} \quad (6)$$

where n is the number of pairs of (x, y) points.

The smaller the variability of the residual values around the regression line relative to the overall variability, the better is our prediction. For example, if there is no relationship between the x and y variables, then the ratio of the residual variability of the y variable to the

original variance is equal to 1. If x and y are perfectly related, then there is no residual variance and the ratio of variance would be 0. In most cases, the ratio would fall somewhere between these extremes, that is, between 0 and 1. One minus this ratio is referred to as R^2 or the coefficient of determination. This value is immediately interpretable in the following manner. If we have an R^2 of 0.4, then we know that the variability of the y values around the regression line is 1.0–0.4 times the original variance; in other words we have explained 40% of the original variability, and are left with 60% residual variability. Ideally we would like to explain most if not all of the original variability. The R^2 value is an indicator of how well the model fits the data (e.g., R^2 close to 1.0 indicates that we have accounted for almost all of the variability with the variables specified in the model).

All statistical computations were performed with SPSS (version 9.0) software.

Results and discussion

Characteristics of the main pollination period of ragweed pollen for the 5-year data set examined as well as their averages, are shown in Table 1. Both the starting date of the pollination period (between 20 June and 13 July) and the finishing date (11–29 October) vary widely. The duration, average daily count and total count, except for 1998, show definite fluctuations (the total count: the sum of daily counts per cubic metre of air per year). On the other hand, 5 years might not be enough to detect clear trends.

Other parameters of ragweed pollen

Figure 1 shows daily ragweed pollen counts for the five data sets examined. Furthermore, averages for each pollination period (1997–2001) are also displayed (dotted lines). The curves display 5-day running averages; vertical lines indicate the main pollination period. Maximum values can be observed between 15 August and 15 September. This 1-month period is characterised by the highest ratio of ragweed pollen to total pollen release: 1997: 79.3%, 1998: 77.3%, 1999: 86.6%, 2000: 83.5%, 2001: 86.9%.

Five-year daily averages of the ragweed pollen counts for the period between 15 July and 15 October were analysed with the help of the Makra test. According to the χ^2 -test, the data set of these average values fits the normal

Table 1 Characteristics of ragweed pollen in Szeged for their main pollination period, according to Nilsson and Persson (1981)

Characteristics	1997	1998	1999	2000	2001	Average
Starting date	9 July	13 July	6 July	20 June	7 July	5 July
Finishing date	29 October	14 October	23 October	22 October	11 October	20 October
Duration (days)	113	94	110	95	97	102
Average diurnal count (pollen grains m^{-3})	61	29	67	88	93	68
Total count (pollen grains $m^{-3} year^{-1}$)	7,994	3,859	8,847	11,592	12,277	8,914

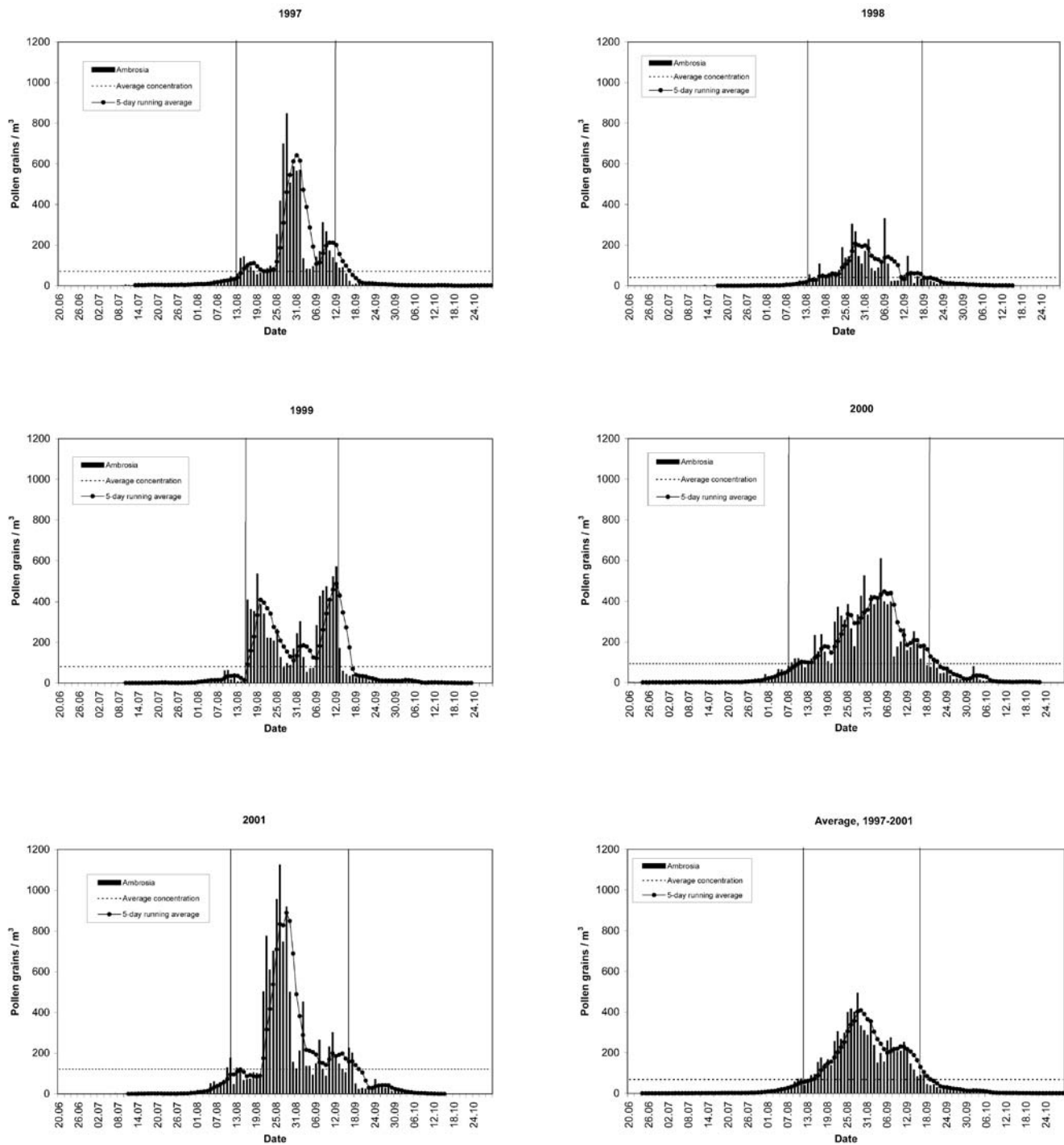


Fig. 1 Ragweed pollen counts and their running means, 1997–2001. *Dotted lines* The averages for the period examined *Vertical lines* The main pollination period

distribution; hence, the Makra test can be applied. The result received is reliable if this subset of the pollination period was chosen, since, on the one hand, the main pollination period differs year by year and, on the other hand, the period between 15 July and 15 October contains almost the annual total of the pollen grains each year. Only a mere fraction (0.35% in 1997; 0.05% in 1998;

0.08% in 1999; 0.38% in 2000 and 0.04% in 2001) can be found beyond it.

With the help of the Makra test we determined whether or not significant differences can be found between the average of an arbitrary subsample of the time series mentioned and that of the whole sample. It was found that averages of the subsamples with periods 15 July–14 Au-

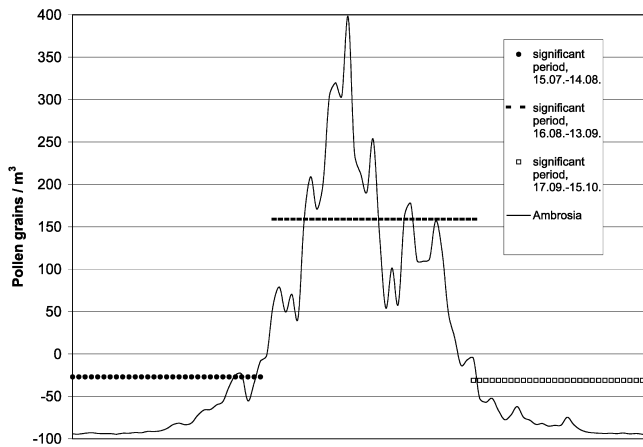


Fig. 2 Subperiods with significantly different averages of ragweed pollen counts from the mean of the entire data series, i.e. the “breaks”; diurnal average pollen counts, Szeged, 15 July–15 October, 1997–2001

gust, and 17 September–10 October are significantly lower than that of the whole sample, while the average of the period between 16 August and 13 September is significantly higher than that of the whole sample. This means that, according to the data set examined, the period between 16 August and 13 September can be considered to be the one most polluted by ragweed pollen in the air, hence the most dangerous one for pollinosis (Fig. 2). The result, obtained by this method indicates the same period as the most dangerous one for the highest pollen concentration, established according to Fig. 1.

The usefulness and the added value of the Makra test is that, by applying it, we can determine whether or not the averages of non-independent data sets, namely, the average of a selected subsample, and that of the whole data set differ significantly.

One might ask, why the Makra test is used and not the two-sample test or the F -test. The main difference between the methods is that the Makra test compares two non-independent samples (the whole sample and any of its subsamples) which, after elementary steps, become

independent of each other whereas the two-sample test and the F -test compare two different and independent samples.

Connection of ragweed pollen concentration with meteorological elements

The pollen season starts about the middle of July and ends about the middle of October, reaching a maximum in the middle of this 12-week time span. During this period temperature and vapour pressure have a systematic trend towards lower values. The highest temperatures are measured at the beginning of the pollen season, together with low pollen concentrations, and the lowest temperatures at the end of the pollen season, again with low pollen concentrations. If this systematic trend of the meteorological variables is not considered, the factor analysis applied might be influenced by it and the multiple regression model might not be as good as it could be.

In order to analyse any connection between ragweed pollen concentration and the meteorological elements, multivariate statistical analysis was applied with the help of factor analysis.

Factor analysis was performed for the 465-day-long period containing 5-year daily values of the examined 12 variables between July 15 and October 15 1997–2001. The mean seasonal variation was subtracted from the original daily values of each variable. According to the χ^2 -test the deseasonalised data sets fit the normal distribution; therefore, the method can be applied.

Factor loadings of the rotated component matrix are shown in Table 2. After performing factor analysis, according to the Guttman criterion, four factors were retained. Eigenvalues of the retained and rotated components as well as the explained and cumulative variances are also shown in Table 2. The four retained factors explain 84.4% of the total variance of the 12 original variables.

When performing factor analysis on standardised data, factor loadings received are correlation coefficients between the original variable and its co-ordinates (factor

Table 2 Factor loadings of the rotated component matrix. Loadings higher than $|0.091|$ (the threshold value for the 95% significance level) are written in bold type. The variables represent deseasonalised data sets

Variables	Factor 1	Factor 2	Factor 3	Factor 4
Daily pollen concentration	-0.083	0.204	0.090	-0.522
Mean air temperature (T_{mean})	0.689	0.699	-0.091	-0.037
Saturation vapour pressure (E)	0.706	0.671	-0.100	-0.040
Dew point temperature (T_d)	-0.175	0.901	0.005	-0.032
Vapour pressure (VP)	-0.008	0.956	-0.130	-0.057
Maximum air temperature (T_{max})	0.685	0.633	0.060	-0.097
Minimum air temperature (T_{min})	0.262	0.498	-0.794	-0.039
Potential evapotranspiration (PE)	0.935	0.242	-0.038	-0.013
Relative humidity (RH)	-0.944	0.220	-0.051	-0.014
Irradiance (I)	0.742	-0.199	0.224	0.077
Daily temperature range ($\Delta T = T_{\text{max}} - T_{\text{min}}$)	0.457	0.170	0.846	-0.063
Wind speed (WS)	-0.091	0.149	0.093	0.854
^a Eigenvalue	5.136	2.750	1.226	1.010
Explained variance(%)	42.803	22.921	10.214	8.415
Cumulative variance(%)	42.803	65.724	75.938	84.354

^a rotated sums of the squared loadings

scores) belonging to the rotated axes. Statistical significance of a factor loading – as a correlation coefficient – can be calculated by the following formula:

$$t = \sqrt{\frac{r^2(n-2)}{1-r^2}} \quad (7)$$

where r is the given factor loading, n is the number of the element pairs [$n = 465$ is the length of the 5-year (1997–2001) daily deseasonalised values of the 12 variables, $n - 2$ is the number of degrees of freedom and t is the parameter to be calculated, which follows Student's t -distribution. If the absolute value of t (in the case of $n - 2 = 463$ degrees of freedom and 5% significance level) is higher than the threshold value of Student's t -distribution ($t_{threshold} = 0.091$) then we can say that the null hypothesis – according to which the time series of the original variable and that of its factor scores are independent – is not fulfilled. Hence, if values of factor loadings are higher than 0.091 then the *a priori* hypothesis, according to which the two time series are independent, is rejected. For the sake of an easier survey, factor loadings higher than $|0.091|$ are indicated by bold figures (Bartzokas and Metaxas 1993, 1995; Sindosi et al. 2003).

Connections between the 12 variables examined will now be analysed according to the factor loadings of the retained and rotated factors (Table 2).

Factor 1

This factor explains 42.8% of the total variance and, on the basis of its significant factor loadings, contains the following daily parameters: mean air temperature (T_{mean}), saturation vapour pressure (E), dew point temperature (T_d), maximum and minimum temperatures (T_{max} and T_{min}), potential evapotranspiration (PE), relative humidity (RH), irradiance (I), daily temperature range ($\Delta T = T_{max} - T_{min}$), and wind speed (WS). Factor 1 reveals the opposite relationship between, on the one hand, potential evapotranspiration and, on the other hand, dew point temperature (the temperature at which the air becomes saturated), vapour pressure and relative humidity. Since the factor loading of the daily pollen concentration is not significant, no relation with the meteorological variables considered (however many of them have significant factor loadings) can be established.

Factor 2

This factor explains 22.9% of the total variance and comprises each of the 12 variables. Only the irradiance (I) has negative factor loading. Temperature (T_{mean} , T_{min} , T_{max}) and humidity (E , T_d , VP, PE) variables are in direct proportion. Higher temperature involves higher potential evapotranspiration (PE) and, thus, higher vapour pressure (VP), too; hence, dew point temperature (T_d) will also be higher. The factor loading of the daily pollen concentra-

tion is relatively low; however it is significant and is in direct proportion to the variables mentioned, except for irradiance (I).

In the light of the above results, why might hot and moist weather promote high pollen concentrations in the air? The role of temperature is clear; it is in direct proportion to pollen concentration. On the other hand, the effect of humidity might be complex. According to Giner et al. (1999), a nightly relative humidity in excess of 60% appeared to influence the atmospheric pollen concentration negatively during the day but, when the relative humidity was higher than 80% in the morning, pollen concentrations increased again. Another factor might be that, in the case of high relative humidity, pollen grains with an uneven surface can stick together more easily, causing the sampler to trap more pollen grains. Also, it has been observed in Szeged that, on the day following a rainy day (when relative humidity is higher), pollen concentration suddenly increases.

Factor 3

This factor explains 10.2% of the total variance and contains the mean air temperature (T_{mean}), saturation vapour pressure (E), vapour pressure (VP), minimum air temperature (T_{min}), irradiance (I), daily temperature range ($\Delta T = T_{max} - T_{min}$) and wind speed (WS). In this factor, daily temperature range, together with minimum temperature, has the highest loading. The factor loading of the daily pollen concentration is not significant; hence, its relationship to the variables mentioned can not be interpreted.

Factor 4

This factor explains 8.4% of the total variance and comprises only daily pollen concentration, maximum temperature (T_{max}) and wind speed (WS). High loadings of daily pollen concentration (-0.522) and wind speed (0.854) indicate that only part of the variance of daily pollen concentration is controlled by the above-mentioned temperature and humidity variables.

A study of factors 2 and 4 raises the question: why can wind promote high pollen concentrations in some cases and not in others? Under which conditions would pollen accumulate in the plant and under which conditions would pollen be released? The higher concentration of airborne pollen sometimes depends less on phenological phases and meteorological conditions than on transportation by the wind. Wind direction was not considered when factor analysis was performed (because of the disturbing effect of a building, situated next to the meteorological station). Pollen concentration depends not only on wind speed but on wind direction too. If winds blow from a source area, the pollen concentration increases; on the other hand, wind from other directions would modify the pollen counts less. The pollen concentration may also increase

Table 3 Regression analysis of the daily pollen number (p) as well as that of the difference of the daily pollen number and its mean seasonal variation (Δp). Variables represented in each case were considered in the first step of the regression analysis, while those indicated by bold type have remained until the last step of the regression. $^1\Delta p_{1-3}$ is the difference between the average pollen number for the preceding 3 days and its mean seasonal variation. Error = $\sum_{\text{time interval}} |(\text{real value} - \text{estimate value})/(\text{real value})|$. Values of the error function can only be used to compare various regression equations.

Dependent variable	Deseasonalised data		Original data		
	Case 1	Case 2	Case 3	Case 4	Case 5
Dependent variable	Δp	Δp	p	p	p
Independent variables	ΔT_{\min} ΔT_{\max} ΔT_{mean} ΔI ΔE ΔPE ΔVP ΔT_d ΔWS or WS ΔRH or RH	ΔT_{\min} ΔT_{\max} ΔT_{mean} ΔI ΔE ΔPE ΔVP ΔT_d ΔWS ΔRH $^1 \Delta p_{1-3}$	T_{\min} T_{\max} T_{mean} I E PE VP T_d WS RH	T_{\min} T_{\max} T_{mean} I E PE VP T_d WS RH p'	T_{\min} T_{\max} T_{mean} I E PE VP T_d WS RH p_{1-3} p'
R (expl. var.)	0.201	0.664	0.234	0.772	0.858
Error (1997–2000)	6.358	4.443	8.875	6.903	5.723
Error (2001)	11.759	3.330	7.880	4.104	4.731

through the re-suspension of high-atmosphere pollen that could have been released several hours earlier (Giner et al. 1999; Fehér and Járjai-Komlódi 1997).

Assessment of the daily average number of ragweed pollen grains

Further statistical analysis was performed to detect the influence of meteorological parameters on the daily ragweed pollen counts for the years examined. For the calculations, data for the period 1997–2000 were used, while the year 2001 was considered to be the reference year; namely, the assessment was performed for this year.

Let us introduce some notations. Let p be the daily pollen number, p' the mean seasonal variation of the daily pollen number (1997–2000) and p_{1-3} the average pollen number for the preceding 3 days. The daily sums of precipitation (namely: rainy days) were not considered (e.g. Giner et al. 1999). The reason for this is that the role of precipitation is complicated just because of the negative effect of rain intensity on pollen counts (Galán et al. 2000). According to Fornaciari et al. (1992), on the one hand, the best correlation was obtained by comparing pollen concentrations (Urticaceae) and meteorological parameters on non-rainy days and, on the other hand, correlating rainfall and pollen concentration is very difficult.

Multiple regression analysis was performed for the above-mentioned variables, using the “Enter” method. This method lets us select how independent variables are entered into the analysis. The Enter method enters all variables at the same time.

The multiple regression analysis can have two aims: (a) the estimation of future data on the basis of past ones; (b) to detect relations between the independent variables considered and the dependent variable. In the latter case, of course, we also receive an estimation for the chosen dependent variable (p), and the daily meteorological variables can also be used for estimating the daily number

of pollen grains. We chose this latter aim. According to which, p' and p_{1-3} and the meteorological variables were used in the regression procedure in order to detect the goodness of their connection with p .

Since p' is the mean seasonal variation of the daily pollen number (1997–2000), it is independent of the daily pollen number (p) of the reference year (2001). Furthermore, p' contributes to the quality of the regression considerably, since its standardized coefficient is much higher than that of the other variables.

The daily pollen concentration (p) depends on the average pollen number for the preceding 3 days (p_{1-3}). On the other hand, our aim is to detect the goodness of their relation; hence, p_{1-3} is applied in the regression procedure. Besides, application of p_{1-3} in the regression model gives better estimation for p .

When performing regression analysis, firstly the variable pollen number (p) or its modified variant (Δp) is considered to be the dependent variable, while the meteorological variables or their modified variants are counted as independent variables (in some cases not only meteorological variables are considered to be independent variables). The modified variable is defined as the difference of the original variable and its mean seasonal variation. (They are signed as follows: Δp , ΔT_{mean} , ΔT_{\max} , ΔT_{\min} , ΔE , ΔVP , ΔT_d , ΔI , ΔWS , ΔRH , ΔPE .)

Significance values of each independent variable are evaluated during the regression process. The less the significance value belonging to a variable, the better the role of the variable in the regression is. If the significance value belonging to the variable is relatively great, the variable cannot significantly improve the accuracy of the estimate.

The regression analysis is performed by using several steps. The variable having the greatest significance value will be omitted from the next step of the regression analysis. The procedure is carried out step-by-step until the significance values belonging to each variable become less than about 0.2. Five different cases were examined according to the selection of dependent and independent

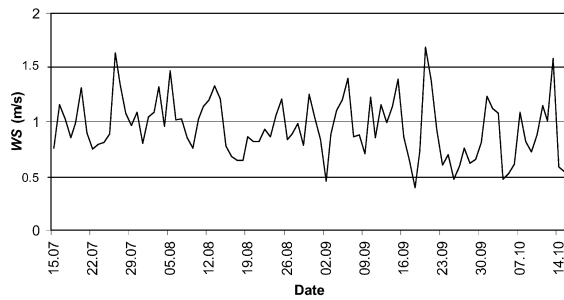


Fig. 3 The mean seasonal variation of the daily wind speed (WS)

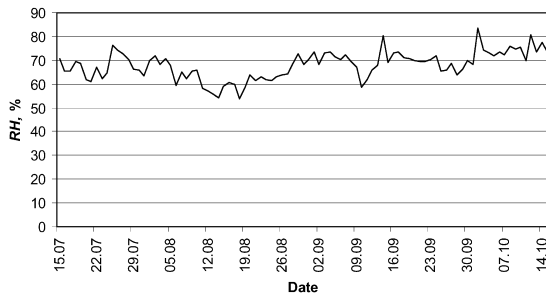


Fig. 4 The mean seasonal variation of the daily relative humidity (RH)

variables. The results (cases 1–5) are summarized in Table 3.

Since WS and RH as original variables have no seasonality (trend) (Fig 3–4), they were used in the regression analysis instead of variables ΔWS and ΔRH . According to the preliminary expectations, the result was the same (see case 1, Table 3). WS and RH seem not to play a role in the regression analysis in determining the daily pollen counts.

Except for case 2, each resulting regression equation comprises the following variables: I, E, PE. Hence they might have a greater role in daily pollen concentration. Variables T_d and VP occur in three cases; therefore, they might have a limited role compared to the former variables. If Δp_{1-3} is considered to be an independent variable, together with meteorological variables (deseason-

$$\Delta p = 0.089 + 12.751\Delta T_{\text{mean}} - 6.385\Delta T_{\text{max}} - 2.708\Delta T_{\text{min}} + 0.729\Delta p_{1-3} \quad (\text{case 2})$$

$$p = -27.510 - 0.123I - 35.017E + 88.243PE + 32.770VP - 1.756T_{\text{min}} + 0.464p' + 0.588p_{1-3} \quad (\text{case 5}) \quad (8)$$

alised data), then the regression equation contains only the temperature variables of meteorological ones (case 2, Table 3).

The equation resulting from the regression was applied to the data. When applying the regression analysis, the error was only calculated for those days, on which the daily pollen number exceeded 10. The reason for this is as follows. When comparing differences between the averaged daily and the estimated daily pollen numbers, then, considering the accuracy of the forecast, it is not very

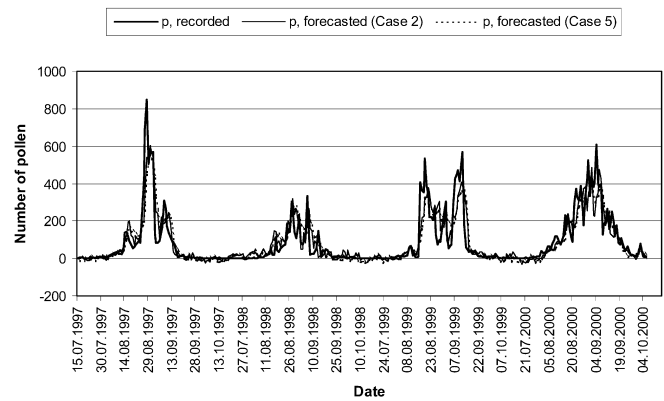


Fig. 5 Recorded and forecasted daily pollen numbers 1997–2000

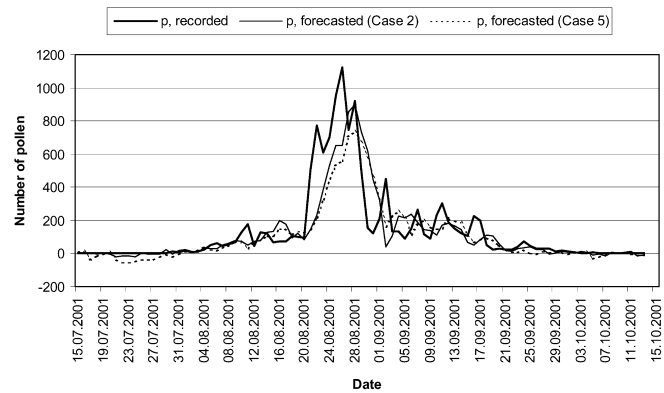


Fig. 6 Recorded and forecasted daily pollen numbers 2001

important if, in the case of a daily pollen count below 10 grains/m^3 , the error of the assessment is large (e.g. the procedure forecasts 2 or 3 pollen grains/ m^3 instead of 1). As mentioned above, the threshold value for clinical symptoms is considered to be $20 \text{ pollen grains/m}^3$ air (Jäger 1998)].

After performing the regression analysis only two cases are emphasised, since the lowest errors (Table 3) belong to following equations.

If the estimated daily pollen numbers (p , regression Eq. 8) are compared to the original ones, the equations seem to follow the trends tightly (Figs. 5–6). Both figures clearly indicate, on the one hand, that maxima are underestimated and, on the other hand, have negative estimates too. Estimates for case 2 show higher extremes than those for case 5 (see e.g. dates 14 August 1997; 18 August 1998; 04 September 2000; 02 September 2001; 28 August 2001). Results represented by Figs. 5–6 reflect the explained variances (R , Table 3).

For purpose of comparison, input data for case 2 and case 5 are the same; however, different variables were prepared from them (original data, deseasonalised data) (Table 3).

Conclusion

Parameters of ragweed pollen [maximum daily concentration per year; total number per year; first observation day; last observation day; duration (days); average daily number; number of days per year exceeding the threshold value for clinical symptoms ($20\text{--}30\text{--}50$ pollen grains $\text{m}^{-3}\text{day}^{-1}$)], fluctuate considerably; however, the short observation period does not allow to perform trend analysis.

The ragweed pollen load of Szeged is at its most serious between 16 August and 13 September, hence this is the most dangerous period for pollinosis. This period of highest pollen concentration, established by the Makra test, confirms the results of empirical pollen calculations.

Application of factor analysis reduced the dimensions of the original data set (daily ragweed pollen concentration and the 11 meteorological variables examined), facilitating the detection of connections among them. After factor analysis, four factors were retained according to the Guttman criterion. These four factors explain 84.4% of the total variance of the original 12 variables.

When factor analysis is performed with deseasonalised data instead of the original ones, the most characteristic changes in factor loadings belonging to the variables considered are as follows. Daily pollen concentration has significant loadings in factors 2 and 4 instead of factors 1, 3 and 4. The parameters play, to various extents, a positive role in increasing daily pollen concentrations, except for irradiance, which seems to be inversely related to daily pollen number. Minimum air temperature indicates a clear positive connection with daily pollen concentration. A significant connection between the daily pollen number and wind speed, with both positive and negative seems to explain both high and low pollen concentration as a consequence of high wind speed. This might hint at the ambivalent role of wind speed.

Surprisingly, if seasonality is subtracted, then the explained variance of the regression decreases. According to our preliminary expectations the relationship between daily pollen number and the meteorological variables would have been more clear-cut. However, this is not the case. Seasonality does not play such a great role in the time interval we have examined. It might be that a statistical description of pollen data simplifies complex processes like phenology and dispersion of pollen in the air.

This is a preliminary study based on data from only 5 years. The resulting regression equation is only an approximate indication of which variables are useful for predicting ragweed pollen counts. Further data are required to improve the model accuracy and, hence, to allow refined conclusions to be drawn. In further research, phenological phases of ragweed will also be taken into

account, that influence the formation and ripening of the pollen in the plant. A phenological model for ragweed, which calculates the timing of phenological phases on the basis of temperature sums, will be helpful to improve the accuracy of the prediction model. Another direction for further research might be the analysis of the relation between the different types of air mass established for the Carpathian Basin and pollen concentrations.

Acknowledgements The authors would like to express their gratitude to Helfried Scheifinger (Zentralanstalt fuer Meteorologie und Geodynamik, Vienna, Austria) for valuable help and advice, and to Szilvia Horváth and (Department of Climatology and Landscape Ecology, University of Szeged, Hungary), Gábor Motika (Environmental Protection Inspectorate of Lower-Tisza Region, Szeged, Hungary) and Predrag Radisic (University of Novi Sad, Laboratory of Palinology and Ecology, Novi Sad, Serbia-Montenegro) for useful hints on the topic. This study was supported by the Hungarian National Foundation for Scientific Research (OTKA no. T 034765).

References

- Bartzokas A, Metaxas DA (1993) Covariability and climatic changes of the lower troposphere temperatures over the Northern Hemisphere. *Nouvo Cimento* 16C:359–373
- Bartzokas A, Metaxas DA (1995) Factor analysis of some climatological elements in Athens, 1931–1992: covariability and climatic change. *Theor Appl Climatol* 52:195–205
- Comtois P (1998) Ragweed (*Ambrosia* sp.): the phoenix of allergophytes. In: Spietsma FT (ed) *Satellite Symposium Proceedings: Ragweed in Europe*. 6th International Congress on Aerobiology, Perugia, Italy, 31 August – 5 September 1998. Alk-Abello, Horsholm, pp 3–5
- Fehér Z, Járαι-Komlódi M (1997) An examination of the main characteristics of the pollen seasons in Budapest, Hungary (1991–1996). *Grana* 36:169–174
- Fornaciari M, Bricchi E, Greco F, Fascini D, Giannoni C, Frenguelli G, Romano B (1992) Daily variations of Urticaceae pollen count and influence of meteorological parameters in East Perugia during 1989. *Aerobiologia* 8:407–413
- Galán C, Alcázar P, Cariñanos P, Garcia H, Domínguez-Vilches E (2000) Meteorological factors affecting daily Urticaceae pollen counts in southwest Spain. *Int J Biometeorol* 43:191–195
- Galán C, Cariñanos 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
- Giner MM, Carrión Garcia JS, Garcia Sellés J (1999) Aerobiology of *Artemisia* airborne pollen in Murcia (SE Spain) and its relationship with weather variables: annual and intradiurnal variations for three different species. Wind vectors as a tool in determining pollen origin. *Int J Biometeorol* 43:51–63
- Jäger S (1998) Global aspects of ragweed in Europe. In: Spietsma FT (ed) *Satellite Symposium Proceedings: Ragweed in Europe*. 6th International Congress on Aerobiology Perugia, Italy, 31 August – 5 September 1998. Alk-Abello, Horsholm, pp 6–8
- Járαι-Komlódi M, Juhász M (1993) *Ambrosia elatior* (L.) in Hungary (1989–1990). *Aerobiologia* 9:75–78
- Jato MV, Rodríguez FJ, Seijo MC (2000) *Pinus* pollen in the atmosphere of Vigo and its relationship to meteorological factors. *Int J Biometeorol* 43:147–153
- Jolliffe IT (1990) Principal component analysis: a beginner's guide I. Introduction and application. *Weather* 45:375–382
- Jolliffe IT (1993) Principal component analysis: a beginner's guide II. Pitfalls, myths and extensions. *Weather* 48:246–253

- Juhász M (1995) New results of aeropalynological research in Southern Hungary. Publications of the Regional Committee of the Hungarian Academy of Sciences, Szeged 5:17–30
- Käpylä M, Penttinen A (1981) An evaluation of the microscopical counting methods of the tape in Hirst-Burkard pollen and spore trap. *Grana*, 20:131–141
- Makra L, Tar K, Horváth Sz (2000) Some statistical characteristics of the wind energy over the Great Hungarian Plain. *Ambient Energy* 21:85–96
- Makra L, Horváth Sz, Pongrácz R, Mika J (2002) Long term climate deviations: an alternative approach and application on the Palmer drought severity index in Hungary. *Phys Chem Earth* 27:1063–1071
- Nilsson S, Persson S (1981) Tree pollen spectra in the Stockholm region (Sweden), 1973–1980. *Grana* 20: 179–182
- Sindosi OA, Katsoulis BD, Bartzokas A (2003) An objective definition of air mass type affecting Athens Greece; the corresponding atmospheric pressure and air pollution levels. *Environ Technol* 24:947–962
- Tar K, Makra L, Horváth Sz, Kircsi A (2001) Temporal change of some statistical characteristics of wind speed over the Great Hungarian Plain. *Theor Appl Climatol* 69:69–76