

AEROPALYNOLOGY: AN EXCEL-BASED TOOL FOR STATISTICAL ANALYSIS AND COMPUTATION OF POLLEN SEASON PARAMETERS FROM AIRBORNE POLLEN CONCENTRATIONS

AEROPALYNOLOGY: NÁSTROJ V PROSTREDÍ EXCEL NA ŠTATISTICKÚ ANALÝZU A VÝPOČET PARAMETROV PEĽOVEJ SEZÓNY Z KONCENTRÁCIÍ PEĽU V OVZDUŠÍ

Michal Hrabovský*

*Comenius University in Bratislava, Faculty of Natural Sciences, Department of Botany,
Révová 39, 81102 Bratislava, Slovakia*

Received 15 April 2026; Received in revised form 1 May 2026; Accepted 13 May 2026

Abstract

AeroPalynology, an Excel-based tool, was developed for statistical analysis and computation of pollen season parameters from airborne pollen concentrations. Although advanced statistical environments such as R or Python offer extensive analytical capabilities, many practitioners in aerobiology are not familiar with these platforms or prefer more accessible solutions. To address this gap, an Excel add-in was designed to integrate commonly used statistical methods with tools specifically tailored for aerobiological applications. The add-in implements widely used non-parametric methods for comparing two or more datasets, including tests suitable for independent and related samples, as well as correlation and trend analyses. In addition, it provides built-in functions for calculating key parameters of the pollen season, enabling standardized and reproducible evaluation of aerobiological time series. AeroPalynology is intended as a practical and user-friendly solution for researchers and practitioners, facilitating routine data analysis without the need for specialized programming skills. By combining statistical functionality with domain-specific tools in a familiar spreadsheet environment, the add-in supports efficient analysis and interpretation of pollen data.

Key words: Microsoft Excel; aerobiology; pollen season; non-parametric statistics; data analysis

Introduction

The analysis of pollen seasons represents a crucial component of aerobiological research, as it enables the quantification of temporal patterns in airborne pollen concentrations. Accurate determination of the start, peak, and end of the pollen season is essential, particularly in the context of public health, allergology, and environmental monitoring, since it provides the basis for assessing population exposure to aeroallergens and for improving preventive and forecasting strategies (Ščevková, Dušička 2022, 2023, 2025).

Aerobiological datasets are typically characterized by high variability, non-linearity,

* Corresponding author: michal.hrabovsky@uniba.sk

and deviations from normal distribution, largely due to seasonal biological cycles and the strong influence of meteorological conditions on pollen release and dispersion. For this reason, non-parametric statistical tests are commonly applied in aerobiological data processing, as they do not assume normality and are robust to outliers and skewed distributions. These methods enable more reliable comparisons of seasonal characteristics and support robust evaluation of long-term trends in airborne pollen data (Žilka et al. 2024, Hrabovský et al. 2025, Ščevková et al. 2025).

In this context, dedicated computational tools were developed to streamline the processing and analysis of aerobiological datasets, facilitating standardized and efficient evaluation of pollen seasons, such as Excel-based applications designed for structured data handling and interpretation.

Methods

Custom functions and macros were developed in Microsoft Visual Basic for Applications (VBA) within MS Excel and tested using previously published data (Hrabovský et al. 2025, Ščevková et al. 2025). The functions are provided as supplementary material and are included in an MS Excel add-in available at: <https://fns.uniba.sk/pracoviska/biologicka-sekcia/kbo/aplikacie/>.

Results and Discussion

Seven macros implementing non-parametric statistical tests were developed as part of the AeroPalynology tool. Each macro corresponds to a specific statistical procedure commonly used in aerobiological data analysis. The **Spearman()** macro calculates Spearman's rank correlation coefficient to assess monotonic relationships between variables arranged in two columns. The **KruskalWallis()** macro performs the Kruskal–Wallis test for comparing more than two independent groups arranged in separate columns. The **MannWhitney()** macro implements the Mann–Whitney U test for comparison of two independent samples arranged in two columns. The **Wilcoxon()** macro performs the Wilcoxon rank-sum test for unpaired comparisons of two datasets arranged in two columns. The **Wilcoxon_paired()** macro applies the Wilcoxon signed-rank test for paired samples arranged in two columns, with paired observations aligned row-wise. The **MannKendall()** macro performs the Mann–Kendall trend test for the detection of monotonic trends in time series data contained in a single column. The **Ancova()** macro implements non-parametric analysis of covariance for group comparisons with covariate adjustment. The dependent variable (y) is provided in the first set of columns, while covariates are provided in a second set of columns, with observations paired row-wise across both datasets. All macros can be executed independently or through a higher-level wrapper macro AeroPalynology(), which provides centralized access to the full analytical workflow.

In the analysis of pollen seasons, several quantitative descriptors are required to characterize their timing, duration, and intensity. Timing parameters include the start day, end day, peak day (defined as the day of maximum pollen concentration and expressed

as the day of the year counted from 1 January, DOY), and the first high day (defined as the first day with a daily mean pollen concentration equal to or exceeding the threshold value). Duration is represented by the length of the pollen season, expressed in days. Intensity parameters include the peak value (expressed in pollen/m³), the Annual Pollen Integral (APIn, calculated as the sum of mean daily pollen concentrations during the year and expressed in pollen*day/m³), the Seasonal Pollen Integral (SPIn, calculated as the sum of mean daily pollen concentrations during the main pollen season and expressed in pollen*day/m³), and the number of high days. Together, these parameters provide a comprehensive description of pollen exposure dynamics, capturing both the temporal structure and the magnitude of airborne pollen load. A set of eight custom Excel functions was developed to automatically compute these key pollen season parameters. The analysed time series always begins on 1 January of the respective year. For the purpose of pollen season calculations, missing values (NA) are treated as zero concentration values; therefore, NA and 0 are considered equivalent in the computational procedure. The main pollen season definition is based on the proportion of the cumulative annual pollen sum included between the lower and upper percentage thresholds. This proportion is expressed as a decimal value corresponding to the selected cut-off limits. For example, a 0–100% definition corresponds to 1.00, a 5–95% season definition corresponds to 0.90, a 3–97% definition corresponds to 0.94, and a 10–90% definition corresponds to 0.80. The developed functions are as follows:

POLLENSEASON(range, season)

Determines the main pollen season **duration** (in days) based on the input data range and selected season definition. For example =POLLENSEASON(A1:A365, 0.9) returns the duration of 5–95% season.

POLLENSEASONSTART(range, season, asDate)

Returns the **start** of the main pollen season based on the selected season definition. The result can be returned either as DOY or as a calendar date when asDate = TRUE. For example =POLLENSEASONSTART(A1:A365, 0.9) returns the start day of season; =POLLENSEASONSTART(A1:A365, 0.9, TRUE) returns the date of start day of season.

POLLENSEASONEND(range, season, asDate)

Returns the **end** of the main pollen season based on the selected season definition. The result can be returned either as DOY or as a calendar date when asDate = TRUE. For example =POLLENSEASONEND(A1:A365, 0.9) returns the end day of season; =POLLENSEASONEND(A1:A365, 0.9, TRUE) returns the date of end day of season.

POLLENSEASONPEAK(range, season)

Identifies the **peak** pollen concentration within the defined main pollen season. For example, =POLLENSEASONPEAK(A1:A365, 0.90) returns the maximum pollen concentration within the 5–95% pollen season, while =POLLENSEASONPEAK(A1:A365, 1.00) returns the annual maximum value.

POLLENSEASONPEAKDAY(range, season, asDate)

Returns the **peak day** of pollen concentration within the defined main pollen season, optionally as a calendar date when *asDate* = TRUE. By default, the result is returned as DOY. For example, =POLLENSEASONPEAKDAY(A1:A365, 0.9) returns the peak day of season, while =POLLENSEASONPEAKDAY(A1:A365, 0.9, TRUE) returns the corresponding calendar date.

POLLENSEASONSUM(range, season)

Calculates the total pollen integral for the defined season (**APIn** or **SPIn** depending on input configuration). For example, =POLLENSEASONSUM(A1:A365, 1.00) returns the annual pollen integral (APIn), while =POLLENSEASONSUM(A1:A365, 0.90) returns the seasonal pollen integral (SPIn).

POLLENSEASONTHRESHOLD(range, season, threshold)

Calculates the number of **high days** exceeding a defined pollen concentration threshold. For example, =POLLENSEASONTHRESHOLD(A1:A365, 0.90, 80) returns the number of days during which pollen concentration exceeds 80 pollen/m³ within the defined main pollen season. When season = 1.00, the calculation corresponds to the full annual dataset.

POLLENSEASONTHRESHOLDDAY(range, season, threshold, asDate)

Identifies the first high day, defined as the first day when pollen concentration is equal to or exceeds the specified threshold value, optionally returned as a calendar date when *asDate* = TRUE. For example, =POLLENSEASONTHRESHOLDDAY(A1:A365, 0.90, 80) returns the first high day within the defined main pollen season when pollen concentration is equal to or exceeds 80 pollen/m³, while =POLLENSEASONTHRESHOLDDAY(A1:A365, 0.90, 80, TRUE) returns the corresponding calendar date.

ISSEASON(cell, range, season, asBinary)

Determines whether a cell within the specified range belongs to the pollen season. The function returns TRUE or FALSE. If *asBinary* is set to TRUE, the function returns 1 or 0 instead. The pollen season boundaries are calculated using the specified season coefficient.

ISSEASONTABLE(dates, range, season, asBinary, truncateFirstDays, truncateLastDays)

Returns a dynamic array identifying whether each observation belongs to the pollen season. The function groups the input time series by calendar year and calculates pollen season boundaries independently for each year. The output array contains TRUE or FALSE values for each row in the dataset. If *asBinary* is set to TRUE, the function returns 1 or 0 instead. Optionally, a fixed number of days can be truncated from the beginning and/or end of each year before pollen season detection is performed. Truncated days are internally replaced by zero values. For example, =ISSEASONTABLE(A:A, B:B, 0.90, FALSE, 120, 0) truncates the first 120 days of each year before calculating pollen season statistics.

POLLENSEASONTABLE(dates, range, season, truncateFirstDays, truncateLastDays)

Returns a dynamic array table with yearly pollen season statistics. The function groups data by **year** and calculates **start**, **end** and **duration** of pollen season, **peak** pollen concentration and its **day**, **SPIn** and **APIIn**. Each row corresponds to one calendar year in the dataset. Optionally, the function can truncate a fixed number of days from the beginning and/or end of each year before the pollen season calculations are performed. Truncated days are internally replaced by zero values. For example, =POLLENSEASONTABLE(A:A, B:B, 0.90, 0, 120) truncates the last 120 days of each year before calculating pollen season statistics.

Conclusion

The MS Excel add-in containing the developed functions and macros provides a standardized and automated framework for processing aerobiological time series data. It enables consistent calculation of pollen season parameters, statistical evaluation, and threshold-based analyses within a single integrated environment. This approach improves reproducibility, reduces manual processing errors, and facilitates efficient analysis of large aerobiological datasets.

Acknowledgements

The author would like to thank doc. RNDr. Jana Ščevková, PhD. for her assistance in reviewing and clarifying the definitions and interpretation of pollen season parameters.

References

- Hrabovský, M., Ščevková, J., Rendeková, A., Zahradníková, E., Jarolímek, I., 2025: Long-term changes in the allergenic potential of ruderal vegetation in a Central European urban area. *Regional Environmental Change*, 25:35.
- Ščevková, J., Dušička, J., 2022: Pollen season 2019 in the Bratislava area. *Acta Botanica Universitatis Comenianae*, 59:11–17
- Ščevková, J., Dušička, J., 2023: Pollen season 2021 in the Bratislava area. *Acta Botanica Universitatis Comenianae*, 61:21–27
- Ščevková, J., Dušička, J., 2025: Pollen season 2024 in the Bratislava area. *Acta Botanica Universitatis Comenianae*, 65:3–8
- Ščevková, J., Tropeková, M., Dušička, J., Lafféřsová, J., Rybníček, O., Zahradníková, E., Hrabovský, M., 2025: Rising allergenic risk in cities: a progressive pollen calendar analysis in response to climate change in Central Europe. *Regional Environmental Change*, 25:103.
- Žilka, M., Hrabovský, M., Dušička, J., Zahradníková, E., Gahurová, D., Ščevková, J., 2024 Comparative analysis of airborne fungal spore distribution in urban and rural environments of Slovakia. *Environmental Science and Pollution Research*, 31(54):63145–63160.

Abstrakt

Bol vyvinutý Excelový nástroj AeroPalynology určený na analýzu peľových sezón a spracovanie aerobiologických dát. Hoci pokročilé štatistické prostredia, ako sú R alebo Python, ponúkajú rozsiahle analytické možnosti, niektorí pracovníci v oblasti aerobiológie tieto platformy neovládajú alebo uprednostňujú dostupnejšie riešenia. Na vyplnenie tejto medzery bol navrhnutý Excelový doplnok, ktorý integruje bežne používané štatistické metódy s nástrojmi špecificky prispôbenými pre aerobiologické aplikácie. Doplnok implementuje široko používané neparametrické metódy na porovnávanie dvoch alebo viacerých súborov dát, vrátane testov pre nezávislé aj párové vzorky, ako aj korelačné a trendové analýzy. Okrem toho obsahuje vstavané funkcie na výpočet kľúčových charakteristík peľovej sezóny, čo umožňuje štandardizované a reprodukovateľné hodnotenie aerobiologických časových radov. AeroPalynology je určený ako praktické a používateľsky priateľské riešenie pre výskumníkov a odborníkov z praxe, ktoré umožňuje rutinnú analýzu dát bez potreby pokročilých programátorských zručností. Spojením štatistických funkcií s doménovo špecifickými nástrojmi v známom tabuľkovom prostredí add-in podporuje efektívnu analýzu a interpretáciu peľových dát.

Kľúčové slová: Microsoft Excel; aerobiológia; peľová sezóna; neparametrická štatistika; analýza dát