



The evaluation of pollen concentrations with statistical and computational methods on rooftop and on ground level in Vienna – How to include daily crowd-sourced symptom data



Maximilian Bastl^{a,b,*}, Katharina Bastl^a, Kostas Karatzas^c, Marija Aleksic^a, Reinhard Zetter^b, Uwe Berger^a

^a Aerobiology and Pollen Information Research Unit, Department of Oto-Rhino-Laryngology, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria

^b Department of Paleontology, University of Vienna, Geozentrum UZA II, Althanstraße 14, 1090 Vienna, Austria

^c Environmental Informatics Research Group, Department of Mechanical Engineering, Aristotle University, 54124 Thessaloniki, Greece

ARTICLE INFO

Keywords:

Symptom data
Sampling height
Pollen concentrations
Computational methods

ABSTRACT

Background: It is recommended to position pollen monitoring stations on rooftop level to assure a large catchment area and to gain data that are representative for a regional scale. Herein, an investigation of the representativeness of pollen concentrations was performed for 20 pollen types in the pollen seasons 2015–2016 in Vienna for rooftop and ground level and was compared with weather data and for the first time with symptom data.

Methods: The complete data set was analyzed with various statistical methods including Spearman correlation, ANOVA, Kolmogorov–Smirnov test and logistic regression calculation: Odds ratio and Yule's Q values. Computational intelligence methods, namely Self Organizing Maps (SOMs) were employed that are capable of describing similarities and interdependencies in an effective way taking into account the U-matrix as well. The Random Forest algorithm was selected for modeling symptom data.

Results: The investigation of the representativeness of pollen concentrations on rooftop and ground level concerns the progress of the season, the peak occurrences and absolute quantities. Most taxa examined showed similar patterns (e.g. *Betula*), while others showed differences in pollen concentrations exposure on different heights (e.g. the Poaceae family). Maximum temperature, mean temperature and humidity showed the highest influence among the weather parameters and daily pollen concentrations for the majority of taxa in both traps.

Conclusion: The rooftop trap was identified as the more adequate one when compared with the local symptom data. Results show that symptom data correlate more with pollen concentrations measured on rooftop than with those measured on ground level.

Background

Diseases related to pollen allergies are a major problem for a considerable fraction of the population. Pollen allergies affect about one million people in Austria according to the first Austrian allergy report.¹ The therapy of allergies is based on prevention of allergen exposure, provisions to suppress symptoms e.g. medication and specific immunotherapy.² Pollen information services support pollen allergy sufferers in allergy therapy by providing pollen forecasts and pollen related information and therefore contribute to allergen prevention. The fundament

of pollen information and pollen forecasts is high quality pollen concentration data.³ Such data sources are available in the vast majority by volumetric pollen and spore traps of the Hirst design⁴ in Europe. Pollen measurement stations are usually situated on rooftops to cover a large catchment area and to give a representative sample of regionally distributed pollen.^{5,6} Measurements on ground level are not recommended since the results would be affected by an overrepresentation of pollen producing sources in the immediate vicinity of the pollen trap.⁵ However, pollen allergy sufferers spend most of their time on ground level and are naturally influenced by local flora surrounding them.

Abbreviations: ZAMG, Zentralanstalt für Meteorologie und Geodynamik; PHD, Patient's Hayfever Diary; SLI, Symptom load index; SOM, Self Organizing Map.

* Corresponding author. Aerobiology and Pollen Information Research Unit, Department of Oto-Rhino-Laryngology, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria.

E-mail addresses: maximilian.bastl@meduniwien.ac.at (M. Bastl), katharina.bastl@meduniwien.ac.at (K. Bastl), kkara@eng.auth.gr (K. Karatzas), marija.prentovic@meduniwien.ac.at (M. Aleksic), reinhard.zetter@univie.ac.at (R. Zetter), uwe.berger@meduniwien.ac.at (U. Berger).

<https://doi.org/10.1016/j.waojou.2019.100036>

Received 10 December 2018; Received in revised form 4 April 2019; Accepted 17 April 2019

1939-4551/© 2019 The Author(s). Published by Elsevier Inc. on behalf of World Allergy Organization. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Several studies indicate that most airborne pollen types are of greater importance on micro or local scales and only a small fraction of airborne pollen is transported on regional scales or above.⁷ Hence, studies during the 80ies and 90ies of the last century performed investigations of the representativeness on pollen concentrations by Hirst type pollen traps at ground and rooftop level combined with weather parameters including also ground level measurements.^{8–14} The results of these studies are partly contradictory depending on the analyzed taxa, the methodology, climatic conditions, biogeographical regions and urban influences. Other studies focus on the concentrations in street canyons of urbanized areas, which seem to be comparable to rooftop levels¹⁵ or even lower in studies with personal samplers.¹⁶ The investigations of the representativeness of pollen concentrations of different heights were hitherto only compared with weather parameters, but never with daily crowd-sourced symptom data of pollen allergy sufferers. Crowd-sourced symptom data from on-line diaries is available only since the last ten years^{17,18} and was never implemented in a study comparing pollen-monitoring stations of different heights before. This study combines for the first time pollen concentration data of the most abundant taxa of two pollen traps in different heights (rooftop level and ground level) with meteorological parameters and crowd-sourced symptom data in a suburban district of Vienna by use of statistical and computational intelligence methods to improve the knowledge of local vegetation influence on pollen allergy sufferers and the possible impact on pollen information services.

Methods

Pollen sampling and analyzing

Two pollen traps of the Hirst design⁴ (Burkard Manufacturing Co.)

were used for this study. The first pollen sampler is located on the rooftop of the Karl-Kreil building on the property of the National Austrian Meteorological Service, Vienna (Zentralanstalt für Meteorologie und Geodynamik, ZAMG, 48°14'57"N 16°21'24"E) (Fig. 1) in an elevation of 14 m above ground. This pollen sampler (ATWIEN) is active continuously since 2003 and represents the reference device for pollen concentrations and pollen information within the city of Vienna. The second pollen sampler is situated in the phenological garden of the ZAMG (ATWIE2) at human height (in 1.6 m above ground; Fig. 1) and was only active during the study period of 2015/2016. The Euclidean distance between both samplers is approximately 100 m and both devices are situated in an open suburban area with no direct adjacent buildings but with gardens and individual buildings nearby in distances between 50 and 100 m. Both measurement stations were active during the main pollen season in 2015 and 2016. Whereas the rooftop trap performed continuous sampling in both years, the ground level trap started later in 2015 (on the 15th of March) and was closed during the months of October, November and the first half of December 2015, when pollination is less likely in Austria (personal observation and unpublished data). Only data from time periods in which both sites were active was included into the statistical analysis. However, in 2016 some days of data are missing in both traps. In the ground level trap (ATWIE2) the 19th and the 20th April is missing due to wrong manipulation of the drum in calendar week 16. Two days of data in July are missing in both traps due to a delayed drum change within the holiday season in both traps in calendar week 29. The drums of both pollen measurement samplers were routinely changed twice a week including a flow control to assure a correct mode of operation. The analysis of the tapes and the calculation of daily pollen concentrations followed the horizontal transversers reading method (including a percentage of reading of more than 10% of the slide), which is described in detail



Fig. 1. Overview map of the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) and the surrounding area extracted from Google Earth. The rooftop pollen monitoring station is represented by the light grey pin, the ground level station by the dark grey pin.

in the minimum recommendations for analyzing aerobiological samples.⁶ All samples were evaluated by the same analyst (MB) in order to exclude an inter-observer error. The pollen season of each taxon was defined by using the standardized season definition in the pollen database of the European Aeroallergen Network (EAN): the season starts at the day with 1% of the cumulative annual total pollen amount and ends at the day of 95% of the total annual pollen amount.

Symptom data

The symptom data included in this study originates from the Patient's Hayfever Diary (PHD), a free web-based online tool available via website (<https://www.pollendiary.com/Phd/>) or the "pollen" mobile application in Austria¹⁷ and were used as a "proxy" to understand symptoms. Symptom data from the PHD were already used in several scientific studies to compare pollen and symptom data. e.g. Ref. ^{19–25} The users of the PHD fill in a daily questionnaire and report the severity of specific symptoms per organ (eyes, nose and lungs) thus compiling a symptom score. The questionnaire asks for medication use to adapt the symptom score and for a zip code to assign the user to a biogeographical region. For this study the data from all users of Vienna were requested for the time of the study period in 2015 and 2016 including all organs (eyes, nose, lungs) and medication use. After this process daily symptom load indices (SLIs) were calculated as described in Bastl et al.¹⁹ The daily symptom data were normalized to attain values between 0 and 10. The crowd-sourced symptom data were not filtered within this study, except for the Vienna city zip codes. Hence, each user is included in the calculations if he/she entered symptoms in the pollen diary at least once during the study period.

Weather parameters

Meteorological data were obtained from the rooftop of the ZAMG on a daily basis. The average values of the following parameters were used: temperature (maximum, Tmax; minimum – Tmin; and mean; Tmean, in °C), total precipitation (PP, in mm), and humidity (H, in %) during the whole observation period when both pollen samplers were active.

Statistical analysis

The R software package in version 3.2.2 was used for statistical analysis. The Kolmogorov–Smirnov test for normality assessed the distribution of the pollen data. Since the results were negative, Spearman correlation analysis was performed in order to determine the strength and direction of the relationship between pollen concentrations recorded at rooftop and ground level. A linear regression model was used to describe the relationship between these variables, indicating R² and p-values.

Based on daily changes in the concentration of pollen in both traps during the course of the pollen season, the evaluated pollen types were divided into groups using the logistic regression. As a dependent variable in the model the direction of change in the data of the pollen trap on ground level was used and the independent variable was the direction of change of pollen data on the rooftop (+1, as positive change, and –1, as negative change of direction in relation to the previous day). The obtained categories are described with the Odds ratio and Yule's Q values. The analysis of variance (ANOVA) was carried out to determine the relative influence of the meteorological parameters on the variance of pollen in both traps within both years. The results of the statistical analyses are displayed in the significance levels of 0.05, 0.01 and 0.001.

Computational intelligence methods

The investigation of the relationships and dependencies among the parameters of the studied dataset (pollen concentration data from the ground level station (ATWIE2) and the roof top station (ATWIEN),

meteorological and symptom data for years 2015 and 2016) made use of the "black box" approach.²⁶ No previous knowledge was taken into account for the analysis concerning any relationship between ground or elevated pollen concentration data and symptom load data. Therefore, a twofold approach was performed: (a) an unsupervised computational intelligence method (thus without any employment with the data that leads to better performance and therefore learning). For this purpose Self Organizing Maps (SOMs) were used, which help in the visualization of two dimensions and the behavior of parameters interacting in many dimensions, and thus allow to assess which parameters have interrelationships, dependencies and are described with similar patterns (thus which station is more "related" to symptoms); (b) a supervised learning computational intelligence method: Random Forests (RF). The idea of this approach is to have a suitable and powerful algorithm, and then check which of the two stations lead to a better modeling of symptom data and to a better description of symptoms.

SOMs have been successfully used in the past in the analysis of related data.²⁷ The SOM method is based on neural networks consisting of a two-dimensional array of randomly weighted neurons.²⁸ Data are presented to the network and are matched with one of the neurons (the winning neuron). This procedure is repeated in a learning loop until a similarity criterion (usually a Euclidean distance) is fulfilled, causing the network topology to adjust and eventually form clusters of similar attributes. The Euclidean distance between neighboring neurons is represented by the unified distance matrix (U-matrix), and is therefore used as the basic visualization method for SOM because it depicts relations between neighboring data.^{29,30} Thus, high values in the U-matrix correspond to large distances between neighboring neurons while low (here black) values correspond to neurons, which have a low distance and formulate a cluster of similar neurons. Each parameter within the dataset is therefore "mapped" to a SOM, which represents a fictitious (mathematically robust) space that has no relationship to any physical space (thus without any axis), consisting of nodes arranged in a hexagonal grid. The visual inspection of each SOM allows the identification of "patterns", i.e. areas of high or low values for each parameter. The similarity of these patterns reflects that the same parameters have been mapped in the same way in the same area of their SOMs. The data can be described in terms of similarities and interdependencies in an effective way by taking into account the U-matrix. Calculations were made in Matlab[®] by employing the SOM toolbox of the Helsinki University of Technology.

The RF algorithm is a data-driven modeling approach,^{31,32} and is applied herein in order to identify the level of relationship between one of the pollen monitoring stations and the symptom data. The hypothesis to be evaluated is that if one of the two stations is more influential in the description of the symptom data, then data from this would lead to better symptom modeling results in comparison to the data coming from the alternative station and being used as model input. RF was applied with the aid of WEKA (a data mining software;³³), which served as the computational workbench for its development and testing. Algorithm details as well as model development and evaluation via a ten-fold cross validation procedure have already been described.³⁰

Results

This study was carried out in order to increase the knowledge on the representativeness of pollen concentrations measured at different heights in combination of crowd-sourced symptom data (symptom scores).

Pollen season parameters on ground level and rooftop

The airborne pollen of the 20 most widespread taxa in terms of quantitative occurrence in the air in Vienna were analyzed. The taxa included are: Alder (*Alnus*), hazel (*Corylus*), poplar (*Populus*), willow (*Salix*), elm (*Ulmus*), ash (*Fraxinus*), birch (*Betula*) and hornbeam (*Carpinus*) for the early flowering trees; oak (*Quercus*), beech (*Fagus*), plane tree (*Platanus*), walnut (*Juglans*) and pine tree (*Pinus*) for the late spring

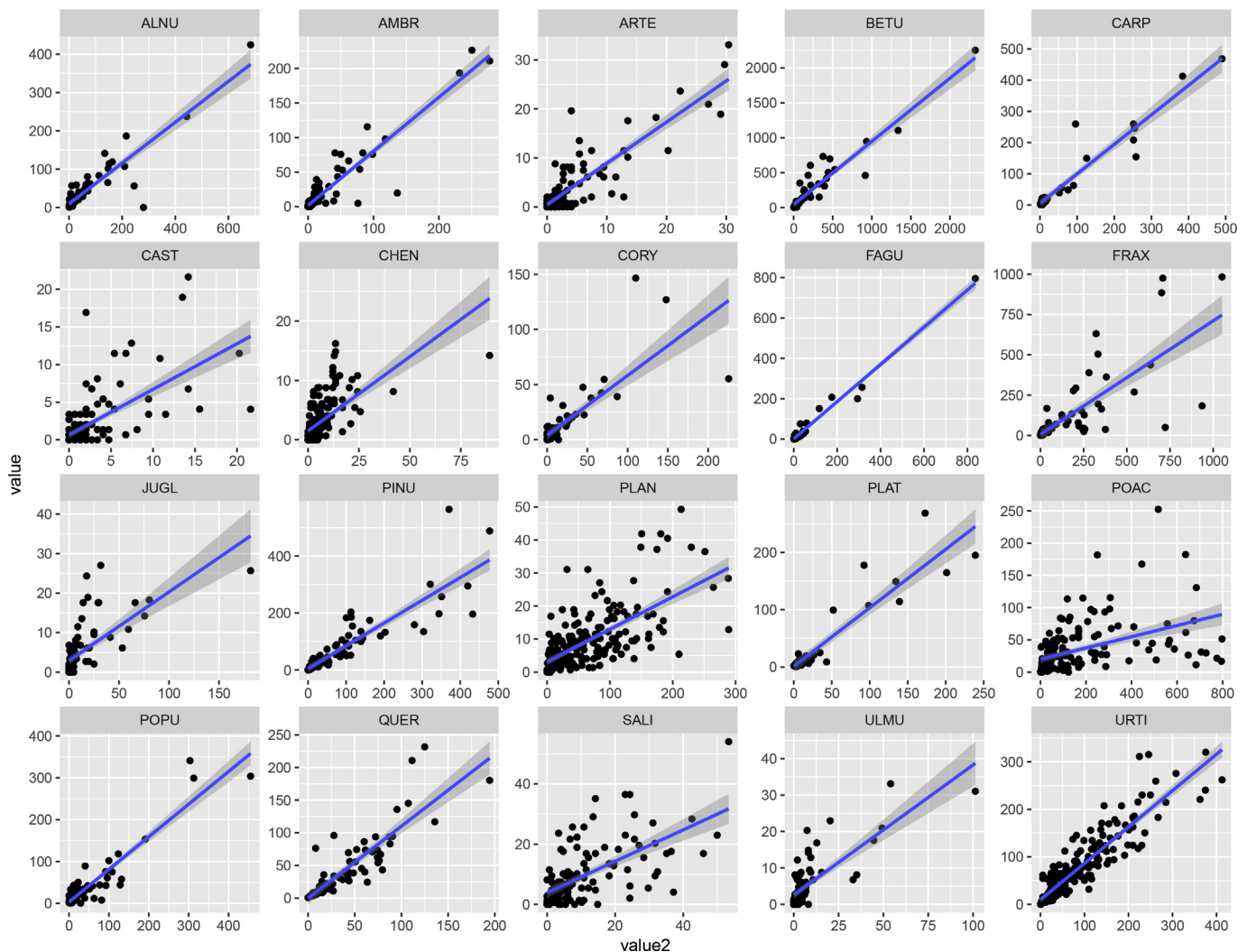


Fig. 2. Relationship between the daily pollen concentration data of twenty taxa on rooftop (Value) and on ground level (value 2). Shortcodes of all taxa corresponds to the first four characters of the taxon (e.g. ALNU = *Alnus*).

trees; pollen from the grass family (Poaceae), plantain (*Plantago*), the nettle family (Urticaceae) and sweet chestnut (*Castanea*) for summer; and pollen from the amaranth family (Amaranthaceae), mugwort (*Artemisia*) and ragweed (*Ambrosia*) for the late summer/early autumn. Start and end day, as well as the duration of the main pollen season and the seasonal pollen integral (SPIN; ³⁴) for both years were chosen as seasonal parameters for all taxa (see [Supplements](#)). Only *Alnus* and *Corylus* could not be compared in 2015 since the ground level pollen trap initiated its operation too late in the season. Overall, the pollen concentration levels of most of the taxa were higher in 2016 in Vienna compared to the values from the year 2015. There are also taxa that had higher pollen levels in 2015 such as *Fraxinus*, *Quercus*, *Pinus* and the Poaceae family ([Supplements](#)). The longest seasonal duration in both pollen measurement stations was observed for *Plantago* with 136 days (2015) and 103 days (2016) on the rooftop and 116 days (2015) and 97 days (2016) on ground level ([Supplements](#)). The greatest difference in the duration based on the same season definition for both pollen traps was recorded in the Poaceae family group. The difference between rooftop and ground level trap amounts to 29 days in 2015 and 34 days in 2016 for this taxon. The highest values in the rooftop trap in both years were recorded for *Betula* and the Urticaceae family, whereas the Poaceae family and the Urticaceae family showed the highest seasonal pollen values on ground level ([Supplements](#)).

Comparison of daily pollen concentrations

The statistical analysis of the daily pollen concentrations showed a significant positive relationship between the data gained at rooftop and on ground level ([Fig. 2](#)). The highest similarity could be observed in *Betula*, *Carpinus*, *Fagus* and Urticaceae pollen ($R^2 > 0.9$, $p < 0.001$), considering the relation of amount and presence of pollen during the season. The smallest association was observed in Poaceae pollen ($R^2 = 0.445$, $p < 0.001$).

The difference in the daily changes of direction in both traps was calculated with Odds Ratio, Yule's Q and the daily change of direction ([Tables 1 and 2](#)). The Yule's Q values ranged from 0.045 (CAST) to 0.993 (BETU). In most of the cases (12 pollen types), those values were higher than 0.8. Based on the differences in the changes of direction of the daily pollen concentrations the pollen types evaluated were divided into 5 groups ranging from a very high relationship to no relationship ([Table 2](#)). The highest relationship in the change of direction was indicated for *Alnus*, *Betula*, *Carpinus*, *Fagus*, and *Quercus* (Yule's Q ≥ 0.9), and the lowest for *Castanea* (Yule's Q < 0.1). The second group included *Ambrosia*, *Fraxinus*, *Pinus*, Poaceae, *Populus* and Urticaceae. These taxa also show a high relationship (Yule's Q values range from 0.8 to 0.9). All other pollen types show a medium relationship in the direction of change. No pollen type was assigned into the low relationship category.

Table 1

Differences in the daily change of direction of the studied pollen taxa in both pollen monitoring stations, including Odds Ratio, Yule's Q and the Same Direction. Shortcodes of all taxa corresponds to the first four characters of the taxon (e.g. ALNU = *Alnus*).

Pollen types	Odds Ratio	Yule's Q	Same Direction
ALNU	33.75	0.94	0.84
AMBR	12.81	0.86	0.78
ARTE	3.46	0.55	0.65
BETU	275.00	0.99	0.94
CARP	52.00	0.96	0.88
CAST	1.10	0.05	0.51
CHEN	3.41	0.55	0.65
CORY	7.58	0.77	0.73
FAGU	27.87	0.93	0.84
FRAX	9.02	0.80	0.75
JUGL	6.50	0.73	0.72
PINU	13.17	0.86	0.78
PLAN	2.58	0.44	0.61
PLAT	15.00	0.88	0.79
POAC	9.00	0.80	0.75
POPU	16.35	0.89	0.79
QUER	19.31	0.90	0.81
SALI	3.69	0.57	0.66
ULMU	6.11	0.72	0.71
URTI	14.67	0.87	0.79

Table 2

Categorization of pollen types based on the changes of direction in both pollen monitoring station, divided into the categories very high relationship, high relationship, medium relationship, low relationship and no relationship.

Categories	Range of Yule's Q	Pollen types	Additional information on the taxa
very high relationship	≥0.9	ALNU, BETU, CARP, FAGU, QUER	All taxa widespread in Vienna
high relationship	≥0.8–0.9	AMBR, FRAX, PINU, PLAT, POAC, POPU, URTI	FRAX, PINU, PLAT, POAC, POPU, URTI widespread in Vienna AMBR inhomogeneous, mostly long-range transport
medium relationship	≥0.4–0.8	ARTE, CHEN, CORY, JUGL, PLAN, SALI, ULMU	ARTE, CHEN, PLAN common in Vienna CORY, JUGL, SALI, ULMU locally scattered
low relationship	≥0.1–0.4	–	–
no relationship	<0.1	CAST	Rare and very local within Vienna

Pollen and weather parameters

The highest number of high significant correlations was detected in the parameters of temperature (Tmax, Tmean) and humidity (H). The maximum temperature (Tmax) shows a strong influence in both traps and was confirmed for *Betula*, *Amaranthaceae*, *Fraxinus*, *Plantago*, *Populus*, *Quercus*, *Salix* and *Urticaceae* ($p < 0.001$). The effect of Tmax on the rooftop is stronger for *Poaceae*, *Platanus*, *Corylus* and *Ulmus*, while the influence on *Ambrosia* and *Castanea* was higher on ground level. The same effects could also be obtained for the mean temperature (Tmean), but compared to Tmax the significances are lower. The only exception is *Juglans* demonstrating an influence of Tmean in the statistical analysis. Furthermore, no significant correlation for Tmean could be detected in the case of *Platanus* and *Ulmus*. The influence of the minimum temperature (Tmin) is in general higher on the rooftop, with the exception of the *Poaceae* family (Tmin significant in the garden). The statistical outcome demonstrates significant correlations between pollen data and humidity. Humidity had the same significance level for both traps for *Fraxinus* ($p < 0.001$), but showed a stronger impact on the roof top pollen data set

(Table 3) in most of the cases. Humidity did not show significant effects on the *Poaceae* family and *Populus* pollen registered in the ground trap. Precipitation only had significant influence on *Plantago* and *Platanus* in both measurement sites, and on *Artemisia* in the ground level trap ($p < 0.05$).

Pollen and symptom data

More than 20.000 data sets of crowd-sourced symptom data were included into the calculations of this study for every year. The monthly number of data entries as well as the monthly SLI values is displayed in detail in Table 4. Moreover, a comparison graph of daily pollen concentration data (ATWIEN and ATWIE2; total daily pollen concentrations) and daily SLI values is displayed in Fig. 3. The computational intelligence methods were used in order to combine the symptom data with the pollen concentration data of both pollen monitoring stations and the weather parameters. The most important six plant taxa inducing pollen allergies in the eastern part of Austria have been selected for this study³⁵: *Ambrosia*, *Artemisia*, *Betula*, *Fraxinus*, *Plantago* and the *Poaceae* family. The SOMs analysis (Fig. 4) showed that the six pollen types examined are well mapped demonstrating a distinct pattern per type. *Ambrosia*, *Artemisia*, *Betula* and *Plantago* demonstrate identical patterns for both stations whereas *Fraxinus* shows a slight difference between the stations, with ATWIE2 being slightly higher in terms of values. The *Poaceae* family demonstrates the biggest difference in its profile between the two stations, with ATWIE2 presenting with higher values. This difference was slight and could be explained by local influences.

Symptom data with and without medication demonstrate an identical profile. This indicates that the real symptoms (not the medication) are the explanation for the patterns found. In general, the symptom data were comparable by correlation within both pollen-monitoring stations in the SOMs analysis.

The Random Forest algorithm for symptom forecasting identified ATWIEN as the most significant parameter in terms of providing with pollen measurements as model inputs in comparison to ATWIE2. Table 5 presents the model results for the various subsets used as model inputs. It is evident that by using ATWIEN data, the symptom load index without medication can be forecasted with a correlation coefficient reaching 0.75, just 5% below the correlation coefficient achieved by using the whole dataset. This coefficient drops to 0.60 when using ATWIE2 data instead (i.e. 25% less than with the full dataset).

Discussion

Comparison of daily pollen concentrations

The significant correlations for the tree taxa of *Betula*, *Carpinus* and *Fagus* (Fig. 2) corresponds with most of the hitherto performed studies in literature, which took more than one taxa into account or focused on the "vertical variation" of tree pollen.^{8,9,12–14} Zwander⁸ and Rantio-Lehtimäki et al.⁹ suggested that the pollen concentrations of trees are more homogeneously distributed and have a higher pollen production than those of shrubs and grasses. The high numbers of trees in the city of Vienna and the surrounding areas is therefore an additional explanation. *Betula*, *Carpinus* and *Fagus* are used as park and alley trees on a regular basis in the city. Moreover, *Carpinus* and *Fagus* are highly represented in the forests close to Vienna. Also pollen from the *Urticaceae* family performed with a high correlation in the linear regression model. This is contradicting most findings in literature where pollen from the *Urticaceae* were either more abundant on ground level^{8,9} or on rooftop level^{10,11}. Bryant et al.³⁶ assumed that pollen of the *Urticaceae* family is more easily transported upwards due to its small size. Therefore, it would also be possible that the distribution of *Urticaceae* pollen is also more homogeneously distributed in the air and correlates with the pattern of the tree taxa if no local sources are close to the ground level pollen trap. The small association of the *Poaceae* pollen corresponds to previous findings since

Table 3

Correlations between pollen concentration data of both pollen monitoring stations and meteorological parameters including maximum temperature (Tmax), minimum temperature (Tmin), mean temperature (Tmean), precipitation (PP) and humidity (H).

Pollen types	Tmax		Tmin		Tmean		PP		H	
	Roof	Garden	Roof	Garden	Roof	Garden	Roof	Garden	Roof	Garden
ALNU	0.019 *	0.048 *	0.018 *	NS	0.005 **	0.020 *	NS	NS	NS	NS
AMBR	0.008 **	0.001 ***	NS	NS	NS	0.010 *	NS	NS	NS	NS
ARTE	NS	NS	NS	NS	NS	NS	NS	0.043 *	NS	NS
BETU	0.000 ***	0.000 ***	0.005 **	0.006 **	0.000 ***	0.000 ***	NS	NS	NS	NS
CARP	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
CAST	0.011 *	0.004 **	NS	NS	0.023 *	0.016 *	NS	NS	NS	NS
CHEN	0.000 ***	0.000 ***	0.027 *	NS	0.000 ***	0.002 **	NS	NS	NS	NS
CORY	0.007 **	0.013 *	0.003 **	0.031 *	0.001 **	0.006 **	NS	NS	NS	NS
FAGU	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
FRAX	0.000 ***	0.000 ***	0.003 **	NS	0.000 ***	0.000 ***	NS	NS	0.000 ***	0.000 ***
JUGL	NS	NS	0.000 ***	0.000 ***	0.025 *	0.013 *	NS	NS	0.000 ***	0.015 *
PINU	NS	NS	NS	NS	NS	NS	NS	NS	0.022 *	0.042 *
PLAN	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.046 *	0.028 *	0.000 ***	0.015 *
PLAT	0.006 **	0.014 *	NS	NS	NS	NS	0.013 *	0.011 *	NS	NS
POAC	0.000 ***	NS	NS	0.002 **	0.001 **	0.013 *	NS	NS	0.000 ***	NS
POPU	0.000 ***	0.000 ***	0.001 ***	0.003 **	0.000 ***	0.000 ***	NS	NS	0.035 *	NS
QUER	0.001 ***	0.000 ***	NS	NS	0.044 *	0.008 **	NS	NS	0.004 **	0.005 **
SALI	0.000 ***	0.000 ***	0.002 **	NS	0.000 ***	0.001 **	NS	NS	0.000 ***	0.006 **
ULMU	0.004 **	NS	NS	NS	NS	NS	NS	NS	NS	NS
URTI	0.000 ***	0.000 ***	0.000 ***	0.001 ***	0.000 ***	0.000 ***	NS	NS	0.000 ***	0.003 **

***- p<0.001, **- p<0.01, *- p<0.05; NS – No significant correlations.

Table 4

Daily symptom data from Vienna of the years 2015 and 2016 on a monthly basis, including the number of data entries and the daily averaged symptom data from the SLI per month. Moreover, the total number of data sets and the averaged SLI from January to September of each year is displayed.

Year 2015	Data entries per month	SLI	Year 2016	Data entries per month	SLI
January 2015	321	1,84	January 2016	370	1,46
February 2015	912	2,19	February 2016	1298	4,12
March 2015	3739	3,60	March 2016	3083	3,81
April 2015	6102	4,87	April 2016	6516	5,44
May 2015	5062	4,40	May 2016	3621	4,04
June 2015	3745	4,66	June 2016	3207	4,47
July 2015	2009	3,59	July 2016	1936	3,96
August 2015	1681	3,87	August 2016	1660	4,12
September 2015	1411	4,36	September 2016	1434	4,32
Total:	24982	3,71		23125	3,97

the limited height of grasses leads to an increase in pollen concentrations on ground level.^{8,9}

The categories of the Yule's Q analysis (Table 2) support the linear regression model (Fig. 1) and show overall very high to moderate relationships when both pollen traps are compared. The highest relationships were observed for the tree taxa of *Betula*, *Alnus*, *Carpinus*, *Fagus* and *Quercus* and correspond to the findings in former literature. High relationships were observed for *Fraxinus*, *Pinus*, *Platanus*, *Populus*, *Ambrosia*, *Poaceae* and the *Urticaceae* family (Tables 1 and 2). The daily distribution pattern shows a close relationship in the daily course of pollen concentrations during the grass pollen season although grass pollen concentrations were much higher in the ground level trap. Such a close relationship could originate in the natural exposure of the local grass composition on the property of the ZAMG. The close relationship in the ragweed pollen concentrations can be explained by long-range transport since *Ambrosia* is not growing in large quantities in the city area of Vienna but is mostly transported from outside of the city. The moderate category was represented by taxa, which can be associated with strong local influence (Tables 1 and 2). Trees such as *Corylus*, *Salix* and *Ulmus* are not frequently distributed in Vienna. The same is true for the weeds such as

Artemisia, the *Amaranthaceae* family and *Plantago*, which show an increased pollen concentration level on ground level and also a different change of direction pattern due to local influence especially in the ground level trap. *Castanea* is the only taxa with no relationship at all in the Yule's Q analysis (Table 2). This can be explained due to the rare occurrence of *Castanea* trees in Vienna in general and the thus limited distribution causing its pollination being a very local phenomenon in particular.

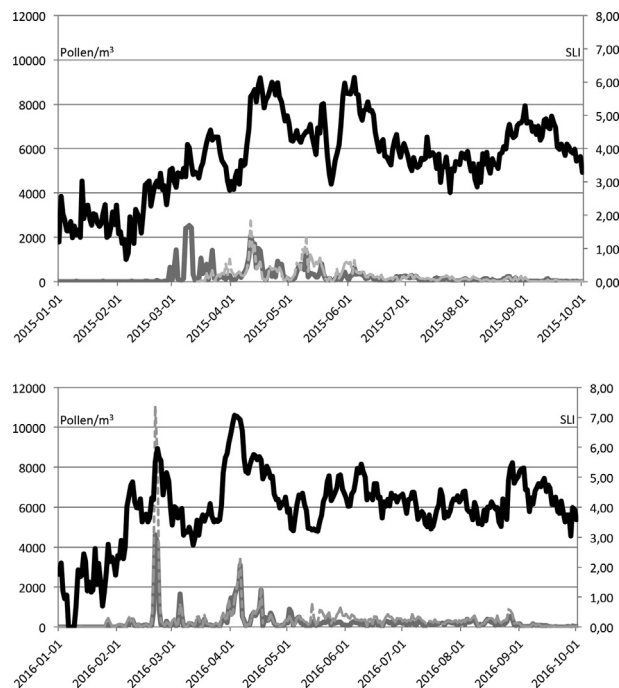


Fig. 3. Comparison graph of daily averaged symptom data (SLI; black line) and daily pollen concentrations (total of all taxa) of the rooftop pollen monitoring station (ATWIEN; dark grey line) and the ground level pollen monitoring station (ATWIE2; light grey dashed line) of the years 2015 (upper graph) and 2016 (lower graph).

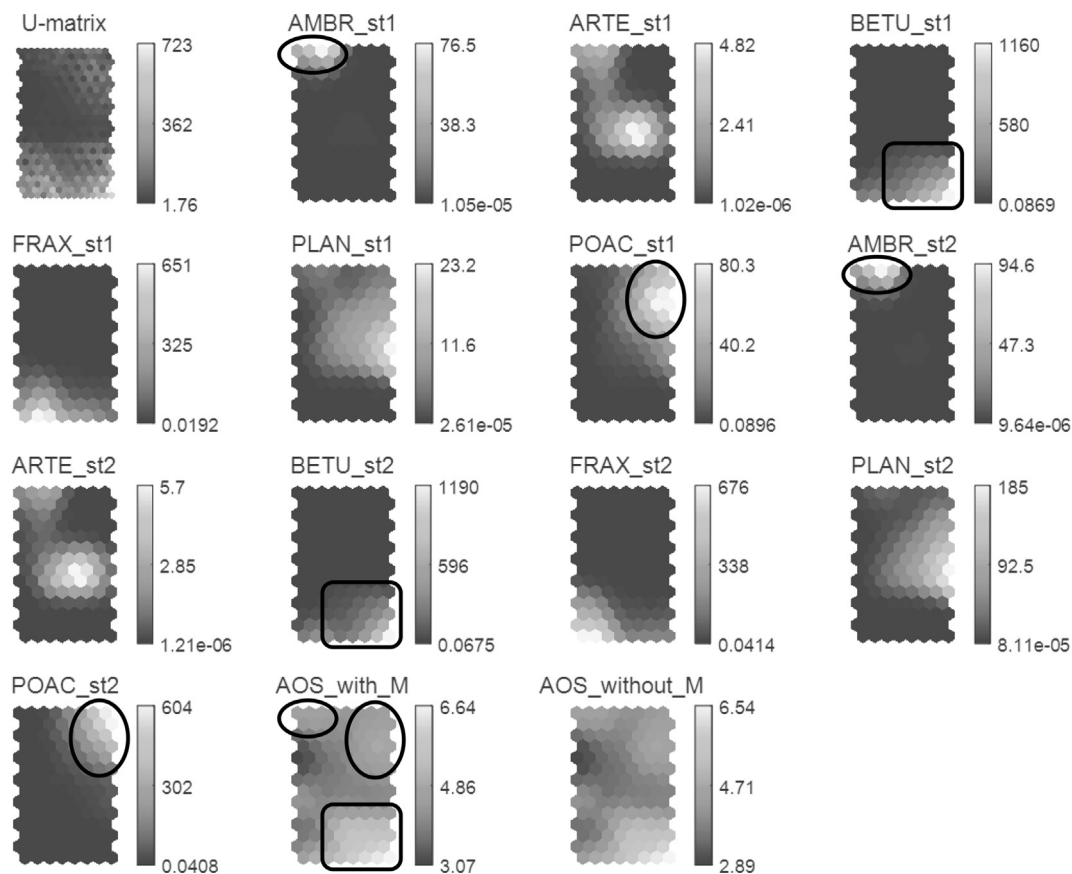


Fig. 4. The Self Organizing Maps (SOMs) of all data for both study years for *Ambrosia* (AMBR), *Artemisia* (ARTE), *Betula* (BETU), *Fraxinus* (FRAX), *Plantago* (PLAN) and the Poaceae family (POAC) for ATWIEN (st1), and ATWIE2 (st2). Moreover the average symptoms (AOS) with and without medication (M) are displayed. Each SOM corresponds to one of the parameters under investigation.

Pollen and weather parameters

Temperature and humidity had a significant influence on pollen concentrations in both traps, but showed no clear differences between them (Table 3) similar to former comparison studies.^{9,10,12} Precipitation showed nearly no correlation in the ANOVA except for plantain and plane tree (Table 3). Alcázar et al.¹² also reported low influence of meteorological factors such as precipitation, humidity or temperature when comparing pollen concentrations of different heights. Rantio-Lehtimäki et al.⁹ reported that wind speed did not show any differences either in the pollen concentrations of pollen traps on ground and rooftop level. The results herein confirmed the results from literature and showed that the influence of meteorological parameters in general is small and demonstrate an equal effect on the pollen types monitored in both pollen traps and thus on both altitude levels (rooftop and ground).

Pollen and symptom data

The analysis of the SOM's suggests that the symptom load index seems to be mostly attributed to *Betula*, followed by the Poaceae family and

Table 5

Performance statistics of the Random Forest-oriented model for the forecasting of symptom data (without medication) and for various subsets of the initial data set.

	Full dataset	Excluding meteo	Only station 1 (Excluding meteo and station 2)	Only station 2 (Excluding meteo and station 1)
Correlation coefficient	0.7921	0.7719	0.7491	0.6054
MAE	0.5386	0.5502	0.5725	0.6958
RMSE	0.7277	0.7509	0.7815	0.9397

Ambrosia in the case of the rooftop station. For the ground level station *Betula* and also *Ambrosia* seems to be related to symptoms whereas the relationship of the Poaceae family to symptoms is less reflected in the ground level trap. The importance of *Betula*, Poaceae and *Ambrosia* is confirmed in prevalence studies in Austria³⁵ as well as studies regarding the pollen information consumption.²² The highest prevalence rates of pollen allergies in Austria are attributed to birch pollen allergy with more than 40% and to grass pollen allergy with more than 50%.³⁵ The prevalence rate of *Ambrosia* is lower; however, the allergenicity of *Ambrosia* is high and may be reflected in the symptom data as well. *Artemisia*, *Fraxinus* and *Plantago* are of minor importance regarding the analysis of the symptom data. However, there is a slight difference in the relationship between *Fraxinus* and *Betula*, which can be explained by the overlapping pollination period of both tree taxa. *Fraxinus* pollen could have a more local pattern and therefore, may not be distributed as homogeneously as other tree pollen (personal observation). This could also explain the higher values of *Fraxinus* pollen in the ground level trap.

The Random Forest algorithm for symptom forecasting identified the rooftop level trap as the most significant parameter in terms of providing with pollen measurements as model inputs in comparison to the ground level station. This result leads to the conclusion that not only pollen concentration data, but also symptom data are comparable on a larger scale and can be attributed to a single pollen trap in one city on rooftop level taking into account that they represent an aggregated-averaged symptom load (see also Ref. ³⁷). This outcome should be taken into account together with contradictory findings on the pollen distribution in one city³⁸ suggesting that symptom levels may be different in various locations in one city. Although some of the taxa examined in this study clearly showed higher pollen concentrations on ground level (e.g. the Poaceae family or *Plantago*) they did not demonstrate sufficient influence

on the average symptom data in the whole observation area. Therefore, current recommendations on the position of pollen monitoring stations especially its position on flat rooftops⁶ can be supported also regarding the averaged symptom data in a city.

Limitations

The pollen data obtained for this study is following the minimum recommendations for the evaluation of pollen concentration data.⁶ Fluctuations in the daily pollen concentrations of different monitoring stations are the normal case since pollen-monitoring stations represent point measurements.³⁹ Therefore, not only the daily pollen concentrations but also the differences in the daily change of direction (increase/decrease of pollen levels during the season) have been analyzed in this study to increase the significance of the pollen concentration data. The PHD data are crowd-sourced, which results in an easy and fast data access and high numbers of data sets but less information on the subject profiles. Since a large number of taxa were presented in this study no user filtering was performed in order to include the maximum number of datasets for the statistical analysis and the computational methods. Usually users are only entering data when they experience symptoms and do not enter data during the whole year. In addition the medical history of users remain unknown, as the PHD conforms to data protection law. Therefore, crowd-sourced symptom data cannot be compared directly with symptom data of patients and are used as a proxy to understand symptoms as shown in former studies.^{19–25}

Conclusion

The analysis of two Hirst type pollen traps, one on rooftop level and one on ground level, allowed an in-depth investigation of the representativeness of ground-level and rooftop pollen traps of many pollen types including several tree, grass and weed taxa throughout the pollen seasons of 2015–2016. Whereas the seasonal pattern (start, end and timing of peaks) was consistent for most of the pollen types, the absolute levels showed more variation.

Data presented herein revealed that (1) symptom data correlate more with the rooftop trap than with the ground level trap in a suburban area in Vienna, (2) local influences could be observed for the ground level trap (e.g. by Poaceae and *Plantago* pollen), (3) those local influences do not have a significant impact on the overall symptom data, (4) computational intelligence methods like the SOMs proved useful besides standard statistical methods when evaluating the representativeness of different pollen monitoring stations at different heights and (5) pollen concentrations of the most prevalent aeroallergens seem to be homogeneously distributed. These conclusions have a far-reaching significance: Computational intelligence methods may be applied more often to such datasets to explore variations in pollen concentrations and symptoms. Pollen concentrations measured on rooftop – as it is a standard aerobiological procedure at the moment – is justified also when overall symptom data are in focus. Results show a strong local aspect and they are only valid for the city of Vienna. However, this study gives guidance in the methods to compare symptom and pollen data suggesting that one pollen trap per city (for cities of the size of Vienna, with no strong orography) could be used as a good indication concerning the distribution of pollen in combination with overall symptoms.

Declarations

Ethics approval and consent to participate

Only crowd-sourced and anonymized symptom data were used, so no ethics approval was needed.

Consent for publication

All authors confirm their consent for publication.

Availability of data and material

Users were guaranteed anonymity, so now raw data or individual user data are available.

Competing interests

All authors report no competing interests.

Funding

No funding was received for this study.

Author contributions

MB drafted the main body of the manuscript and prepared several figures. KK performed the computational calculations, drafted some of the figures and supervised the statistical analyses. MA took part in the statistical analyses. KB, KK, MA, RZ and UB all took part in the design of the study, the interpretation of the results and in writing the manuscript. UB and RZ supervised the study. All authors approved the final version of the manuscript.

Acknowledgements

The authors want to thank Alexander Kowarik for the support in statistical analysis of the data as well as Christoph Jäger for the technical background of the databases used (European Aeroallergen Network, Patient's Hay-fever Diary). We are indebted to the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) for providing the weather data used in this study and want to thank especially Harald Seidl for his efforts.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.waojou.2019.100036>.

References

1. Dorner T, Rieder A, Lawrence K, Kunze M. *Österreichischer Allergiebericht. Verein Altern mit Zukunft*. 2006. Wien.
2. Schmid-Grendelmeier P. Pollen allergy and immunotherapy. *Ther Umsch*. 2012;69:239–248.
3. Bastl K, Berger M, Bergmann K-C, Kmenta M, Berger U. The medical and scientific responsibility of pollen information services. *Wien Klin Wochenschr*. 2017;129(1–2):70–74. <https://doi.org/10.1007/s00508-016-1097-3>.
4. Hirst JM. An automatic volumetric spore trap. *Ann Appl Biol*. 1952;39(2):257–265. <https://doi.org/10.1111/j.1744-7348.1952.tb00904.x>.
5. De Weger LA, Bergmann K-C, Rantio-Lehtimäki A, et al. Impact of pollen. In: Sofiev M, Bergmann K-C, eds. *Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts*. Dordrecht: Springer; 2013:161–215.
6. Galán C, Smith M, Thibaudon M, et al. Pollen monitoring: minimum requirements and reproducibility of analysis. *Aerobiologia*. 2014;30:385–395. <https://doi.org/10.1007/s10453-014-9335-5>.
7. Sofiev M, Belmonte J, Gehrig R, et al. Airborne pollen transport. In: Sofiev M, Bergmann K-C, eds. *Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts*. Dordrecht: Springer; 2013:127–159.
8. Zwander H. Der Blütenstaubgehalt der Luft in Atemhöhe im Vergleich mit Luftschichten in 27 Meter Höhe. *Carinth II*. 1985;175/95:12–25.
9. Rantio-Lehtimäki A, Koivikko A, Kupias R, Mäkinen Y, Pohjola A. Significance of sampling height of airborne particles for aerobiological information. *Allergy*. 1991;46:68–76.
10. Galán C, Alcázar-Teno P, Domínguez-Vilches E, Villamandos de la Torre F, García-Pataleon FI. Airborne pollen grain concentrations at two different heights. *Aerobiologia*. 1995;11:105–109.
11. Alcázar P, Galán C, Carinanos P, Domínguez-Vilches E. Vertical variation in Urticaceae airborne pollen concentration. *Aerobiologia*. 1998;14:131–134.
12. Alcázar P, Galán C, Carinanos P, Domínguez-Vilches E. Effects of sampling height and climatic conditions in aerobiological studies. *J Invest Allergol Clin Immunol*. 1999;9(4):253–261.
13. Leuschner RM. Comparison between pollen counts at ground and at roof level in Basel (Switzerland). *Aerobiologia*. 1999;15:143–147.

14. FThM Spieksma, van Noort P, Nikkels H. Influence of nearby stands of *Artemisia* on street-level versus roof-top-level ratio's of airborne pollen quantities. *Aerobiologia*. 2000;16:21–24.
15. Peel RG, Kennedy R, Smith M, Hertel O. Do urban canyons influence street level grass pollen concentrations? *Int J Biometeorol*. 2014;58(6):1317–1325. <https://doi.org/10.1007/s00484-013-0728-x>.
16. Peel RG, Hertel O, Smith M, Kennedy R. Personal exposure to grass pollen: relating inhaled dose to background concentration. *Ann Allergy Asthma Immunol*. 2013;111(6):548–554. <https://doi.org/10.1016/j.anai.2013.09.002>.
17. Kmenta M, Bastl K, Jäger S, Berger U. Development of personal pollen information – the next generation of pollen information and a step forward for hay fever sufferers. *Int J Biometeorol*. 2014;58(8):1721–1726. <https://doi.org/10.1007/s00484-013-0776-2>.
18. De Weger LA, Hiemstra S, Op den Buysch E, van Vliet AJH. Spatiotemporal monitoring of allergic rhinitis symptoms in The Netherlands using citizen science. *Allergy*. 2014;69(8):1085–1091. <https://doi.org/10.1111/all.12433>.
19. Bastl K, Kmenta M, Jäger S, Bergmann KC, EAN, Berger U. Development of a symptom load index: enabling temporal and regional pollen season comparisons and pointing out the need for personalized pollen information. *Aerobiologia*. 2014;30:269–280. <https://doi.org/10.1007/s10453-014-9326-6>.
20. Bastl K, Kmenta M, Geller-Bernstein C, Berger U, Jäger S. Can we improve pollen season definitions by using the symptom load index in addition to pollen counts? *Environ Pollut*. 2015;204:109–116. <https://doi.org/10.1016/j.envpol.2015.04.016>.
21. Karatzas K, Voukantsis D, Jäger S, et al. The patient's hay-fever diary: three years of results from Germany. *Aerobiologia*. 2014;30:1–11. <https://doi.org/10.1007/s10453-013-9303-5>.
22. Kmenta M, Zetter R, Berger U, Bastl K. Pollen information consumption as an indicator of pollen allergy burden. *Wien Klin Wochenschr*. 2016;128:59–67. <https://doi.org/10.1007/s00508-015-0855-y>.
23. Kmenta M, Bastl K, Kramer MF, et al. The grass pollen season 2014 in Vienna: a pilot study combining phenology, aerobiology and symptom data. *Sci Total Environ*. 2016;566–567:1614–1620. <https://doi.org/10.1016/j.scitotenv.2016.06.059>.
24. Voukantsis D, Karatzas K, Jäger S, Berger U, Smith M. Analysis and forecasting of airborne pollen-induced symptoms with the aid of computational intelligence methods. *Aerobiologia*. 2013;29(2):175–185. <https://doi.org/10.1007/s10453-012-9271-1>.
25. Voukantsis D, Berger U, Tzima F, et al. Personalized symptoms forecasting for pollen-induced allergic rhinitis sufferers. *Int J Biometeorol*. 2015;59(7):889–897. <https://doi.org/10.1007/s00484-014-0905-6>.
26. Solomatine DP, Dulal KN. Model trees as an alternative to neural networks in rainfall–runoff modeling. *Hydrol Sci J*. 2003;48(3):399–411.
27. Karatzas K, Riga M, Voukantsis D, Dahl Å. Allergy medication dose and biological weather: a data-oriented investigation. *Int J Artif Intell*. 2012;9(12):90–105.
28. Kohonen T. *Self-Organizing Maps*. second ed. Germany: Springer-Verlag Berlin Heidelberg; 1997.
29. Ultsch A, Siemon HP. Kohonen's Self organizing feature maps for explanatory data analysis. *INNC Paris*. 1990:305–308.
30. Karatzas K, Katsifarakis N. Modelling of household electricity consumption with the aid of computational intelligence methods. *Adv Build Energy Res*. 2018;12(1):84–96. <https://doi.org/10.1080/17512549.2017.1314831>.
31. Liu X, Wu D, Zewdie GK, et al. Using machine learning to estimate atmospheric *Ambrosia* pollen concentrations in Tulsa, OK. *Environ Health Insights*. 2017;11:2–10. <https://doi.org/10.1177/1178630217699399>.
32. Nowosad J. Spatiotemporal models for predicting high pollen concentration level of *Corylus*, *Alnus* and *Betula*. *Int J Biometeorol*. 2016;60(6):843–855. <https://doi.org/10.1007/s00484-015-1077-8>.
33. Eibe F, Hall MA, Witten IH. *The WEKA Workbench. Online Appendix for „Data Mining: Practical Learning Tools and Techniques“*. fourth ed. Morgan Kaufmann; 2016.
34. Galán C, Ariatti A, Bonini M, et al. Recommended terminology for aerobiological studies. *Aerobiologia*. 2017;33:293–295.
35. Hemmer W, Schauer U, Trinca A-M, Neumann C. *Endbericht 2009 zur Studie: Prävalenz der Ragweedpollen-Allergie in Ostösterreich. St.Pölten: Amt der NÖ Landesregierung, Landesamtsdirektion. Amtsdruckerei: Abteilung Gebäudeverwaltung*; 2010.
36. Bryant RB, Emberlin JC, Norris-Hill J. Vertical variation in pollen abundance in North-Central London. *Aerobiologia*. 1989;5:123–137.
37. Bastl K, Kmenta M, Berger M, Berger U. The connection of pollen concentrations and crowd-sourced symptom data: new insights from daily and seasonal SLI data from 2013–2017 in Vienna. *World Allergy Organ J*. 2018;11(1):24.
38. Werchan B, Werchan M, Mücke H-G, et al. Spatial distribution of allergenic pollen through a large metropolitan area. *Environ Monit Assess*. 2017;189(4):169. <https://doi.org/10.1007/s10661-017-5876-8>.
39. Skjøth CA, Sikoparija B, Jäger S, EAN-Network. Pollen sources. In: Sofiev M, Bergmann K-C, eds. *Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts*. Dordrecht: Springer; 2013:9–27.