ACCEPTED VERION WITH MERGED FIGURES AND FORMAT.

Understanding hourly patterns of *Olea* pollen concentrations as tool for the environmental impact assessment

Santiago Fernández-Rodríguez [a,*], José María Maya-Manzano [b], Alejandro Monroy Colín [c], Raúl Pecero-Casimiro [c], Jeroen Buters [b], José Oteros [d]

[a] *Department of Construction, School of Technology, University of Extremadura, Avda. de la Universidad s/n, Cáceres, Spain*
[b] *Center of Allergy & Environment (ZAUM), Member of the German Center for Lung Research (DZL), Technical University and Helmholtz Center Munich, Munich, Germany*
[c] *Department of Plant Biology, Ecology and Earth Sciences, Faculty of Science, University of Extremadura, Faculty of Science, Avda. Elvas s/n, 06071 Badajoz, Spain*
[d] *Department of Botany, Ecology and Plant Physiology, University of Córdoba, Spain*

Keywords:
Olea pollen
Topic:
Hourly data
Aerobiology
Clustering
Neural networks

a b s t r a c t

Bioinformatics clustering application for mining of a large set of olive pollen aerobiological data to describe the daily distribution of *Olea* pollen concentration. The study was performed with hourly pollen concentrations mea- sured during 8 years (2011–2018) in Extremadura (Spain). *Olea* pollen season by quartiles of the pollen integral in preseason (Q1: 0%–25%), in-season (Q2 and Q3: 25%–75%) and postseason (Q4: 75%–100%). Days with pollen concentrations above 100 grains/m^3 were clustered according to the daily distribution of the concentrations. The factors affecting the prevalence of the different clusters were analyzed: distance to olive groves and the moment during the pollen season and the meteorology. During the season, the highest hourly concentrations during the day where between 12:00 and 14:00, while during the preseason the highest hourly concentrations were de- tected in the afternoon and evening hours. In the postseason the pollen concentrations were more homoge- neously distributed during 9-16 h. The representation shows a well-defined hourly pattern during the season, but a more heterogeneous distribution during the preseason and postseason. The cluster dendrogram shows that all the days could be clustered in 6 groups: most of the clusters shows the daily peaks between 11:00 and 15:00 with a smooth curve (Cluster 1 and 3) or with a strong peak (2 and 5). Days included in cluster 9 shows an earlier peak in the morning (before 9:00). On the other hand, cluster 6 shows a peak in the afternoon, after 15:00. Hourly concentrations show a sharper pattern during the season, with the peak during the hours close to the emission. Out of the season, when pollen is expected to come from farther distances, the hourly peak is lo- cated later from the emission time of the trees. Significant factors for predicting the hourly pattern were wind speed and direction and the distance to the olive groves.

1. Introduction

Bioinformatics are key methods in feature selection techniques to develop in the machine learning and data mining fields (Saeys et al., 2007). Big data is characterized by the volume and variability of the data and the management of the process to analyze for information (Marx, 2013; Obermeyer and Emanuel, 2016). Also big data is being used to estimate the information modeling in biology (Gharajeh, 2018; Hiller and Blanke, 2017; Ho and Giannoulatou, 2019; May, 2017; Rojo et al., 2019a; Schadt et al., 2010). A common technique used is clustering (Aghabozorgi et al., 2015) applied in aerobiological studies to group a large data set (Alcázar et al., 2019; Oteros et al., 2013) and for showing potential pathways of vegetation (Fyfe et al., 2018) and potential long-distance sources of pollen (Makra et al., 2016). Also models based on artificial neural networks are being used for forecasting aerobiological parameters (Astray et al., 2016; Sobol and Finkelstein, 2018; Valencia et al., 2019) as high pollen concentration (Nowosad, 2016), remote sensing index as NDVI (González-Naharro et al., 2019) and for real-time bio-aerosol classification (Leśkiewicz et al., 2018).

The level of temporal precision (hourly scale) is important to consider in environmental studies to plan the time scale, especially in urban aerobiological surveys as; in microclimatic gradients provide evidence for trees (Gubler et al., 2018), to assess the environmental impact assessment (Fernández-Rodríguez et al., 2018), or the phenology environmental (Monroy-Colín et al., 2018), for studying the spatial differentiation (Fernández-Rodríguez et al., 2014b; Fernández-Rodríguez et al., 2014c) the effect of meteorological conditions in pollen grains (Borycka and Kasprzyk, 2018; García-Mozo et al., 2017; Maya Manzano et al., 2017a) and fungal spores (Almeida et al., 2018; Maya-Manzano et al., 2016b). Also personal pollen information concerning medical symptoms are getting more important (Kmenta et al., 2014).

One pollen type, together with *Quercus* and *Poaceae*, common in Extremadura is *Olea* (Fernández-Rodríguez et al., 2014d; Maya-Manzano et al., 2016a). *Olea europaea* is an anemophilous (wind pollinated) species in the Mediterranean region (Galán et al., 2008). The genus *Olea* consist of a large group of species and in Europe *Olea europaea* var. *sylvestris* is found as natural vegetation (Marx, 2013) and *Olea europaea* as a cultivate for oil production (Hernández-Ceballos et al., 2012). The area of olive groves in Spain has increased by 2.6% since 2012, especially in three most important olive oil producing areas; Andalucía (60.38%), Castilla la Mancha (15.83%) and Extremadura (10.49%) (AICA, 2019). The land cover with olive groves of Extremadura exceeds 283,416 ha where rainfed crop land (224,798 ha) and irrigated crop land (58,617 ha) are two other important agricultural land cover areas (AICA, 2019; MAPA, 2019). In Extremadura, the surface of olive trees in intensive and superintensive cultivation has grown 53% between 2014 and 2018 (JEx-Agriculture, 2019). Olive production in Extremadura is steadily increasing due to the increased efficiency with the recent groves of olive trees being managed much more intensive than the traditional crops of olives trees (Larbi et al., 2011; Ravetti, 2014; Rodríguez-Ortega et al., 2017). There is a strong correlation between the amount of olive production and amount of *Olea* pollen (Fernandez-Mensaque et al., 1998). The amount of *Olea* pollen in Extremadura is already high (Fernández-Rodríguez et al., 2014b) and the expected increases in olive production is therefore likely to increase *Olea* pollen concentrations.

Allergic rhinitis and conjunctivitis (ARC) affects 16.9–28.5% of adults in Europe (Bauchau and Durham, 2004). *Olea* pollen can cause ARC (Motreff et al., 2014). Meteorological variables and atmospheric pollution can have an indirect effect on the occurrence of ARC episodes. Quantity of pollen grains in the air depends on meteorological parameters such as relative humidity, temperature, wind and rainfall (Breton et al., 2006; Díaz De La Guardia et al., 2006). Olive pollen allergens are considered to be one of the most important causes of respiratory allergic disease in the South-West Mediterranean region (D'Amato et al., 2007), also documented by several papers on the Iberian Peninsula in Portugal (Fernández-Rodríguez et al., 2015; Ribeiro et al., 2013; Sánchez-Mesa et al., 2005) and Spain (De Linares et al., 2007; Villalba et al., 2014).

Olive trees are not only used for olive and olive oil production (Mohammad and Pooryousef, 2011). They are also used as ornamentals in park landscaping (Staffolani et al., 2011), potentially acting as a source to allergic problems in urban environments (Velasco-Jiménez et al., 2014). Urban green areas are important elements in the planning of modern cities (Grote et al., 2016) and olive trees have been cited as frequent plant species used as ornamental in urban environments (Cariñanos and Casares-Porcel, 2011). Olive trees have been included when estimating the allergenic potential of urban green spaces (Cariñanos et al., 2014) and were taken into account to relate visitors' behavior and characteristics of green spaces in Granada (South-West Spain) (Adinolfi et al., 2014). Studies on both local and regional sources to urban olive pollen concentrations are therefore highly relevant.

Olea pollen season in South-West Spain ranges from April to June, when the timing and intensity of the pollen curve is highly influenced by meteorological parameters (Hernández-Ceballos et al., 2012). *Olea* pollen varies throughout the day and depends, in part, on meteorological parameters in the Mediterranean region (Fernández-Rodríguez et al., 2014c; Fernández-Rodríguez et al., 2014d; Galan et al., 1991; Ribeiro et al., 2008) and atmospheric transport from areas abundant with olive grove (Fernández-Rodríguez et al., 2014b). Galan et al. (1991) showed in Córdoba (South-West Spain) that diurnal olive pollen concentrations recorded in Córdoba peaked during the middle of the day (12:00–18:00) and minimum concentrations were witnessed at night till early in the morning in the morning (20:00–10:00) while nighttime peaks were typically associated with a 10–18 h of atmo- spheric transport from known olive grove regions (Fernández- Rodríguez et al., 2014b). There is therefore a complex interplay be- tween local sources and more remote sources. For remote sources the atmospheric transport since the release may be important.

This study aims to model the pattern *Olea* pollen concentration in five cities (Badajoz, Cáceres, Don Benito, Plasencia and Zafra) of Extremadura (SW Iberian Peninsula) since 2011 to 2018. This is done by examining hourly and daily mean observations of airborne *Olea* pollen concentrations, grouped on preseason, season and postseason (25, 50 and 75%). High (above daily 100 grains/m^3) days were analyzed by clustering analysis and artificial neural network considering the patterns in *Olea* pollen and the relationship with the parameters (distance to olive crops, the seasonality and the meteorology), in order to forecast the influence of each factor.

2. Material and methods

2.1. Sampling sites and inventory of potential olive pollen sources

Extremadura is an agricultural region (41,635 km^2) of the South-West Spain, with a population of 1,072,863 (NSI, 2019). Five cities were studied in the region: in the West, Badajoz (BA), in the North, Plasencia (PL), in the Middle, Cáceres (CC) and Don Benito (DB) and in the South, Zafra (ZA) (Fig. 1). The respective heights, in meters, above sea level are 184 m, 415 m, 280 m and 508 m. The agricultural surroundings of Badajoz are mainly fruit trees and olive groves. In Plasencia, land cover is dominated by holm-oak trees and areas used for extensive grazing. Land cover around Don Benito are irrigated crops (e.g. corn, tomato, rice) and fruit trees. Land cover in the vicinity of Zafra is dominated by olive trees and holm-oak trees and areas used for livestock production.

Olive pollen sources were identified using the Corine Land Cover (CLC) 2012 and 2018 Version 20b2 dataset for the studied area (CLC, 2019), being an agricultural area in permanent crops as olive groves (code 223). A Digital Elevation Model (DEM) has been used to explore the effect of the relief of the landscape. The DEM was produced by analysis a global data set by Natural Earth from the US National Park Service (van Loon, 2008) obtained from the Raster Data (https://www.naturalearthdata.com/downloads/) with spatial resolutions of 10 m (large scale data). The inventory of *Olea* pollen sources in the Iberian Peninsula is shown (Fig. 1 names too small font). Furthermore, the influence of areas of 25, 50, 100 and 200 km in radius surrounding the pollen trap for each city were formed to estimate the influence of the olives groevs for the periods 2012 (beginning of the study) and 2018 (end of the study), by using R statistical software and QGIS software (QGIS 3.4.4).

2.2. Climate and meteorological data

Extremadura has a Mediterranean climate affected by a degree of continentality, with an annual mean temperature between 15 °C and 17 °C; Zafra (15.8 °C), Plasencia (16.1 °C), Badajoz and Don Benito (17.1 °C). The highest temperature is reached in summer, mainly in July and August. Total annual rainfall varies depending on latitude (Plasencia 694 mm, Zafra 619 mm, Don Benito 460 mm and Badajoz 447 mm). Daily and Average climatic data (1981–2010) were provided by the Regional Meteorological Center (AEMET, 2020). Wind data where obtained and managed by using the R package "rwind" (Fernández-López and Schliep, 2019).

2.3. Olea pollen data

Data average (00:00 to 23:59 h) and diurnal (hourly) of airborne *Olea* pollen concentration were obtained from 2011 to 2018 (BA 2011–2018; CC 2016–2018; DB 2011–2014 and 2016–2018; PL 2011–2016; ZA 2011–2018) using Burkard 7-day recording volumetric spore traps (Hirst, 1952) situated on the terraces of School Engineering Agrarian (38.89, −6.96), School of Technology (39.48, −6.34) and Hospital Virgen del Puerto (40.04, −6.08) at 16 m, I.E.S Donoso Cortés (38.96, −5.86) at 6 m and Hospital Zafra (38.43, −6.42) at 6 m about ground level. Daily average and diurnal (hourly). Air is sucked into the trap at a rate of 10 l/min through a 2 mm × 14 mm orifice. Behind the orifice the air flows over a rotating drum that moves past the inlet at 2 mm/h and is covered with an adhesive coating, transparent plastic tape. The sampler used in this study was located on an open terrace with the intake at 1.5 m from the floor. Standardized data management procedures were used, similar to those described by the Spanish Aerobiology Network (REA) (Galán et al., 2007). Daily average (00:00 to 23:59) *Olea* pollen concentrations and diurnal variations (hourly concentrations) were expressed as the number of grains per cubic meter of air (pollen grains/m^3).

2.4. Data analysis

The *Olea* pollen seasons were segmented by quartiles of yearly total pollen amount (25, 50, 75%), terming preseason to the first quartile of the annual pollen (i.e. till reaching 25% of the annual pollen), in-season to the second and third quartiles (25%–75%) and postseason to the fourth quartile. The characteristics of *Olea* pollen seasons were described according to recent recommended terminology and include the start and end dates of the *Olea* pollen season, the high dates for start and end period, the duration (in days), the peak of the maximum day *Olea* pollen

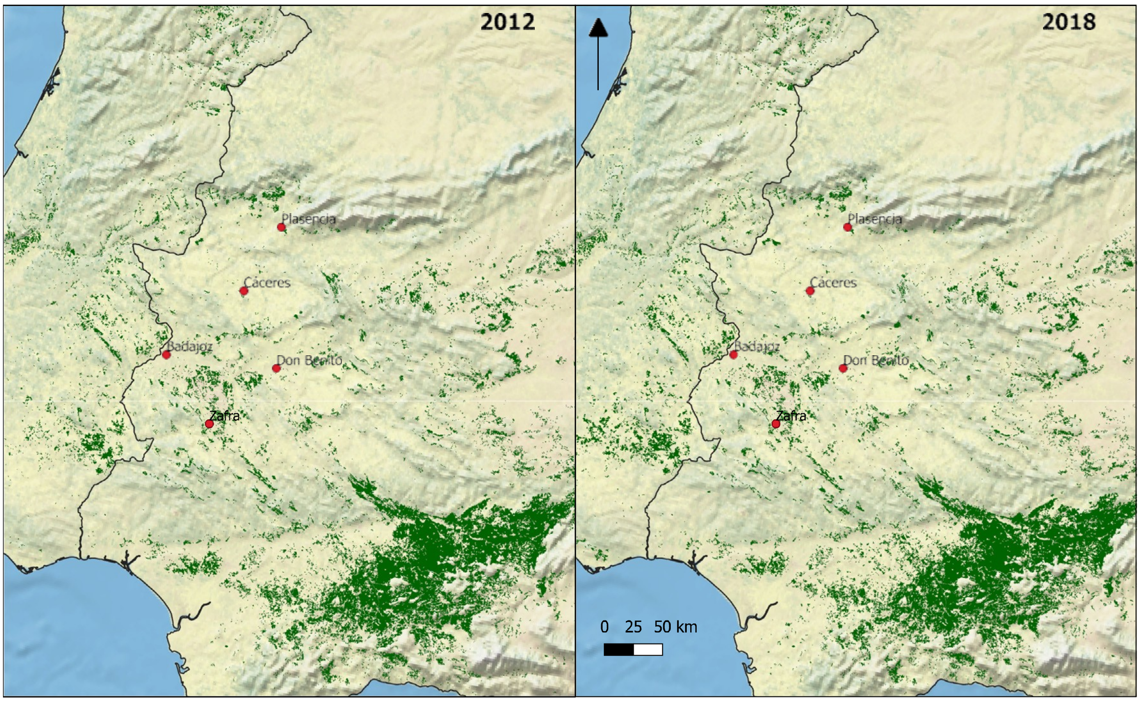


Fig. 1. Mapping sources of olive pollen of the Iberian Peninsula. Olive groves are shown in green colour. Groves areas in 2012 (a) and 2018 (b) are shown.

concentration, the number of days above 100 pollen grains/m^3, and the daily maximum average *Olea* pollen concentration and annual pollen integral (API), which is the sum of daily average airborne *Olea* pollen concentrations recorded during the main pollen season (Galán et al., 2017). This calculation was done using the function "calculate_ps" and the method "clinical" included to the package "Aerobiolgy" in R software (Rojo et al., 2019b). This method considers the start of the olive pollen season as 1st day of 5 days (out of 7 consecutive days) each of these 5 days with ≥20 pollen/m^3 and with a sum of these 5 days of ≥200 pollen/m^3. And the end of the season as the last day of series of 5 days (out of 7 consecutive days) with ≥20 pollen/m3 and with a sum of these 5 days of ≥200 pollen/m3 according to Pfaar et al. (2016)

Only days with a daily average airborne *Olea* pollen concentration N 100 pollen grains/m^3 were selected for this study (Fernández-Rodríguez et al., 2014b). The threshold was an arbitrary number selected for removing days with low daily average values, while still allowing a representative size of the sampled days for analysis or based on daily warning levels (Skjøth et al., 2009). This group of pollen data was analyzed for the mean diurnal (hourly) variation for each of the sites Badajoz, Cáceres, Don Benito, Plasencia and Zafra, respectively. For this purpose, pattern analysis by statistical clustering was applied, to determine the most frequent diurnal (hourly) patterns in *Olea* pollen and the relationship with distance to olive groves, the seasonality and the meteorology (Alcázar et al., 2019; Oteros et al., 2013). This tool could help to identify the potential sources of *Olea* pollen grains in the Iberian Peninsula (Fernández-Rodríguez et al., 2014b; Hernández-Ceballos et al., 2014).

K-means and clustering analysis were applied to the distribution of hourly percentages above the daily 100% for all the data. A Mantel test was applied to determine the optimal k number of clusters in our data set. After clustering pollen hourly concentrations, we also related the hourly patterns with explicatory factors: the day of the year, wind direction, wind speed, amount of olive groves at different distances. This analysis was carried out by training an artificial neural network model using the r package "keras" (Gulli and Pal, 2017).

The results of the influence of each variable on predicting the daily pattern were summarized by the polarity correlation plot. The model is summarized by the accuracy (acc) and the errors summation (loss). In Neural Networks training, the loss value is what the training tries to minimize. It shows the likelihood to assign a low probability to the actual class, while acc gives the percentage of instances that are correctly classified. In our model, each one of the 6 defined daily patters are a class to predict. Training was done with 50% of the data and validation was done with the remaining 50% of the data.

3. Results

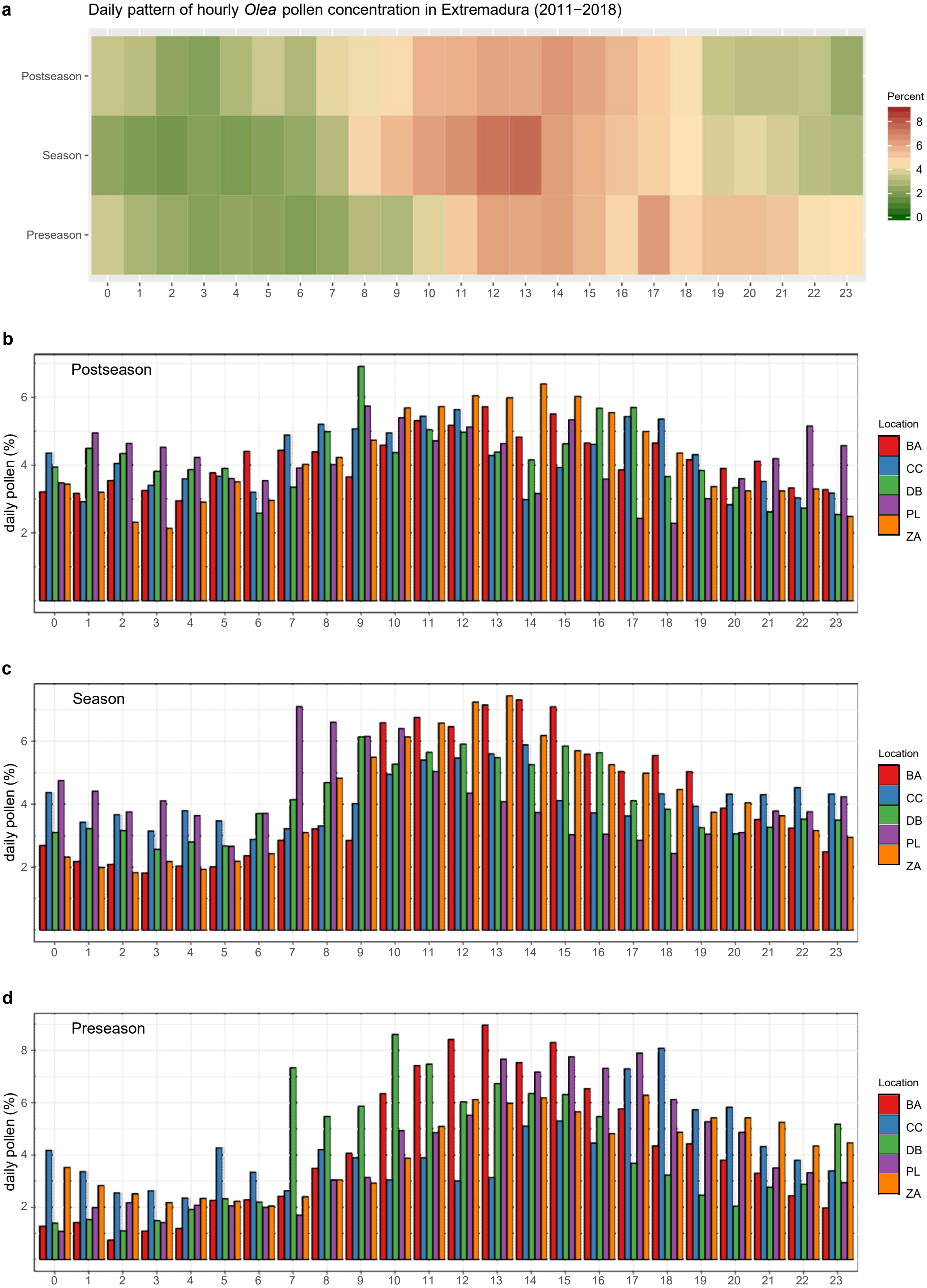
3.1. *Olea* pollen season and area characterization

In Fig. 1 it can be seen that olive groves are located within the area delimited by the studied cities and, furthermore, it is present in areas bordering Extremadura as in Portuguese regions of Portugal Center (at Northwest) and Alentejo (at West), and in Spanish communities of Andalucía (at South and East) and Castilla la Mancha (at East). The highest densities of olives groves in Extremadura is placed between Badajoz, Don Benito (Vegas Bajas and Altas of Guadiana river) and Zafra (Tierra de Barros). In the North, there are lower densities (Vegas Alagón and Tiétar rivers). The local sources of Badajoz are within 100 km distance, local hot spots are mainly placed to the Northwest, West, South-East and South-West. The main sources appear around 50 km

Table 1
Olea pollen season characteristics of Badajoz, Cáceres, Don Benito, Plasencia and Zafra. (Length (days), pollen integral (pollen grains), peak (pollen grains/m^3). Pollen season definition following the clinical method proposed by Pfaar et al. (2016) and calculated with AeRobiology R package (Rojo et al., 2019).

Location	Start date	Start High date	Peak date	End High date	End date	Length	Pollen Integral	Peak	Days N 100
Badajoz	27/04/2011	01/05/2011	11/05/2011	16/05/2011	25/05/2011	29	5608.8	664.8	17
	10/05/2012	13/05/2012	16/05/2012	18/05/2012	05/06/2012	27	2611.8	341.1	11
	07/05/2013	22/05/2013	14/05/2013	26/05/2013	05/06/2013	30	2351.7	396	7
	03/05/2014	05/05/2014	15/05/2014	19/05/2014	26/05/2014	24	3346.2	485.1	10
	03/05/2015	07/05/2015	10/05/2015	20/05/2015	09/06/2015	38	11,571.3	1610.1	15
	13/05/2016	NA	21/05/2016	NA	03/06/2016	22	2250.9	450.9	4
	19/04/2017	23/04/2017	03/05/2017	25/05/2017	27/05/2017	39	14,925.6	1993.5	29
	16/05/2018	17/05/2018	23/05/2018	23/06/2018	26/06/2018	42	8093.7	793.8	30
Cáceres	18/05/2016	24/05/2016	21/05/2016	02/06/2016	13/06/2016	27	2321.1	225.9	10
	23/04/2017	02/05/2017	04/05/2017	21/05/2017	31/05/2017	39	8226.9	1424.7	19
	20/05/2018	22/05/2018	22/05/2018	20/06/2018	26/06/2018	38	3901.5	519.3	15
Don Benito	05/05/2011	05/05/2011	11/05/2011	12/05/2011	13/05/2011	9	3258	1017.9	8
	10/05/2012	16/05/2012	17/05/2012	18/05/2012	10/06/2012	32	3033.9	434.25	7
	17/05/2013	02/06/2013	23/05/2013	05/06/2013	01/07/2013	46	2727	333.9	10
	02/05/2014	05/05/2014	15/05/2014	19/05/2014	19/05/2014	18	4698	559.8	15
	17/05/2016	17/05/2016	21/05/2016	03/06/2016	14/06/2016	29	4330.54	962.1	14
	17/05/2018	20/05/2018	28/05/2018	30/05/2018	31/05/2018	15	3648.6	584.1	13
Plasencia	05/05/2011	10/05/2011	12/05/2011	15/05/2011	26/05/2011	22	2146.5	301.5	11
	19/05/2012	30/05/2012	01/06/2012	01/06/2012	05/06/2012	18	1302.3	290.7	4
	25/05/2013	01/06/2013	05/06/2013	06/06/2013	28/06/2013	35	3284.1	514.8	10
	08/05/2014	10/05/2014	10/05/2014	17/05/2014	20/05/2014	13	1350	183.6	8
	07/05/2015	10/05/2015	12/05/2015	18/05/2015	05/06/2015	30	4273.2	788.4	10
	19/05/2016	26/05/2016	01/06/2016	10/06/2016	14/06/2016	27	4081.5	345.6	17
Zafra	26/04/2011	02/05/2011	16/05/2011	18/05/2011	28/05/2011	33	8657.1	789.3	20
	10/05/2012	17/05/2012	24/05/2012	31/05/2012	06/06/2012	28	3865.5	327.15	16
	08/05/2013	09/05/2013	15/05/2013	07/06/2013	30/06/2013	54	10,720.8	746.1	29
	01/05/2014	04/05/2014	14/05/2014	19/05/2014	21/05/2014	21	8278.2	1353.6	16
	12/05/2015	13/05/2015	14/05/2015	19/05/2015	19/05/2015	8	3110.4	970.2	6
	15/05/2016	21/05/2016	03/06/2016	06/06/2016	07/06/2016	24	5970.6	644.4	20
	01/05/2017	01/05/2017	10/05/2017	26/05/2017	03/06/2017	34	9799.2	1358.1	24
	17/05/2018	27/05/2018	29/05/2018	21/06/2018	25/06/2018	40	11,516.4	1283.4	32

Fig. 2. a) Daily pattern of hourly Olea pollen concentration in Extremadura (2011–2018) at preseason (Q1, 0%–25% of annual pollen), in-season (Q2–Q3, 25%–75% of annual pollen) and postseason (Q4, 75%–100% of annual pollen). Information per station are shown in Fig. 2b (preseason), Fig. 2c (in-season) and Fig. 2d (post-season). Monitoring locations: BA (Badajoz); CC (Cáceres); DB (Don Benito); PL (Palencia) and ZA (Zafra).



a
Daily pattern of hourly Olea pollen concentration in Extremadura (2011−2018)
Postseason
Season
Preseason
Percent
8
6
4
2
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
b
Postseason
daily pollen (%)
Location
BA
CC
DB
PL
ZA
c
Season
daily pollen (%)
Location
BA
CC
DB
PL
ZA
d
Preseason
daily pollen (%)
Location
BA
CC
DB
PL
ZA

in Plasencia mainly at West, Northwest and South-East. For Don Benito and Zafra were found the main olives groves at West, North, South and South-East in the range 100 km. As faraway sources the large quantities of olive trees are near the Guadalquivir Valley in Andalucía (at South-East), which is about 150 km of Zafra, 175 km of Don Benito, 250 km of Badajoz and 350 km of Plasencia. On the other hand, N200 km at North-West of Badajoz there are olives groves in the South and Centre of Portugal.

For the studied years (2011–2018), the timing and the amount of *Olea* pollen recorded in Badajoz, Cáceres, Don Benito Plasencia and Zafra varied annually (Table 1). Average main *Olea* pollen season for all cities ranges from 19th April to 1st July. The duration of the *Olea* pollen season ranged between 8 (Zafra in 2015) and 54 (Zafra in 2013) days. The highest Pollen Integral *Olea* pollen were recorded in Badajoz in 2017, the lowest in Plasencia in 2014. The highest daily average *Olea* pollen concentration was 1993 pollen grains/m^3 recorded in Badajoz on 3 May 2017.

Supplementary material appendix (Fig. S1a–e) show the daily pollen concentration in Badajoz, Cáceres, Don Benito, Plasencia and Zafra. These figures allow to compare the annual distribution between cities. For instance, 2017 was a year with high *Olea* pollen concentration and 2016 recorded similar daily pattern. Throughout the 8-year study in all mentioned cities, daily mean *Olea* pollen concentrations exceeded 100 pollen grains/m^3 on a total of 457 days.

3.2. Clustering hourly pollen concentrations

Fig. 2 shows the daily pattern of hourly *Olea* pollen. Fig. 2a showns which percent of daily total count was present at each hour. We calculated by the average of all the studied hours during the three analyzed periods: pre-season, in-season and post-season. During the season the highest olive pollen concentrations were reached between 12 and 14 h of the day, while during the preseason this was in the afternoon and in the evening. During the postseason the concentrations are more evenly distributed in the central hours of the day (define central hours). The hourly representation shows a defined pattern in the season (Fig. 2c) and more heterogeneous among the stations during the pre-season (Fig. 2d) and postseason (Fig. 2b).

The cluster dendrogram shows the hourly average variation of all days with a daily average *Olea* pollen concentration N 100 pollen grains/m^3 recorded in Badajoz, Cáceres, Plasencia, Don Benito and Zafra (Fig. 3a). This figure represents the cluster dendrogram of all days in 6 set of data: being C1 the highest (271) and C2 the lower (12) matches. Fig. 3b–g indicates the hourly pattern in each set. Analysis showed that there were different hourly peaks; at midday between 11 and 15 h with a smooth curve (cluster 1 and 3) or strong peak (cluster 2 and 5), and earlier, before 9 in the morning, (cluster 4) and later, after 15 in the afternoon (cluster 6).

Supplementary material appendix (Fig. S2) shows a partial explanation of the observed patterns. In the y axis of Fig. S2a we observe the percent that each cluster represents (numbered from 1 to 6) about each part of the season (pre-season, in-season and post-season). The x axis shows the percent of each part of the season (pre-season, in-season and post-season) of the cases of each cluster. For example, the red dot (post-season) referred to cluster 1 (75, 25) shows that 75% of the cases of the post-season period belong to Cluster1, however in Post-season it is only a 25% of the cases. In the same way, cluster 3 means b5% of the cases during the in-season, however almost 100% of the cases reported in cluster 3 are from the in-season period.

Even when the number of days during the in-season are obviously the majority (50%), we observed that some clusters are mainly represented by this period with a much bigger percentage: C3(95%) and C2 (75%). Those are days of the peak during the morning, close to the emission time.

We observed how C5 and C6 are overrepresented during the post-season. Those clusters show the peak during the late afternoon, that could mean that pollen recorded during the post-seasons is coming from farther distances emitted several hours before. C1 is the most common cluster, with a less defined hourly pattern but showing the peak during the middle part of the day where olive emission is expected it is the most common pattern during the whole year.

The Fig. S2b shows the same information but sorting by monitoring place instead of by timing. Curiously, Plasencia, the station located farthest away from most olive groves is overrepresented by C1. On the other hand, Zafra, the station closest to most olive groves in Andalusia, is less defined by C1. Instead, clusters with a better-defined peak are mostly overrepresented by Zafra; i.e. C2 and C5; and under-represented by Plasencia.

Fig. 4 shows the main wind directions and speeds during the studied days (days N100 pollen grains/m^3). We have clustered the days by the timing of the season (Fig. 4a–c) and by hourly cluster (Fig. 4d–i). We have used Badajoz as reference station, as it has the longest time series. As can be observed, days with high pollen during the pre-season are windy days coming from South-West (the Alentejan zone). The wind speed during the season is slower and the winds are coming from prevalently the South-West. Interestingly, during the post-seasonal days, the wind is coming from South-East (the Andalusian groves). About the cluster "Days coinciding with patterns of C3 and C4" are days with strong wind coming from the South-West. In those clusters, the peak is reached earlier during the morning. We observe an interesting phenomenon during days showing a C2 pattern: the wind is coming from South-East.

3.3. Forecasting by artificial neural networks

We developed a forecasting model for the daily pattern based on artificial neural networks. This model allows us to predict the time of the day when the peak will be reached and when the lower concentrations will be reached. Fig. 5a shows the main variable involved in the model: day of the year in combination with the wind speed and the distance to the olive trees as the most important factors to define the hourly distribution of pollen among a day. This can be explained by the fact that during the main pollen season most of the pollen is expected to come from the local sources. However, during out of the main season days, the pollen should arrive from farther distances and so, the daily peak will be delayed with respect to the emission hours. Fig. 5b shows how the model, after 3 epochs can predict the right cluster with an accuracy N75% in both training and testing acc.

4. Discussion

This study focused in the relationship between the distance to the origin of the olives pollen sources and actual *Olea* pollen (hourly and daily) concentration between several cities of SW Iberian Peninsula with heterogeneous densities of potential sources around the sampler points in 8-years study. There are many daily *Olea* pollen studies comparing time series data for a single city with many years, around 20–30 years is more suitable according for climatic comparison 30 years (García-Mozo et al., 2014), 23 (Ruiz-Valenzuela and Aguilera, 2018) and 20 years (Fernández-Rodríguez et al., 2016), but nevertheless less years is being analyzed for comparing between *Olea* pollen monitoring network; 9 (Puljak et al., 2016) and 10 years (Camacho et al., 2017). In this study the hourly *Olea* pollen concentration in a network of stations are considering to study the temporal and spatial pattern of one the most important pollen types in SW Iberian Peninsula (Maya Manzano et al., 2017a). *Olea* flowering season (start and end date) in our study depended on the city and the year; being 29 plus minus how much days on average. Previous studies in the South of Europe indicated similar values (Orlandi et al., 2010; Ruiz-Valenzuela and Aguilera, 2018).

Agricultural cultivation of olives groves is increasing (Erel et al., 2018; Tousa, 2018) in order to produce more olive oil (Colombo and

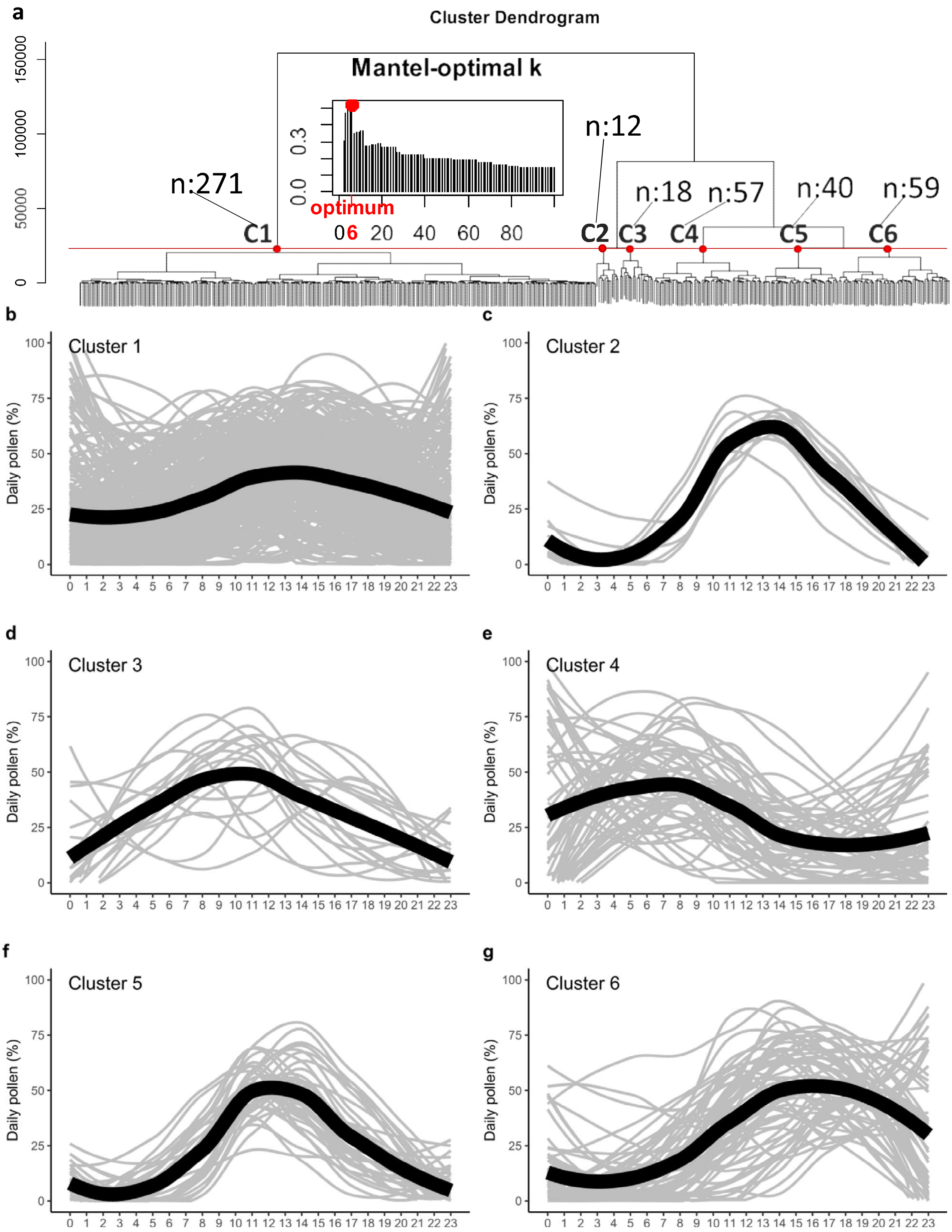


Fig. 3. a) Cluster dendrogram of days N 100 pollen grains/m^3. b–g) Distribution of hourly percentages with respect to the day (100%) on every studied day and grouped by the pollen distribution in clusters from 1 to 6.

Perujo-Villanueva, 2017; Llerena et al., 2012). This tendency is reflected in the pollen spectra in SW Iberian Peninsula cities as Jaén (Ruiz-Valenzuela and Aguilera, 2018) and Córdoba (García-Mozo et al., 2014) in Andalucía, and Badajoz in Extremadura (Fernández-Rodríguez et al., 2016). The amounts of airborne *Olea* recorded in Évora, Portugal in the Alentejo region, in the border with Extremadura (Camacho et al., 2017). Our study is in line with this increase, reflected in the annual pollen integral of studies cities, being Zafra the city with

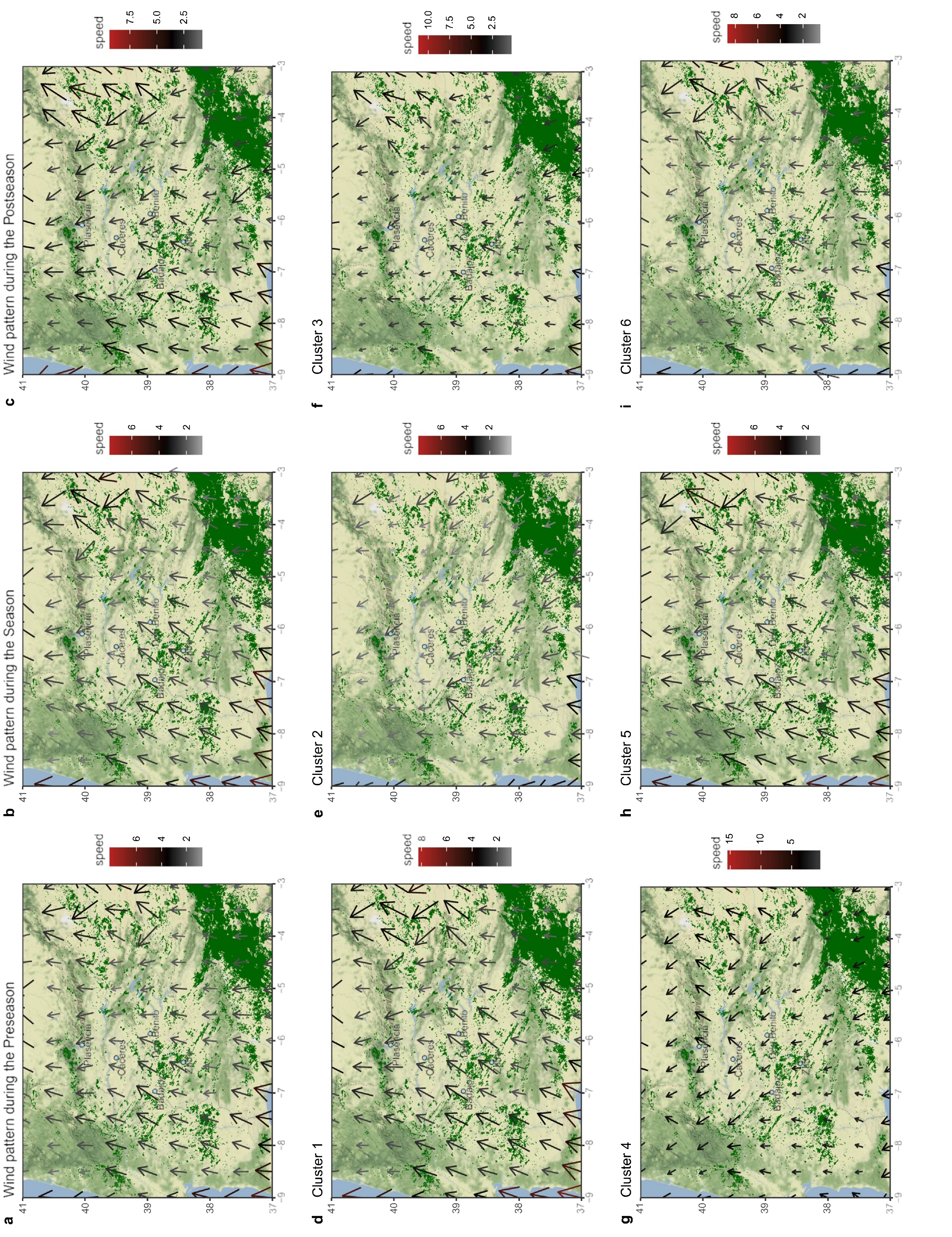


Fig. 4. Average wind direction and wind speed along Extremadura during the studied days (days with N 100 olive pollen grains/m^3). Days are grouped by timing: a) preseason (Q1, 0%–25% of annual pollen), b) in-season (Q2–Q3, 25%–75% of annual pollen) and c) postseason (Q4, 75%–100% of annual pollen). d–i) Days are also grouped in clusters from 1 to 6.

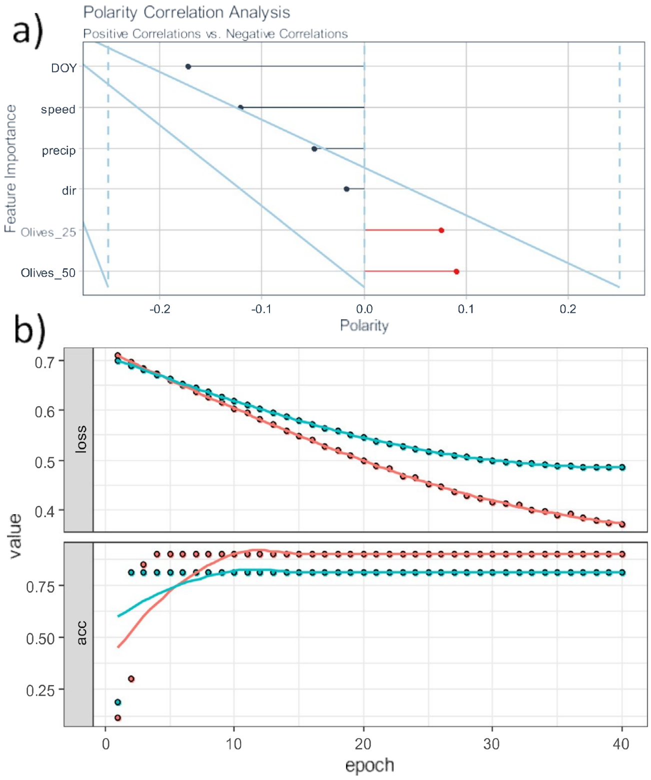


Fig. 5. a) Polarity Correlation Analysis: Main predictor variables in the Artificial Neural Network model. b) Summary parameters of Artificial neural network model applied to classify the clusters: acc and loss values during the training (50% of the cases, red line) and during the validation (50% of the cases, blue line).

the highest values and Plasencia with the lowest average annual values. These results can be explained due to the different densities of olives groves around both cities (Fig. 1). This figure shows a change of olives groves between 2012 and 2018. Buffers around 50 kms show an increase in Badajoz (5.57%), Cáceres (0.04%), Don Benito (1.42%), Plasencia (0.09%) and Zafra (0.64%). This evolution is less evident in the figure than in the *Olea* pollen concentration data. According to the features of Corine Land Cover 2018 the vector CLC database and land uses (in this case olives groves-223) were classified with a 25 ha minimum mapping unit. In the last years the olives groves in intensive and super intensive irrigated (Erel et al., 2018; Tousa, 2018) in Extremadura (JEx-Agriculture, 2019), 54% (2014–2018), and surrodings areas such as Alentejo, Castilla la Mancha and Andalucía (MAPA, 2019). There are many olives groves in Extremadura with less surface (Llerena et al., 2012; Martínez et al., 2014) than the minimum required by CLC for representing in the map. Other important factors to consider is the number of olive trees/ ha depending of crop influenced by the water irrigation, from 80 to 100 (traditional), 200–500 (intensive) to 1500–2500 (super intensive) (Tous et al., 2010). On the other hand, checks based on real observations in the countryside and Corine Land Cover 2018 of recent super intensive plantation olives in Extremadura, around to Badajoz and Don Benito, showing different numbers (221 and 222).

Maya Manzano et al. (2017a) did not find a clear pattern in an hourly *Olea* annual survey between Extremadura and Castilla la Mancha, however land use of olive tree crops in a 50 km area found relationship with *Olea* airborne pollen. In the present study, to model the spatial differentiation of hourly *Olea* pollen concentration and the effect of meteorological conditions was selected the high concentration days. The season period has the highest values at mid-day (10-17 h, being 12-14 h the maximum), while for preseason and postseason the values were more distributed around the afternoon and evening (12–23 h and 10–18 h, respectively), with local differences between the studied cities. This could be attributed to the local effect of the olives sources related with the phenology (Garcia-Mozo et al., 2009; Orlandi et al., 2005). In our case, we adduce in the preseason, due to later hourly pattern, the importance of the far away potential sources olives. Post season could have a similar explanation as pre-season. However, clusters with high central hourly values are probably from local sources as the main contributor. Similar results were found considering the phenology data (Cariñanos et al., 2004; Tormo-Molina et al., 2011) and *Olea* pollen data (Maya Manzano et al., 2017b) in the South Iberian Peninsula. This idea related to the moving daily curve and change potential origin was used in the base of the methodology of back trajectory studies applied on Aerobiology, such as in *Olea* in the study area (Fernández-Rodríguez et al., 2014b). In this previous study were selected 3 episodes to identify several potential sources in a single city for 3 years.

The present study tries to find a regional model, using machine learning, for the region with more cities and years studied. On average of days 100 N pollen grains/m^3 Zafra presented the highest value (20) and Plasencia the least (10). This threshold was used to select *Olea* pollen episodes of high concentration (Galera et al., 2018) for this pollen type in the studied area (Fernández-Rodríguez et al., 2014b). Bilińska et al. (2017) to find source regions of pollen arriving selected between days of low and high values. If, in addition, the time pattern is included, a greater level of detail is available.

Clustering analysis of hourly data was used with this purpose. On the basis of previous aerobiological analyses in the study (Fernández-Rodríguez et al., 2014a) and other close by places (Alcázar et al., 2019; Hernández-Ceballos et al., 2011; Hernández-Ceballos et al., 2014), could be considered episodes of long-distance or regional scale transport to study the temporal cycle of airborne spore concentration and their attribution a to local sources (Skjøth et al., 2012). Furthermore, the regional scale transport of olive pollen can result in increased night-time concentrations of this important aeroallergen (Fernández-Rodríguez et al., 2014a). Long distance transport (LDT) of airborne pollen grains and attempted to identify their sources, e.g.: from pollen types from trees as *Olea* spp.(Fernández-Rodríguez et al., 2014a; Hernández-Ceballos et al., 2011; Hernández-Ceballos et al., 2012; Hernández-Ceballos et al., 2014). The peak of cluster 4 in the morning and the cluster 6 in the afternoon and evening could be due to a pattern related to the LDT of *Olea* pollen, while cluster 2–5, and 1–3 with the main concentration at mid-day are more related with nearby sources of olive trees.

The main limitation of the Corine Land Cover 2018 is the surface mapping minimum unit (25 ha), as there are many existing olives groves with a low extension in Extremadura and surroundings areas. A new line of research is to identify these smaller groves by remote sensing and using geostatistical techniques and geographic information systems for modeling olive flowering (Knezević et al., 2017; Navas-López et al., 2018). This tool provides a reliable interpretation of aerobiological data and of the factors involved in the airborne dispersal of pollen (Rojo and Pérez-Badia, 2015). Recently, there is an increased use of mapping sources on allergenic pollen vegetation related with human health (McInnes et al., 2017), or for modeling olive pollen intensity (Rojo et al., 2016). Remote sensing is related with the mapping sources, being an important advance as a complementary tool for detecting changes in urban areas (Mertes et al., 2015) and in crop mapping (Skakun et al., 2017). The use of remote sensing and GIS has been applied to produce an inventory of grass pollen (Skjøth et al., 2013) and to correlate parameter of images of satellite (NDVI) with oak pollen type (González-Naharro et al., 2019). This kind of satellite information parameter could help in further studies to identify local and small sources of olive pollen, which could be presented in parceled surfaces of small extension, but they could contribute as an important equivalent unit of the CLC catalogue.

The last result of our manuscript is a model based on artificial neural networks to predict the daily pattern of pollen. Artificial Neural Network has been used extensively in aerobiology (Nowosad et al., 2018; Oteros et al., 2013). However, here we aim to predict the hourly pattern of pollen using this technology. To know in advance the hourly distribution of pollen could have significant impact in allergy prevention and public health.

5. Conclusions

The season has the highest values of olive pollen between 12 and 14 h of day, while for preseason the highest olive pollen concentrations are in the afternoon and evening and in postseason more distributed evenly over the central hours of the day. The hourly representation shows a defined pattern during the season and more diffuse pattern in the preseason and postseason. The cluster (c) dendrogram shows 6 sets of hourly average variation with different hourly peaks; at midday between 11 and 15 h with a smooth curve (cluster 1 and 3) and strong peak (cluster 2 and 5), and earlier, before 9 in the morning, (cluster 4) and later, after 15 in the afternoon (cluster 6).

CRediT authorship contribution statement

Santiago Fernández-Rodríguez: Conceptualization, Data curation, Formal analysis, Funding acquisition. José María Maya-Manzano: Project administration, Resources, Software, Supervision. Alejandro Monroy Colín: Investigation, Methodology. Raúl Pecero-Casimiro: Investigation, Methodology. Jeroen Buters: Validation, Visualization. José Oteros: Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was made possible by the research projects PRI16029 and aid for Research Groups GR18113 financed by the Regional Government, Junta de Extremadura (Spain) and the European Social Fund, the José Castillejo Grant from Spanish Govermnent for international mobility (CAS 18/00047) in the Center of Allergy & Environment (ZAUM), Member of the German Center for Lung Research (DZL), Technical University and Helmholtz Center Munich, Germany.

References

Adinolfi, C., Suárez-Cáceres, G.P., Cariñanos, P., 2014. Relation between visitors' behaviour and characteristics of green spaces in the city of Granada, south-eastern Spain. Urban Forestry and Urban Greening 13, 534–542.

AEMET, 2020. Climate Statistics. Badajoz Airport (1981–2010). http://www.aemet.es/es/portada Agencia Estatal de Meteorología, Gobierno de España.

Aghabozorgi, S., Seyed Shirkhorshidi, A., Ying Wah, T., 2015. Time-series clustering - a decade review. Inf. Syst. 53, 16–38.

AICA, 2019. Agency of Information and Food Control of Ministry of Agriculture, Fish, Food and Environment of Gobernment of Spain. http://www.aica.gob.es/.

Alcázar, P., Ørby, P.V., Oteros, J., Skjøth, C., Hertel, O., Galán, C., 2019. Cluster analysis of variations in the diurnal pattern of grass pollen concentrations in Northern Europe (Copenhagen) and Southern Europe (Cordoba). Aerobiologia 35, 269–281.

Almeida, E., Caeiro, E., Todo-Bom, A., Ferro, R., Dionísio, A., Duarte, A., et al., 2018. The influence of meteorological parameters on Alternaria and Cladosporium fungal spore concentrations in Beja (Southern Portugal): preliminary results. Aerobiologia 34, 219–226.

Astray, G., Fernández-González, M., Rodríguez-Rajo, F.J., López, D., Mejuto, J.C., 2016. Airborne castanea pollen forecasting model for ecological and allergological implementation. Sci. Total Environ. 548-549, 110–121.

Bauchau, V., Durham, S.R., 2004. Prevalence and rate of diagnosis of allergic rhinitis in Europe. Eur. Respir. J. 24, 758–764.

Bilińska, D., Skjøth, C.A., Werner, M., Kryza, M., Malkiewicz, M., Krynicka, J., et al., 2017. Source regions of ragweed pollen arriving in south-western Poland and the influence of meteorological data on the HYSPLIT model results. Aerobiologia 1–12.

Borycka, K., Kasprzyk, I., 2018. Hourly pattern of allergenic alder and birch pollen concentrations in the air: spatial differentiation and the effect of meteorological conditions. Atmos. Environ. 182, 179–192.

Breton, M.C., Garneau, M., Fortier, I., Guay, F., Louis, J., 2006. Relationship between climate, pollen concentrations of Ambrosia and medical consultations for allergic rhinitis in Montreal, 1994–2002. Sci. Total Environ. 370, 39–50.

Camacho, I.C., Caeiro, E., Ferro, R., Camacho, R., Câmara, R., Grinn-Gofroń, A., et al., 2017. Spatial and temporal variations in the Annual Pollen Index recorded by sites belonging to the Portuguese Aerobiology Network. Aerobiologia 33, 265–279.

Cariñanos, P., Casares-Porcel, M., 2011. Urban green zones and related pollen allergy: a review. Some guidelines for designing spaces with low allergy impact. Landsc. Urban Plan. 101, 205–214.

Cariñanos, P., Galan, C., Alcázar, P., Domínguez, E., 2004. Airborne pollen records response to climatic conditions in arid areas of the Iberian Peninsula. Environ. Exp. Bot. 52, 11–22.

Cariñanos, P., Casares-Porcel, M., Quesada-Rubio, J.M., 2014. Estimating the allergenic potential of urban green spaces: a case-study in Granada, Spain. Landsc. Urban Plan. 123, 134–144.

CLC, 2019. Corine Land Cover. https://land.copernicus.eu/pan-european/corine-land-cover.

Colombo, S., Perujo-Villanueva, M., 2017. Analysis of the spatial relationship between small olive farms to increase their competitiveness through cooperation. Land Use Policy 63, 226–235.

D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., et al., 2007. Allergenic pollen and pollen allergy in Europe. Allergy 62, 976–990.

De Linares, C., Nieto-Lugilde, D., Alba, F., Díaz De La Guardia, C., Galán, C., Trigo, M.M., 2007. Detection of airborne allergen (Ole e 1) in relation to Olea europaea pollen in S Spain. Clin. Exp. Allergy 37, 125–132.

Díaz De La Guardia, C., Alba, F., De Linares, C., Nieto-Lugilde, D., López Caballero, J., 2006. Aerobiological and allergenic analysis of Cupressaceae pollen in Granada (Southern Spain). J. Investig. Allergol. Clin. Immunol. 16, 24–33.

Erel, R., Yermiyhu, Y., Ben-Gal, A., Dag, A., 2018. Olive fertilization under intensive cultivation management. Acta Hortic. (1217), 207–224.

Fernández-López, J., Schliep, K., 2019. rWind: download, edit and include wind data in ecological and evolutionary analysis. Ecography 42, 804–810.

Fernandez-Mensaque, P.C., González Minero, F.J., Morales, J., Tomas, C., 1998. Forecasting olive (Olea europaea) crop production by monitoring airborne pollen. Aerobiologia 14, 185–190.

Fernández-Rodríguez, S., Skjøth, C.A., Tormo-Molina, R., Brandao, R., Caeiro, E., Silva-Palacios, I., et al., 2014a. Identification of potential sources of airborne Olea pollen in the Southwest Iberian Peninsula. Int. J. Biometeorol. 58, 337–348.

Fernández-Rodríguez, S., Skjøth, C.A., Tormo-Molina, R., Brandao, R., Caeiro, E., Silva-Palacios, I., et al., 2014b. Identification of potential sources of airborne Olea pollen in the Southwest Iberian Peninsula. Int. J. Biometeorol. 58, 337–348.

Fernández-Rodríguez, S., Tormo-Molina, R., Maya-Manzano, J.M., Silva-Palacios, I., Gonzalo-Garijo, Á., 2014c. Comparative study of the effect of distance on the daily and hourly pollen counts in a city in the south-western Iberian Peninsula. Aerobiologia 30, 173–187.

Fernández-Rodríguez, S., Tormo-Molina, R., Maya-Manzano, J.M., Silva-Palacios, I., Gonzalo-Garijo, Á., 2014d. A comparative study on the effects of altitude on daily and hourly airborne pollen counts. Aerobiologia 30, 257–268.

Fernández-Rodríguez, S., Adams-Groom, B., Silva-Palacios, I., Caeiro, E., Brandao, R., Ferro, R., et al., 2015. Comparison of Poaceae pollen counts recorded at sites in Portugal, Spain and the UK. Aerobiologia 31 (1), 1–10.

Fernández-Rodríguez, S., Durán-Barroso, P., Silva-Palacios, I., Tormo-Molina, R., Maya-Manzano, J.M., Gonzalo-Garijo, Á., 2016. Regional forecast model for the Olea pollen season in Extremadura (SW Spain). Int. J. Biometeorol. 60, 1509–1517.

Fernández-Rodríguez, S., Cortés-Pérez, J.P., Muriel, P.P., Tormo-Molina, R., Maya-Manzano, J.M., 2018. Environmental impact assessment of Pinaceae airborne pollen and green infrastructure using BIM. Autom. Constr. 96, 494–507.

Fyfe, R.M., Woodbridge, J., Roberts, C.N., 2018. Trajectories of change in Mediterranean Holocene vegetation through classification of pollen data. Veg. Hist. Archaeobotany 27, 351–364.

Galan, C., Tormo, R., Cuevas, J., Infante, F., Dominguez, E., 1991. Theoretical daily variation patterns of airborne pollen in the southwest of Spain. Grana 30, 201–209.

Galán, C., Cariñanos, P., Alcázar, P., Dominguez-Vilches, E., 2007. Spanish Aerobiology Network (REA) Management and Quality Manual. Servicio de Publicaciones Universidad de Córdoba 978-84-690-6353-8.

Galán, C., García-Mozo, H., Vázquez, L., Ruiz, L., Díaz De La Guardia, C., Domínguez-Vilches, E., 2008. Modeling olive crop yield in Andalusia, Spain. Agron. J. 100, 98–104.

Galán, C., Ariatti, A., Bonini, M., Clot, B., Crouzy, B., Dahl, A., et al., 2017. Recommended terminology for aerobiological studies. Aerobiologia 33, 293–295.

Galera, M.D., Elvira-Rendueles, B., Moreno, J.M., Negral, L., Ruiz-Abellón, M.C., García-Sánchez, A., et al., 2018. Analysis of airborne Olea pollen in Cartagena (Spain). Sci. Total Environ. 622-623, 436–445.

Garcia-Mozo, H., Orlandi, F., Galan, C., Fornaciari, M., Romano, B., Ruiz, L., et al., 2009. Olive flowering phenology variation between different cultivars in Spain and Italy: modeling analysis. Theor. Appl. Climatol. 95, 385–395.

García-Mozo, H., Yaezel, L., Oteros, J., Galán, C., 2014. Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. Sci. Total Environ. 473-474, 103–109.

García-Mozo, H., Hernández-Ceballos, M.A., Trigo, M.M., Galán, C., 2017. Wind dynamics' influence on south Spain airborne olive-pollen during African intrusions. Sci. Total Environ. 609, 1340–1348.

Gharajeh, M.S., 2018. Biological big data analytics. Adv. Comput. 109, 321–355.

González-Naharro, R., Quirós, E., Fernández-Rodríguez, S., Silva-Palacios, I., Maya-Manzano, J.M., Tormo-Molina, R., et al., 2019. Relationship of NDVI and oak (Quercus)pollen including a predictive model in the SW Mediterranean region. Sci. Total Environ. 676, 407–419.

Grote, R., Samson, R., Alonso, R., Amorim, J.H., Cariñanos, P., Churkina, G., et al., 2016. Functional traits of urban trees: air pollution mitigation potential. Front. Ecol. Environ. 14, 543–550.

Gubler, M., Henne, P.D., Schwörer, C., Boltshauser-Kaltenrieder, P., Lotter, A.F., Brönnimann, S., et al., 2018. Microclimatic gradients provide evidence for a glacial refugium for temperate trees in a sheltered hilly landscape of Northern Italy. J. Biogeogr. 45, 2564–2575.

Gulli, A., Pal, S., 2017. Deep Learning with Keras. Packt Publishing Ltd.

Hernández-Ceballos, M., García-Mozo, H., Adame, J., Domínguez-Vilches, E., De la Morena, B., Bolívar, J., et al., 2011. Synoptic and meteorological characterisation of olive pollen transport in Córdoba province (south-western Spain). Int. J. Biometeorol. 55, 17–34.

Hernández-Ceballos, M.A., García-Mozo, H., Adame, J.A., Galán, C., 2012. Last advances in the study of olive airborne pollen dynamic. Olive Consumption and Health, pp. 155–176.

Hernández-Ceballos, M.A., Skjøth, C.A., García-Mozo, H., Bolívar, J.P., Galán, C., 2014. Improvement in the accuracy of back trajectories using WRF to identify pollen sources in southern Iberian Peninsula. Int. J. Biometeorol. 58, 2031–2043.

Hiller, J.S., Blanke, J.M., 2017. Smart cities, big data, and the resilience of privacy. Hastings Law J. 68, 309–356.

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

Ho, J.W.K., Giannoulatou, E., 2019. Big data: the elements of good questions, open data, and powerful software. Biophys. Rev. 11.

JEx-Agriculture. Junta de Extremadura, 2019. Consejería de Medio Ambiente y Políticas Agrarias. http://www.juntaex.es/con03/agricultura.

Kmenta, M., Bastl, K., Jäger, S., Berger, U., 2014. Development of personal pollen information—the next generation of pollen information and a step forward for hay fever sufferers. Int. J. Biometeorol. 58, 1721–1726.

Knezević, M., Zivotić, L., Perović, V., Topalović, A., Todorović, M., 2017. Impact of climate change on olive growth suitability, water requirements and yield in Montenegro. Italian Journal of Agrometeorology 22, 39–52.

Larbi, A., Ayadi, M., Dhiab, A.B., Msallem, M., Caballero, J.M., 2011. Olive cultivars suitability for high-density orchards. Span. J. Agric. Res. 9, 1279–1286.

Leśkiewicz, M., Kaliszewski, M., Włodarski, M., Młyńczak, J., Mierczyk, Z., Kopczyński, K., 2018. Improved real-time bio-aerosol classification using artificial neural networks. Atmospheric Measurement Techniques 11, 6259–6270.

Llerena, J.L., Montaño, A., Garrido, I., 2012. Study of olive oil mills and table olive industries in Extremadura. Acta Hortic. (949), 573–578.

van Loon, A.J., 2008. Geological education of the future. Earth Sci. Rev. 86, 247–254.

Makra, L., Matyasovszky, I., Tusnády, G., Wang, Y., Csépe, Z., Bozóki, Z., et al., 2016. Biogeographical estimates of allergenic pollen transport over regional scales: common ragweed and Szeged, Hungary as a test case. Agric. For. Meteorol. 221, 94–110.

MAPA, 2019. Ministerio de Agricultura, Pesca y Alimentación. Gobierno de España. https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/esyrce/.

Martínez, M., Fuentes, M., Franco, N., Sánchez, J., De Miguel, C., 2014. Fatty acid profiles of virgin olive oils from the five olive-growing zones of extremadura (Spain). JAOCS, Journal of the American Oil Chemists' Society 91, 1921–1929.

Marx, V., 2013. The big challenges of big data. Nature 498, 255–260.

May, M., 2017. Big data, big picture: metabolomics meets systems biology. Science 356, 646–648.

Maya Manzano, J.M., Fernández Rodríguez, S., Vaquero Del Pino, C., Gonzalo Garijo, Á., Silva Palacios, I., Tormo Molina, R., et al., 2017a. Variations in airborne pollen in central and south-western Spain in relation to the distribution of potential sources. Grana 56, 228–239.

Maya Manzano, J.M., Sadyś, M., Tormo Molina, R., Fernández Rodríguez, S., Oteros, J., Silva Palacios, I., et al., 2017b. Relationships between airborne pollen grains, wind direction and land cover using GIS and circular statistics. Sci. Total Environ. 603–613.

Maya-Manzano, J.M., Fernández-Rodríguez, S., Smith, M., Tormo-Molina, R., Reynolds, A., Silva-Palacios, I., et al., 2016a. Airborne Quercus pollen in SW Spain: identifying favourable conditions for atmospheric transport and potential source areas. Sci. Total Environ. 571, 1037–1047.

Maya-Manzano, J.M., Muñoz-Triviño, M., Fernández-Rodríguez, S., Silva-Palacios, I., Gonzalo-Garijo, A., Tormo-Molina, R., 2016b. Airborne Alternaria conidia in Mediterranean rural environments in SW of Iberian Peninsula and weather parameters that influence their seasonality in relation to climate change. Aerobiologia 32, 95–108.

McInnes, R.N., Hemming, D., Burgess, P., Lyndsay, D., Osborne, N.J., Skjøth, C.A., et al., 2017. Mapping allergenic pollen vegetation in UK to study environmental exposure and human health. Sci. Total Environ. 599-600, 483–499.

Mertes, C.M., Schneider, A., Sulla-Menashe, D., Tatem, A.J., Tan, B., 2015. Detecting change in urban areas at continental scales with MODIS data. Remote Sens. Environ. 158, 331–347.

Mohammad, S.M., Pooryousef, M., 2011. A study on traditional usage of olives. J. Appl. Sci. Res. 7, 1186–1187.

Monroy-Colín, A., Silva-Palacios, I., Tormo-Molina, R., Maya-Manzano, J.M., Rodríguez, S.F., Gonzalo-Garijo, Á., 2018. Environmental analysis of airborne pollen occurrence, pollen source distribution and phenology of Fraxinus angustifolia. Aerobiologia 34, 269–283.

Motreff, Y., Golliot, F., Calleja, M., Le Pape, A., Fuhrman, C., Farrera, I., et al., 2014. Short-term effect of pollen exposure on drug consumption for allergic rhinitis and conjunctivitis. Aerobiologia 30, 35–44.

Navas-López, J.F., León, L., Rapoport, H.F., Moreno-Alıás, I., Medina, M.G., Santos, C., et al., 2018. Flowering phenology and flower quality of cultivars 'Arbequina', 'Koroneiki' and 'Picual' in different environments of southern Spain. Acta Hortic. (1229), 257–262.

Nowosad, J., 2016. Spatiotemporal models for predicting high pollen concentration level of Corylus, Alnus, and Betula. Int. J. Biometeorol. 60, 843–855.

Nowosad, J., Stach, A., Kasprzyk, I., Chłopek, K., Dąbrowska-Zapart, K., Grewling, Ł., et al., 2018. Statistical techniques for modeling of Corylus, Alnus, and Betula pollen concentration in the air. Aerobiologia 34, 301–313.

NSI, 2019. Population by Cities. National Institute of Statistics, Madrid, Spain.

Obermeyer, Z., Emanuel, E.J., 2016. Predicting the future-big data, machine learning, and clinical medicine. N. Engl. J. Med. 375, 1216–1219.

Orlandi, F., Ruga, L., Romano, B., Fornaciari, M., 2005. Olive flowering as an indicator of local climatic changes. Theor. Appl. Climatol. 81, 169–176.

Orlandi, F., Garcia-Mozo, H., Galán, C., Romano, B., de la Guardia, C.D., Ruiz, L., et al., 2010. Olive flowering trends in a large Mediterranean area (Italy and Spain). Int. J. Biometeorol. 54, 151–163.

Oteros, J., Garcia-Mozo, H., Hervas-Martinez, C., Galan, C., 2013. Year clustering analysis for modelling olive flowering phenology. Int. J. Biometeorol. 57, 545–555.

Pfaar O, Bastl K, Berger U, Buters J, Calderon M, Clot B, et al. Defining pollen exposure times for clinical trials of allergen immunotherapy for pollen-induced rhinoconjunctivitis—an EAACI Position Paper. Allergy 2016.

Puljak, T., Mamić, M., Mitić, B., Hrga, I., Hruševar, D., 2016. First aerobiological study in Mediterranean part of Croatia (Dalmatia): pollen spectrum and seasonal dynamics in the air of Split. Aerobiologia 32, 709–723.

Ravetti, L.M., 2014. Technology for improving the efficiency of mechanical harvesting in modern olive growing. Acta Hortic. (1057), 221–229.

Ribeiro, H., Oliveira, M., Abreu, I., 2008. Intradiurnal variation of allergenic pollen in the city of Porto (Portugal). Aerobiologia 24, 173–177.

Ribeiro, H., Morales, S., Salmerón, C., Cruz, A., Calado, L., Rodríguez-García, M.I., et al., 2013. Analysis of the pollen allergen content of twelve olive cultivars grown in Portugal. Aerobiologia 29, 513–521.

Rodríguez-Ortega, T., Bernués, A., Olaizola, A.M., Brown, M.T., 2017. Does intensification result in higher efficiency and sustainability? An emergy analysis of Mediterranean sheep-crop farming systems. J. Clean. Prod. 144, 171–179.

Rojo, J., Pérez-Badia, R., 2015. Spatiotemporal analysis of olive flowering using geostatistical techniques. Sci. Total Environ. 505, 860–869.

Rojo, J., Orlandi, F., Pérez-Badia, R., Aguilera, F., Ben Dhiab, A., Bouziane, H., et al., 2016. Modeling olive pollen intensity in the Mediterranean region through analysis of emission sources. Sci. Total Environ. 551-552, 73–82.

Rojo, J., Oteros, J., Pérez-Badia, R., Cervigón, P., Ferencova, Z., Gutiérrez-Bustillo, A.M., et al., 2019a. Near-ground effect of height on pollen exposure. Environ. Res. 160–169.

Rojo, J., Picornell, A., Oteros, J., 2019b. AeRobiology: the computational tool for biological data in the air. Methods Ecol. Evol. 10 (8), 1371–1376.

Ruiz-Valenzuela, L., Aguilera, F., 2018. Trends in airborne pollen and pollen-season-related features of anemophilous species in Jaen (south Spain): a 23-year perspective. Atmos. Environ. 180, 234–243.

Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517.

Sánchez-Mesa, J.A., Serrano, P., Cariñanos, P., Prieto-Baena, J.C., Moreno, C., Guerra, F., et al., 2005. Pollen allerqy in Cordoba city: frequency of sensitization and relation with antihistamine sales. J. Investig. Allergol. Clin. Immunol. 15, 50–56.

Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., Nolan, G.P., 2010. Computational solutions to large-scale data management and analysis. Nat. Rev. Genet. 11, 647–657.

Skakun, S., Franch, B., Vermote, E., Roger, J.C., Becker-Reshef, I., Justice, C., et al., 2017. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. Remote Sens. Environ. 195, 244–258.

Skjøth, C.A., Smith, M., Brandt, J., Emberlin, J., 2009. Are the birch trees in Southern England a source of Betula pollen for North London? Int. J. Biometeorol. 53, 75–86.

Skjøth, C.A., Sommer, J., Frederiksen, L., Gosewinkel Karlson, U., 2012. Crop harvest in Denmark and Central Europe contributes to the local load of airborne Alternaria spore concentrations in Copenhagen. Atmos. Chem. Phys. 12, 11107–11123.

Skjøth, C.A., Ørby, P.V., Becker, T., Geels, C., Schlünssen, V., Sigsgaard, T., et al., 2013. Identifying urban sources as cause of elevated grass pollen concentrations using GIS and remote sensing. Biogeosciences 10, 541–554.

Sobol, M.K., Finkelstein, S.A., 2018. Predictive pollen-based biome modeling using machine learning. PLoS One 13.

Staffolani, L., Velasco-Jiménez, M.J., Galán, C., Hruska, K., 2011. Allergenicity of the ornamental urban flora: ecological and aerobiological analyses in Córdoba (Spain) and Ascoli Piceno (Italy). Aerobiologia 27, 239–246.

Tormo-Molina, R., Silva-Palacios, I., Gonzalo-Garijo, A., Moreno-Corchero, A., Pérez-Calderón, R., Fernández-Rodríguez, S., 2011. Phenological records as a complement to aerobiological data. Int. J. Biometeorol. 55, 51–65.

Tous, J., Romero, A., Hermoso, J.F., 2010. New trends in olive orchard design for continuous mechanical harvesting. Adv. Hortic. Sci. 24, 43–52.

Tousa, J., 2018. High planting density trial with olive cultivar 'Arbequina'. Acta Hortic. 1199, 285–290.

Valencia, J.A., Astray, G., Fernández-González, M., Aira, M.J., Rodríguez-Rajo, F.J., 2019. Assessment of neural networks and time series analysis to forecast airborne Parietaria pollen presence in the Atlantic coastal regions. Int. J. Biometeorol. 63, 735–745.

Velasco-Jiménez, M.J., Alcázar, P., Valle, A., Trigo, M.M., Minero, F., Domínguez-Vilches, E., et al., 2014. Aerobiological and ecological study of the potentially allergenic ornamental plants in south Spain. Aerobiologia 30, 91–101.

Villalba, M., Rodríguez, R., Batanero, E., 2014. The spectrum of olive pollen allergens. From structures to diagnosis and treatment. Methods 66, 44–54.