

Modeling pollen time series using seasonal-trend decomposition procedure based on LOESS smoothing

Jesús Rojo¹  · Rosario Rivero¹ · Jorge Romero-Morte¹ · Federico Fernández-González¹ · Rosa Pérez-Badía¹

Received: 15 February 2016 / Revised: 12 July 2016 / Accepted: 13 July 2016 / Published online: 4 August 2016
© ISB 2016

Abstract Analysis of airborne pollen concentrations provides valuable information on plant phenology and is thus a useful tool in agriculture—for predicting harvests in crops such as the olive and for deciding when to apply phytosanitary treatments—as well as in medicine and the environmental sciences. Variations in airborne pollen concentrations, moreover, are indicators of changing plant life cycles. By modeling pollen time series, we can not only identify the variables influencing pollen levels but also predict future pollen concentrations. In this study, airborne pollen time series were modeled using a seasonal-trend decomposition procedure based on LOcally wEighted Scatterplot Smoothing (LOESS) smoothing (STL). The data series—daily Poaceae pollen concentrations over the period 2006–2014—was broken up into seasonal and residual (stochastic) components. The seasonal component was compared with data on Poaceae flowering phenology obtained by field sampling. Residuals were fitted to a model generated from daily temperature and rainfall values, and daily pollen concentrations, using partial least squares regression (PLSR). This method was then applied to predict daily pollen concentrations for 2014 (independent validation data) using results for the seasonal component of the time series and estimates of the residual component for the period 2006–2013. Correlation between predicted and observed values was $r = 0.79$ (correlation coefficient) for the pre-peak period (i.e., the period prior to the peak pollen concentration) and $r = 0.63$ for the post-peak period. Separate analysis of each of the components of the pollen data series enables the sources of variability to be

identified more accurately than by analysis of the original non-decomposed data series, and for this reason, this procedure has proved to be a suitable technique for analyzing the main environmental factors influencing airborne pollen concentrations.

Keywords Seasonality · Daily variations · Grass pollen · Flowering phenology · Meteorological variables

Introduction

Airborne pollen levels are a key indicator of flowering in wind-pollinated plants. Monitoring of pollen concentrations provides valuable information on plant phenology and productivity and is thus a useful tool in botanical, environmental, and agricultural research (García-Mozo 2011; Fernández-Llamazares et al. 2014). Monitoring is also of public health interest, since pollen is the leading cause of allergies (D’Amato et al. 2007), and the number of allergy-sufferers has considerably increased over the last few decades (Bousquet et al. 2011). Modeling airborne pollen concentrations helps to identify the variables or factors influencing variations over time (Aboulaich et al. 2013; Rojo et al. 2015) and also to predict future pollen levels (Rodríguez-Rajo et al. 2006; Silva-Palacios et al. 2015).

The flowering sequence of anemophilous species determines the airborne pollen concentrations (Estrella et al. 2006), and the pollen spectrum is related to the geographical characteristics, flora, and vegetation of a given site (Kasprzyk 2006; Rojo et al. 2015). Central Spain is characterized by a great diversity and abundance of plant communities and species, which results in high pollen amounts and an important pollen diversity (Pérez-Badía et al. 2010). However, the daily pollen concentrations are also governed by the meteorological

✉ Jesús Rojo
jesus.rojo.ubeda@gmail.com

¹ Institute of Environmental Sciences, University of Castilla-La Mancha, 45071 Toledo, Spain

conditions influencing local pollen emission and dispersal (Sánchez-Mesa et al. 2005). Temperature is one of the most important variables that influence the pollen emission, and other meteorological variables such as rainfall or wind speed and direction influence the pollen dispersal (Rodríguez-Rajo et al. 2003; Aboulaich et al. 2013).

Analysis of aerobiological data series is based on the continuous measurement of airborne biological particle (e.g., pollen and spore) concentrations, with a view to interpreting variations over time (Peel et al. 2014; Hamid et al. 2015). These data can therefore be processed using time-series analysis techniques, and a number of methods have been developed based on distinguishing between the deterministic (seasonal) component and the stochastic (residual) component of data series (Belmonte and Canela 2002; Aznarte et al. 2007; Ocaña-Peinado et al. 2008).

Research has particularly focused on Poaceae pollen (Stach et al. 2008; Kasprzyk and Walanus 2010; Fernández-Rodríguez et al. 2015), for two major reasons: it is the most common cause of pollen allergies in Europe (D'Amato et al. 2007), and the Poaceae family comprises a large number of species with a broad range of distribution, together accounting for roughly 20 % of world plant cover (Mabberley 1987). Due to the numerous grass species involved, the Poaceae pollen season is both very long (Pérez-Badia et al. 2010) and highly complex, with marked variations in concentrations and a considerable randomly varying or stochastic component, which can only be estimated by reference to environmental variables (Sabariego et al. 2012; Pérez-Badia et al. 2013).

One procedure commonly used to analyze time series is data decomposition (Brockwell and Davis 2002), which yields a number of components reflecting different sources of variability (Currie et al. 2011; Petropavlovskikh et al. 2015). Seasonal-trend decomposition based on LOcally wEighted Scatterplot Smoothing (LOESS), known as STL decomposition, is a filtering procedure for decomposing a seasonal times series into three components: trend, seasonal, and remainder or residual. Advantages of this technique include simplicity and speed of computation, robustness of results, and flexibility in the period of the seasonal component (Cleveland et al. 1990). This decomposition procedure, widely used in the natural sciences (e.g., Currie et al. 2011; Vallejos et al. 2013; Bourjea et al. 2015), has recently become more popular in aerobiological research (García-Mozo et al. 2014, 2016; Aguilera et al. 2015; Rojo et al. 2015).

A number of papers report on the prediction of airborne Poaceae pollen concentrations; time-series methods have been used in some cases (Moseholm et al. 1987; Voukantsis et al. 2010; Brighetti et al. 2014) but not in others (Sánchez-Mesa et al. 2002; Aboulaich et al. 2013). Other studies have examined variations in pollen concentrations as a function of changes in meteorological variables (Rojo et al. 2015) or in the life cycle (flowering) of species regarded as the main

sources of pollen emission (Frenguelli et al. 2010; León-Ruiz et al. 2011).

This study sought to chart the correlation between Poaceae pollen data and flowering phenology data for the most abundant and characteristic species of Poaceae in a representative area of central Spain and to model pollen time series in order to identify the meteorological variables shaping each of the components in the pollen data series over the period 2006–2014. It also sought to generate a model to predict daily pollen concentrations, by adding the seasonal component of the series to the residual component estimated on the basis of meteorological variables.

Material and methods

Case study

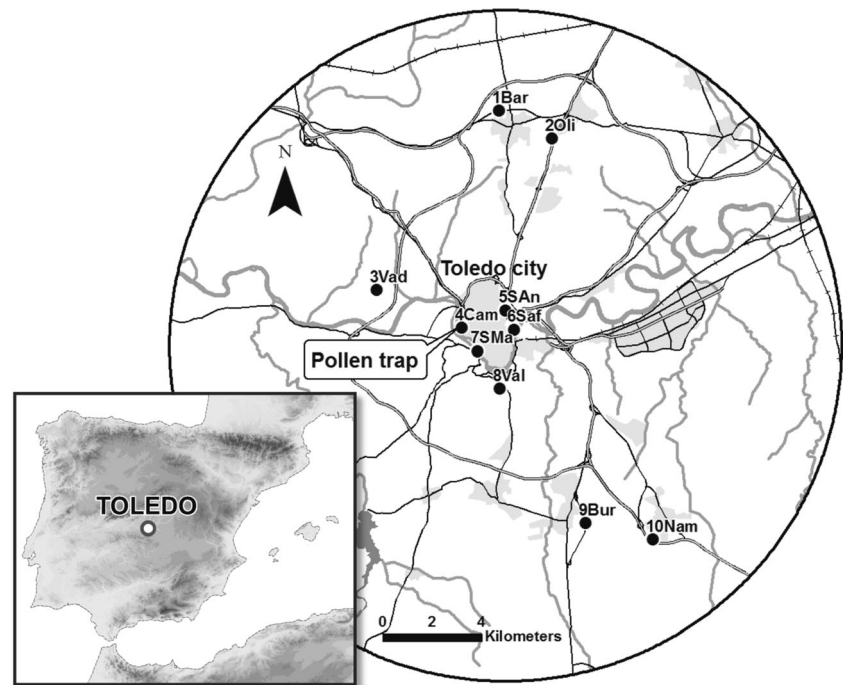
Daily airborne Poaceae pollen concentrations were recorded from 2006 to 2014 using a Hirst-type volumetric trap (Hirst 1952) located in the city of Toledo (Fig. 1), following the protocol recommended by the Spanish Aerobiology Network (Galán et al. 2007). Data on maximum temperature, minimum temperature, and rainfall for the city over the study period were provided by the Spanish Meteorological Agency (AEMET).

Field observations of flowering phenology were carried out in 2013. Phenological sampling was performed at ten selected sampling sites in the vicinity of Toledo, located between 200 m (closest) and 18 km (furthest) from the pollen trap (Table 1), at heights ranging from 450 to 686 m a.s.l. Sites contain various types of habitat characteristic of *Poaceae* species, including the following: holm oak forests with clearings mainly covered by xeric grasslands and shrublands, anthropic areas of ruderal grassland in urban and periurban areas, weed communities associated to the extensive croplands surrounding the city, and finally hygrophilous grasslands with *Poaceae* species.

Phenological data

Weekly phenological observations were based on estimating flowering intensity for the *Poaceae* species studied, measured as the proportion of flowering individuals (%): for this purpose, 25 individuals were selected per species, and the proportion of individuals in bloom each week was recorded (percentage of flowering individuals (PFI)). Using an adaptation of the Biologische Bundesanstalt, Bundessortenamt und Chemische industrie (BBCH) scale for phenological development (Meier 2001), flowering was deemed to begin when the first anthers were visible (BBCH code 61) and to end when all anthers were dehydrated and all pollen had been released from inflorescences (BBCH code 69). A continuous series of

Fig. 1 Study area. Phenological sites in the province of Toledo (central Spain): 1Bar Bargas, 2Oli Olias del Rey, 3Vad Valdelobos, 4Cam Campus University, 5San St. Anton park, 6Saf Safont park, 7Sma St. Martin bridge, 8Val Tagus valley, 9Bur Burguillos, 10Nam Nambroca. The pollen trap is located in the city of Toledo (coordinates: 39° 51' 55" N, 4° 2' 31" W)



daily data was generated from weekly PFI values by linear interpolation, as used in previous phenological research (Oteros et al. 2013b; Rojo and Pérez-Badía 2015).

Species were selected as a function of their abundance or representativeness at each sampling site, using an adaptation of the method recommended by León-Ruiz et al. (2011): 25 individuals/m² were selected for annuals and 25 individuals/10 m² for perennials. As a result, the species sampled varied from site to site, depending on the type of habitat involved

(Table 1). A total of 25 species were studied: *Aegilops geniculata* Roth, *Aegilops triuncialis* L., *Arrhenatherum elatius* (L.) P. Beauv. ex J. Presl & C. Presl subsp. *bulbosum* (Willd.) Schübl. & G. Martens, *Avena barbata* Link, *Avena sterilis* L., *Bromus diandrus* Roth, *Bromus hordeaceus* L., *Bromus rubens* L., *Bromus squarrosus* L., *Bromus tectorum* L., *Cynosurus elegans* Desf., *Dactylis glomerata* L. subsp. *hispanica* (Roth) Nyman, *Echinaria capitata* (L.) Desf., *Elymus pungens* (Pers.) Melderis subsp. *campestris* (Godr.

Table 1 Geographical characteristics of the phenological sites and sampled species

No.	Site	Coordinates	Elevation (m a.s.l.)	Sampled species ^a
1	Bargas	39° 56' 45" N 4° 1' 32" W	581	AvSt, DaHi, LoRi, MiTe, TrPa
2	Olias del Rey	39° 56' 9" N 4° 0' 2" W	623	BrRu, DaHi, LoRi, RoCr, TrPa
3	Valdelobos (Toledo city)	39° 52' 48" N 4° 4' 58" W	514	ArBu, BrSq, DaHi, EcCa, ElCa, RoCr, MaTe
4	Campus (Toledo city)	39° 51' 57" N 4° 2' 37" W	450	BrDi, HoLe, PiMi, PoAn, TrPa
5	St. Anton park (Toledo city)	39° 52' 25" N 4° 1' 16" W	494	AvBa, AvSt, BrDi, BrRu, CyEl, DaHi, HoLe, LoRi, RoCr, TrPa
6	Safont park (Toledo city)	39° 51' 28" N 4° 1' 3" W	470	AvBa, BrDi, PiMi, PoAn, RoCr, TrPa
7	St. Martin bridge (Toledo city)	39° 51' 28" N 4° 2' 3" W	455	PiMi, PoAn, StCa
8	Tagus valley (Toledo city)	39° 50' 41" N 4° 1' 12" W	600	AeGe, ArBu, BrHo, BrTe, DaHi, EcCa, LoRi, StCa, VuBr, VuCi, MaTe
9	Burguillos	39° 47' 43" N 3° 58' 55" W	681	BrRu, BrTe, DaHi, RoCr, TrPa, VuCi
10	Nambroca	39° 47' 24" N 3° 57' 1" W	686	AeGe, AeTr, HoLe, LoRi

^a Abbreviations of sampled species: AeGe *Aegilops geniculata*, AeTr *A. triuncialis*, ArBu *Arrhenatherum elatius* subsp. *bulbosum*, AvBa *Avena barbata*, AvSt *A. sterilis*, BrDi *Bromus diandrus* Roth, BrHo *B. hordeaceus*, BrRu *B. rubens*, BrSq *B. squarrosus*, BrTe *B. tectorum*, CyEl *Cynosurus elegans*, DaHi *Dactylis glomerata* subsp. *hispanica*, EcCa *Echinaria capitata*, ElCa *Elymus pungens* subsp. *campestris*, HoLe *Hordeum murinum* subsp. *leporinum*, LoRi *Lolium rigidum*, MaTe *Macrochloa tenacissima*, MiTe *Micropyrum tenellum*, PiMi *Piptatherum miliaceum*, PoAn *Poa annua*, RoCr *Rostraria cristata*, StCa *Stipa capensis*, TrPa *Trisetum paniceum*, VuBr *Vulpia bromoides*, VuCi *V. ciliata*

& Gren.) Melderis, *Hordeum murinum* L. subsp. *leporinum* (Link) Arcang., *Lolium rigidum* Gaudin, *Macrochloa tenacissima* (L.) Kunth, *Micropyrum tenellum* (L.) Link, *Piptatherum miliaceum* (L.) Coss., *Poa annua* L., *Rostraria cristata* (L.) Tzvelev, *Stipa capensis* Thunb., *Trisetum paniceum* (Lam.) Pers., *Vulpia bromoides* (L.) Gray, and *Vulpia ciliata* Dumort.

Factor analysis was used to reduce dimensionality in the flowering data for the Poaceae species (Rummel 1988; Reymont and Jvreskog 1996), with a view to obtaining a single phenological variable for each species. Given the similarity of flowering curves (PFI) for each species at different sites, the first principal component in the factor analysis accounted in all cases for over 90 % of variance (see example in Fig. 2). A single factor containing 90 % of the PFI information for a given species could thus be obtained by the following equation:

$$F_s = \sum_{i=1}^z a_i X_i$$

where F_s is the factor containing most PFI information for each species s ; a is the coefficient for each variable X , based on the estimated PFI at each sampling site; i is the sampling site; and z is the total number of sampling sites for the species.

Decomposition of aerobiological data

The aerobiological variable (pollen concentration) was transformed by decomposition of the time series (Brockwell and Davis 2002). First, the structure (additive vs. multiplicative models) and seasonality (stationary vs. non-stationary) of the time series were defined using Holt-Winters filtering (Winters 1960) and the Ljung-Box test (Ljung and Box 1978), respectively.

Using an additive model, the daily pollen-concentration time series (Z_t) results from the summation of three components (Harvey and Peters 1990):

$$Z_t = M_t + S_t + R_t \quad t = 1, \dots, T$$

where M_t is the trend, S_t is the seasonality, and R_t is the residual.

This structural model enables each of the components to be isolated and analyzed separately. Here, a seasonal-trend decomposition procedure based on LOESS (STL) was used to analyze daily pollen concentrations over an 8-year period (2006–2013).

STL is a filtering procedure for decomposing time series, based on a sequence of smoothing procedures using a locally weighted regression, commonly known as LOESS (Cleveland et al. 1990). The LOESS smoother is based on fitting a weighted polynomial regression for a given time of observation, where weights decrease with distance from the nearest neighbor (Dagum and Luati 2003). LOESS produces a smoothed estimate ($\hat{y}_j$) that is defined by the following:

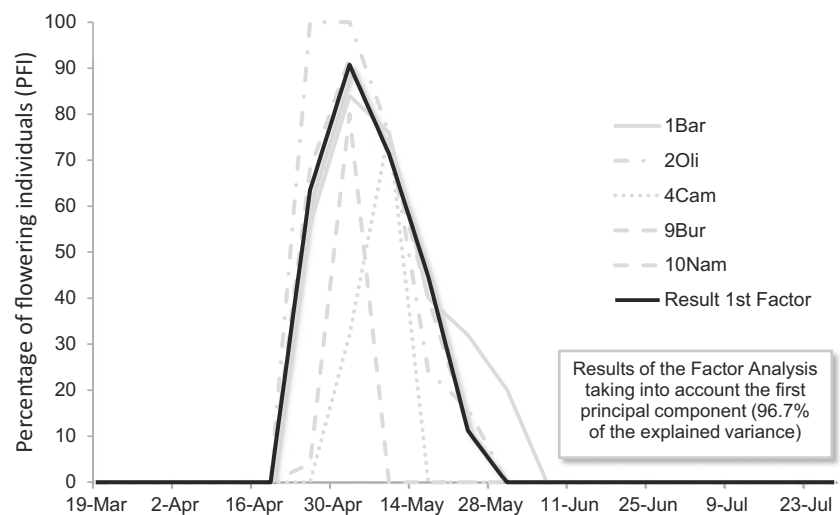
$$\hat{y}_j = \sum_{p=0}^d \beta_{pj} t_j^p \quad j = 1, \dots, n$$

where β_j is the $(d+1)$ -dimensional least squares estimate of the weighted regression, t_j^p is the $(d+1)$ -dimensional vector of the time of observation, j is the number of time lags up to the maximum defined by the smoothing parameter (n), $p = 0, \dots, d$, and d is the degree of the polynomial fitting.

The time series is fitted iteratively until trend and seasonality stabilize, in a multi-step process in which moving averages alternate with LOESS smoothing. At the end of the STL process, the seasonal and trend components are extracted from the data series, in accordance with the following expression:

$$S_t = S_t^{(k+1)} ; M_t = M_t^{(k+1)} ; R_t = Z_t - S_t - M_t$$

Fig. 2 Factor analysis for the species *Rostraria cristata* during the year 2013. Phenological sites: 1Bar Bargas, 2Oli Olías del Rey, 4Cam Campus University, 9Bur Burguillos, 10Nam Nambroca



where k is the number of iterations in which the various steps take place. For a detailed description of the STL method, see Cleveland et al. (1990) and Zhou et al. (2015).

The STL decomposition method requires definition of the smoothing parameter for the seasonal component (n_s) (Cleveland et al. 1990); here, it was set at $n_s \rightarrow \infty$, thus forcing the seasonal component to be periodic and to be constant across all the years in the time series, thus enabling it to be used in predicting future pollen concentrations. Time-series analysis was performed using R software (R Core Team 2015).

Modeling pollen concentrations

Decomposition of the time series into its various components enabled each component to be independently modeled. Models were developed by partial least squares regression (PLSR), using a non-linear kernel algorithm. This method is appropriate for models containing numerous intercorrelated independent variables (Wold et al. 2001).

PLSR creates a block of predictors (X) from the independent variables included in the model (P), using principal components (T). This block is formed by orthogonal factors (i.e., independent of each other), commonly known as latent variables:

$$X = TP + \text{Residuals}$$

Each principal component is calculated from the weight (w) of each independent variable (x) in that factor, computed from the covariance between the response variable (y) and the independent variables:

$$t_i = \sum_{a=1}^p w_{ia} X_a \quad ; \quad w_{ij} = \frac{\text{cov}(x_j, y)}{\sqrt{\sum_{j=1}^p \text{cov}^2(x_j, y)}}$$

$$t_1 = w_{11}x_1, \dots, w_{1p}x_p$$

...

$$t_z = w_{z1}x_1, \dots, w_{zp}x_p$$

where p is the maximum number of independent variables included in the model and z is the maximum number of principal components selected.

Next, the latent variables containing the information from the predictors are used to estimate the response variable:

$$y = TC + \text{Error}$$

where C is a matrix of correlation coefficients for the predictor block and the response variable block.

This study sought to identify potential correlations between data for the trend and seasonality components ($M_t + S_t$) of the data series for the period 2006–2013 with the flowering

phenology variable (PFI) for the grasses studied in 2013. For that purpose, PLSR took into account the flowering of Poaceae species for which a positive significant correlation was observed (Spearman test).

The residual or random component (R_t) was modeled using data for daily pollen concentrations and for meteorological variables measured from 1 day prior to each time point in the series ($x-1$) up to 3 days before that time point ($x-3$). Modeling was carried out from the PS (PS) start date, defined—using an adaptation of the approach recommended by Galán et al. (2001) for olive pollen—as the first day on which at least 10 grains/m³ were recorded with ≥ 10 grains/m³ on the following 2 days. This definition was better suited to the aims of the present study (Jato et al. 2006), in that it enabled the elimination of a large number of earlier days with small intermittent amounts of pollen, which seriously hindered modeling. The residual component was modeled separately for the pre-peak period (i.e., prior to the peak pollen-concentration day) and the post-peak period (i.e., after the peak pollen-concentration day). The model was calibrated for the period 2006–2013.

Finally, daily pollen concentrations for 2014 were predicted using data not used for model calibration. For this purpose, the residual component was estimated using the model and the series was reconstructed by adding the deterministic components (seasonality and trend). The model was validated by comparing the predicted data for 2014 with real pollen concentrations (i.e., raw non-decomposed data) for that year. Relationships and differences between real and predicted pollen curves were determined using the Spearman test (correlation between data series) and the Wilcoxon signed-rank test (differences between paired samples).

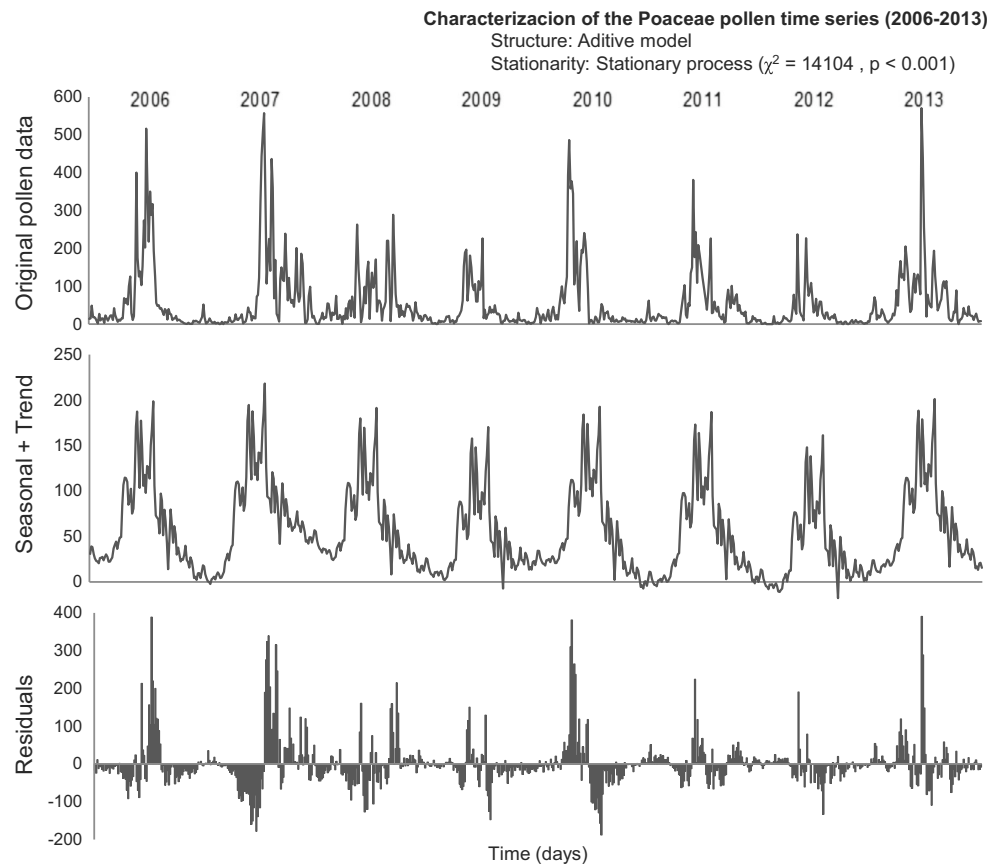
Calibration and validation of the aerobiological data model were performed using the “pls” package (Mevik and Wehrens 2007) and R software (R Core Team 2015).

Results

The Poaceae pollen-concentration series for the period 2006–2013 was an additive time series with a clear cyclic annual pattern (Fig. 3). The seasonality of the series did not vary over time, and no significant trend pattern was observed; for that reason, the trend component was not subjected to separate analysis. The results of the Ljung-Box test confirmed that data matched a stationary seasonal series (Fig. 3).

The daily pollen curve for 2013 reflected the staggered flowering of different grass species; concentrations differed considerably depending on the species in flower at any given moment (Fig. 4). Low peaks were recorded in mid-April 2013, coinciding with the flowering of the earliest Poaceae species such as *Hordeum murinum* subsp. *leporinum* or species of the genera *Avena*, *Bromus*, and *Vulpia*. Pollen

Fig. 3 Poaceae pollen time series for the period 2006–2013. χ^2 statistics from the Ljung-Box test



concentrations rose throughout May, due to the flowering of grasses such as *Macrochloa tenacissima*, *Echinaria capitata*, and *R. cristata* (Fig. 4). The highest concentrations were observed in late May and early June (2013 peak concentration: 26 May), coinciding with the flowering of late species including *Arrhenatherum elatius* subsp. *bulbosum*, *L. rigidum*, *T. paniceum*, *D. glomerata* subsp. *hispanica*, *Piptatherum miliaceum*, and various *Aegilops* species.

Then, from 26 May onward, pollen concentrations began to decline, as flowering gradually ended in the later grass species; by mid-July 2013, the amount of airborne grass pollen had ceased to be significant (Fig. 4). Comparison of percentage of flowering individuals (PFI) data over the flowering period for each species and airborne pollen concentrations for the same period revealed that later-flowering species and *Poa annua*—which flowers over most of the pollen season (Fig. 4)—displayed the strongest correlation with pollen concentrations (Table 2).

The PFIs for all these grass species were used to model pollen concentrations (Fig. 5). Models constructed using PLSR on phenological variables accounted for 35 % of variance in the original (non-decomposed) data series for 2013 (Fig. 5a). However, when the series was decomposed and the residual component was removed, the model improved considerably—accounting for 79 % of variance (Fig. 5b).

Modeling of the residual component, for calibration purposes, showed that this component accounted for 58 % of overall variance in the pre-peak period (Fig. 6) and for 69 % in the post-peak period (Fig. 7). The importance of the variables included in the model, as expressed by regression coefficients, varied between pre-peak and post-peak models, although, in most cases, correlations shared the same sign for both periods (Fig. 8). The most influential variable was pollen concentration on the previous day (*Poac-1*). Maximum temperature on the previous day (*Tmax-1*) also exerted an important positive influence on pollen concentrations, whereas minimum temperature (*Tmin*) and rainfall (*R*) generally had a negative effect (Fig. 8).

Finally, the model was validated using independent pollen data for 2014; the Spearman correlation coefficient between observed and predicted data was $r = 0.79$ for the pre-peak period and $r = 0.63$ for the post-peak period (Fig. 9). The Wilcoxon signed-rank test revealed no significant difference between observed and estimated data series (Fig. 9).

Discussion

Aerobiological data series contain a periodic component which enables them to be modeled as time series

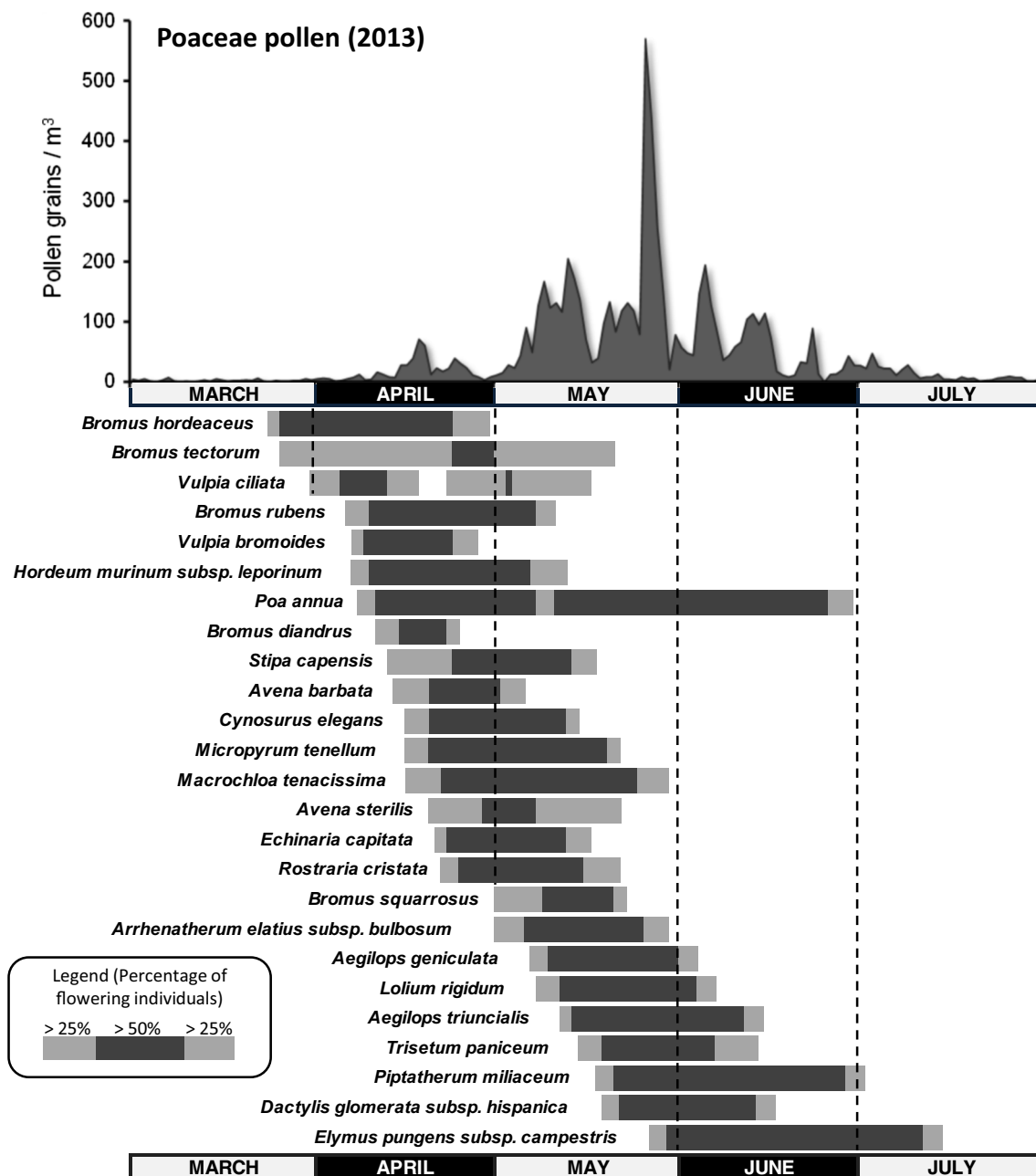


Fig. 4 Comparison between daily pollen concentrations and PFI for the given grass species during the year 2013

(Voukantsis et al. 2010), together with a considerable random component. In this study, airborne pollen time series were modeled using a seasonal-trend decomposition procedure based on LOESS smoothing, which proved well suited for this purpose. While the STL decomposition technique has recently been used by other authors to measure and analyze airborne pollen trends over fairly long periods (García-Mozo et al. 2014; Aguilera et al. 2015), the results obtained here confirm that decomposition of pollen data series is also of great value for interpreting variability in other components—seasonality and residuals—over a smaller timescale (Rojo et al. 2015).

This study represents a methodological advance for the modeling of aerobiological series, with various useful applications: for predicting daily pollen concentrations, for identifying major sources of pollen emission, and for analyzing variations in pollen concentrations as a function of environmental factors such as meteorological variables.

Using STL decomposition procedures, we can isolate seasonality and trend components from data series and also eliminate part of the random noise or residual component which cannot be accounted for by the behavior of the series itself (Brockwell and Davis 2002). In aerobiological data series,

Table 2 Results of Spearman correlation tests between phenological data (PFI) of the species and Poaceae pollen data from original data and the pollen time series (seasonal + trend)

Grass species	Pollen (original data)	Pollen (seasonal + trend)
<i>Aegilops geniculata</i>	0.67*	0.72*
<i>Aegilops triuncialis</i>	0.61*	0.86*
<i>Arrhenatherum elatius subsp. bulbosum</i>	0.51*	0.54*
<i>Dactylis hispanica</i>	0.60*	0.87*
<i>Lolium rigidum</i>	0.63*	0.70*
<i>Piptatherum miliaceum</i>	0.48*	0.80*
<i>Poa annua</i>	0.46*	0.45*
<i>Trisetum paniceum</i>	0.58*	0.88*

* $p < 0.001$

deterministic seasonal and trend components are associated with certain environmental features and biological aspects

from the plant life cycles, which vary as a function of time-scales and geographical factors (Petropavlovskikh et al.

Fig. 5 Results of the partial least square regressions developed for modeling the pollen data from the phenological variables (year 2013). Pollen data: the original pollen data (**a**) and the summations of the seasonal and trend components of the pollen time series (**b**). *MSE* mean square error, *RMSE* root-mean-square error. Sampled species: *AeGe* *Aegilops geniculata*, *AeTr* *Aegilops triuncialis*, *ArBu* *Arrhenatherum elatius subsp. bulbosum*, *DaHi* *Dactylis glomerata subsp. hispanica*, *LoRi* *Lolium rigidum*, *PiMi* *Piptatherum miliaceum*, *PoAn* *Poa annua*, *TrPa* *Trisetum paniceum*

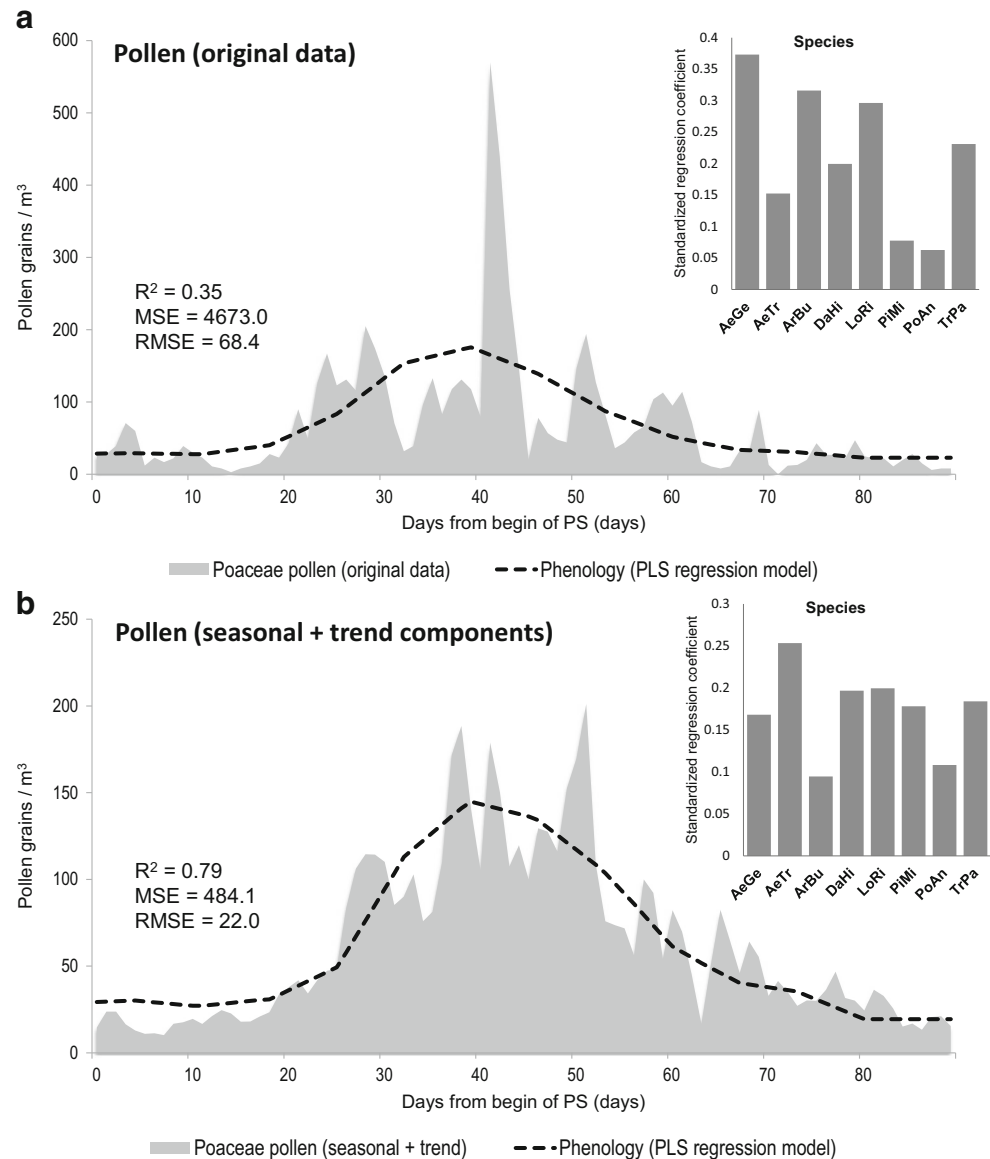
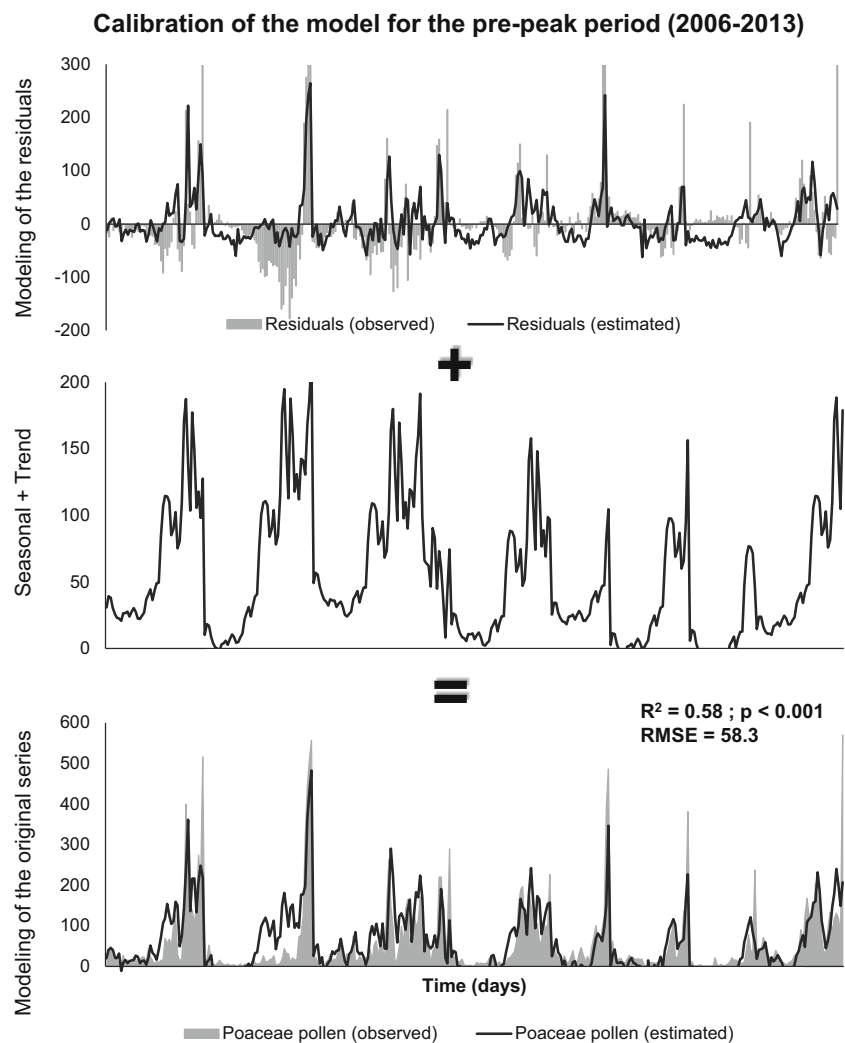


Fig. 6 Calibration of the model for the pre-peak period, obtained by adding the deterministic component (seasonal + trend) to the estimated residuals during the period 2006–2013

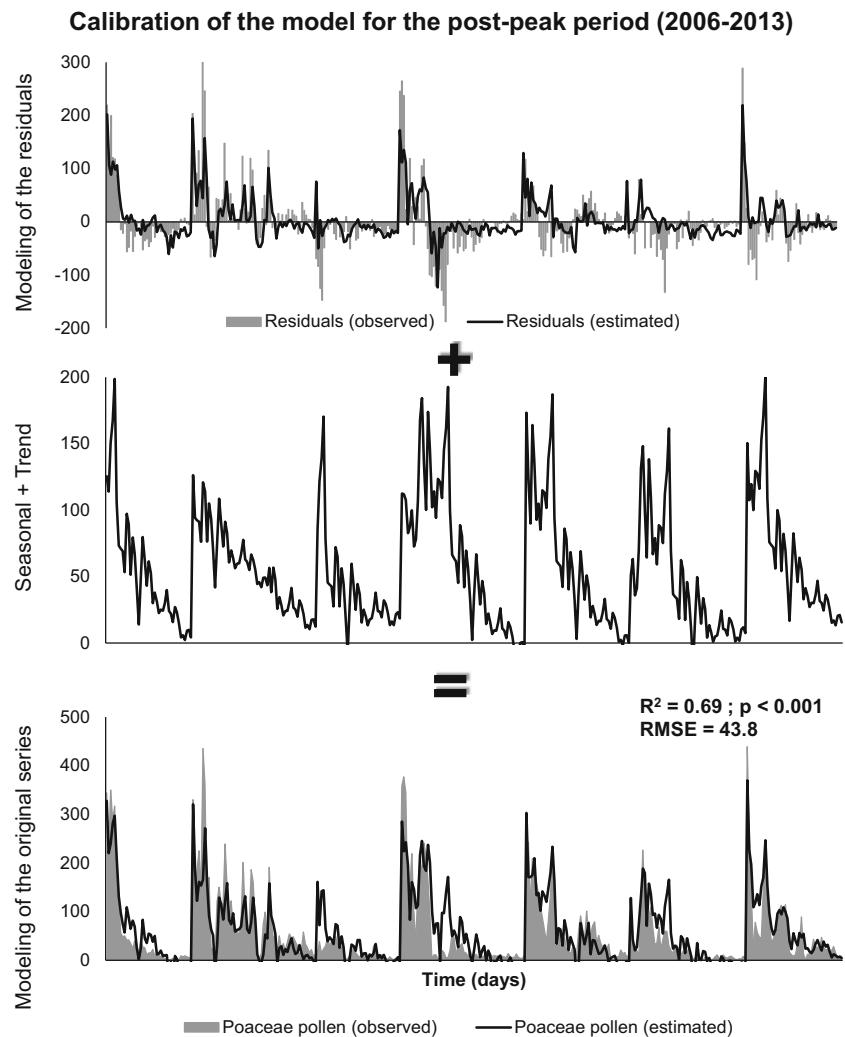


2015). The present study examined the variability of each component separately and then tested for correlations between these components and environmental (meteorological and phenological) parameters over the most appropriate timescale in each case; if data series are analyzed without decomposition, certain interesting features may remain concealed, due to the potential synergy of environmental parameters over different timescales. The present study focused on Poaceae pollen series, but the same method could be used to model the behavior of the pollen spectrum for other species.

Separate processing of each component in the pollen curve with a view to identifying the cause of variations should start by examining shifts in seasonality prompted by the seasonal climate dynamics regulating plant life cycles (Cleland et al. 2007). Analysis of the grass flowering period in the present study showed that the model generated using flowering phenology data for a very small number of species accounted sufficiently for series seasonality.

The key grass species for modeling pollen-curve seasonality proved to be later-flowering species including *Aegilops* spp., *D. glomerata* subsp. *hispanica*, *L. rigidum*, and *T. paniceum*. Similar findings have been reported by other authors in studies of the major contributors to the pollen curve elsewhere in the world (Frenguelli et al. 2010; León-Ruiz et al. 2011). Future research, however, should focus on other aspects of pollen emission by different species, including pollen production (Prieto-Baena et al. 2003; Aboulaich et al. 2009) and pollen grain dispersal capacity (Subiza 2003). In this sense, the major contributors to the Poaceae airborne pollen can be the species belonging to the genera *Trisetum* (*T. paniceum*), *Dactylis* (*D. glomerata*), and *Lolium* (*L. rigidum*). These species produce a huge number of pollen grains per inflorescence (5,000,000–21,000,00 pollen grains per inflorescence, Prieto-Baena et al. 2003). Otherwise, although the flowering of species of the genus *Aegilops* coincided with the highest levels of airborne pollen

Fig. 7 Calibration of the model for the post-peak period, obtained by adding the deterministic component (seasonal + trend) to the estimated residuals during the period 2006–2013



concentrations, these species can have low contribution to the airborne pollen concentrations because their pollen production is below 100,000 pollen grains per inflorescence (Prieto-Baena et al. 2003).

Since the data series used here did not display a significant trend pattern, i.e., it was stationary over time in probability terms (Cryer and Chan 2008), there was no need to analyze the trend component separately. In longer pollen data series,

Fig. 8 Independent variables taken into account when modeling residuals. Regression coefficients of the pollen concentrations for the previous days (*Poac*) and meteorological variables for the previous days: *Tmax* maximum temperature, *Tmin* minimum temperature, *R* rainfall. Data for the previous day (*x*-1), previous 2 days (*x*-2), and previous 3 days (*x*-3) were considered for modeling

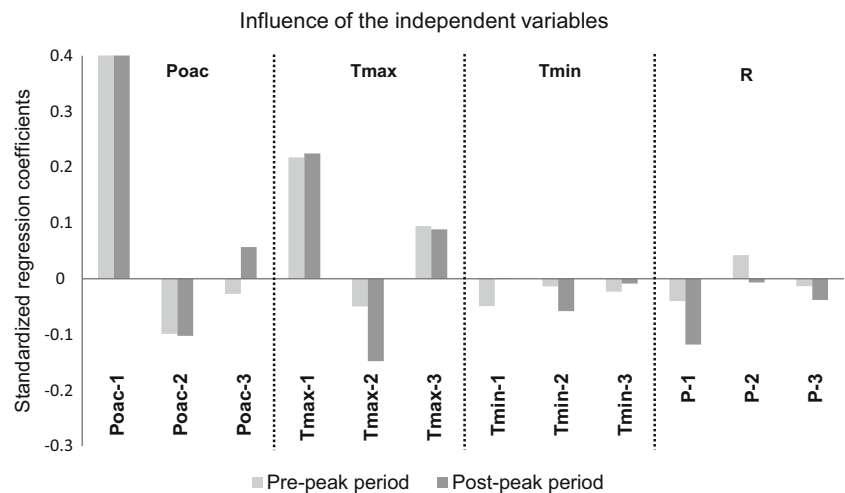
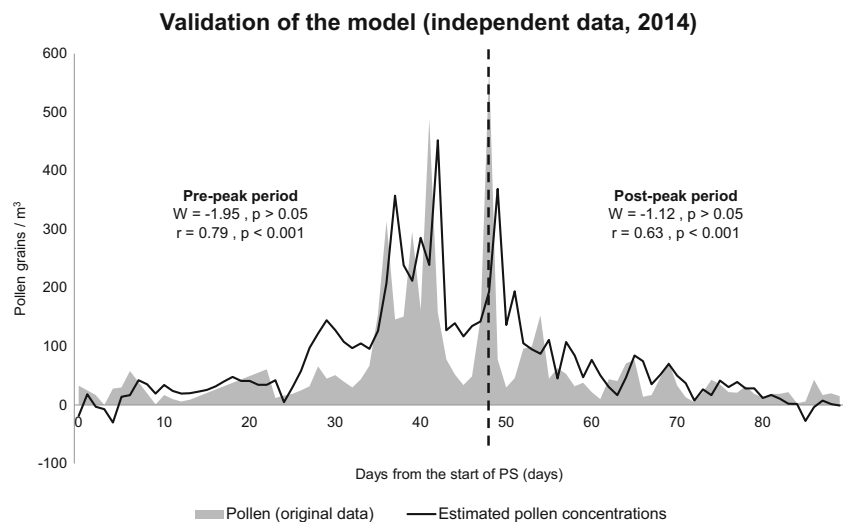


Fig. 9 Estimated Poaceae pollen concentrations obtained by adding the deterministic component (seasonal + trend) to the predicted residual for 2014. r statistics from the Spearman correlation test. W statistics from the Wilcoxon signed-rank test. Dashed line marks the peak pollen-concentration day



however, significant trends have been observed, prompted by variations in weather patterns (Recio et al. 2010; Makra et al. 2011). Reported trends are mostly attributed to changes in plant phenology as a result of environmental changes (Peñuelas et al. 2002; Gordo and Sanz 2009; García-Mozo et al. 2010). Wherever apparent trends are observed, they should be analyzed on a long-term basis by isolating the trend component of the data series (García-Mozo et al. 2010; Aguilera et al. 2015).

Finally, the daily pollen series contained a residual component, reflecting its stochastic nature. This residual displays highly variable behavior due to the influence of short-term, local-scale meteorological conditions (Petropavlovskikh et al. 2015), which, in the Mediterranean region, tend to vary considerably (García-Mozo et al. 2014). On the basis of correlations with a range of meteorological variables (Ranzi et al. 2003), the residual component can be modeled—as here—with a view to identifying the meteorological variables most influencing airborne pollen concentrations (Rojo et al. 2015) and also to predicting future pollen concentrations by estimating the residual component isolated separately from the original data series (Valderrama et al. 2010).

The patterns of Poaceae pollen emission can vary greatly in response to meteorological changes, as it is shown in the results of this study. Therefore, the aerobiological processes are greatly influenced by the meteorological variables (Sánchez-Mesa et al. 2005), and maximum temperature is the most important meteorological variable included in the predictive model with respect to the standardized regression coefficients. However, other important phenomenon such as pollen dispersal is conditioned by other meteorological variables such as rainfall (Rojo et al. 2015), and thus, this variable has been also included in the model of our study.

The model obtained using the STL decomposition procedure and data for 2006–2013 was used to predict airborne

Poaceae pollen concentrations for 2014. The predicted pollen curve correlated well with the observed curve, with no significant difference between distributions. The prediction model for the residuals included meteorological variables and pollen for previous days, focusing on those accounting for the greatest amount of variance (Sawa 1978; Seasholtz and Kowalski 1993). The parsimony principle should always be applied with a view to obtaining models combining simplicity with a high degree of precision (Akaike 1974). Moreover, the calibration of pre-peak and post-peak models with determination coefficients of 0.58 and 0.69, respectively, may be considered quite acceptable (Silva-Palacios et al. 2015) and yielded predictions which correlated closely with observed values for 2014.

The inclusion of a large number of highly correlated predictors may give rise to multicollinearity issues (Graham 2003). Statistical modeling of various aspects of the pollen season often uses numerous pollen-related predictors (Stach et al. 2008; Veriankaitè et al. 2011). PLSR is regarded as a suitable technique for reducing the number of predictors, since it is not sensitive to multicollinearity (Wold et al. 1984); it has recently been used in aerobiological modeling (Oteros et al. 2013a; Brighetti et al. 2014; Rojo et al. 2015). Application of this procedure for the analysis of stochastic processes yields good results, since it reduces other statistical problems linked to model stability and quality of fit (Preda and Saporta 2005).

Finally, the PS start date needs to be carefully defined, since it is a key aspect of modeling. Poaceae have a very long pollen season (Pérez-Badia et al. 2010; Gücel et al. 2013), including numerous days at the start and end of the main season with very low and intermittently varying pollen concentrations. These hinder modeling and, thus, the prediction of future airborne grass pollen concentrations. For a better fit, the pollen season should be reduced to the period over which pollen concentrations are reasonably high (Silva-Palacios

et al. 2015). Moreover, models for predicting daily pollen concentrations—such as that constructed here—can be combined with models generated for predicting the PS start date (Kasprzyk and Walanus 2010, 2014; Pauling et al. 2014).

Conclusions

Decomposition of time series using the STL procedure was found to be a suitable technique for analyzing the phenological and meteorological factors influencing airborne pollen concentrations. Separate analysis of each of the components of the pollen data series enables the sources of variability to be identified more accurately than by analysis of the original non-decomposed data series, since certain environmental factors may mask the influence of other factors operating at different timescales. Furthermore, the results of this work constitute an advance in the knowledge of the major contributors (grass species) to the Poaceae airborne pollen concentrations. The method can also be applied for modeling other pollen types.

References

- Aboulaich N, Achmakh L, Bouziane H, Trigo MM, Recio M, Kadiri M, Cabezedo B, Riadi H, Kazzaz M (2013) Effect of meteorological parameters on Poaceae pollen in the atmosphere of Tetouan (NW Morocco). *Int J Biometeorol* 57:197–205
- Aboulaich N, Bouziane H, Kadiri M, Trigo MM, Riadi H, Kazzaz M, Merzouki A (2009) Pollen production in anemophilous species of the Poaceae family in Tetouan (NW Morocco). *Aerobiologia* 25:27–38
- Aguilera F, Orlandi F, Ruiz-Valenzuela L, Msallem M, Fornaciari M (2015) Analysis and interpretation of long temporal trends in cumulative temperatures and olive reproductive features using a seasonal trend decomposition procedure. *Agric For Meteorol* 203:208–216
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):716–723
- Aznarte JL, Benítez JM, Nieto D, de Linares C, Díaz de la Guardia C, Alba F (2007) Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. *Expert Syst Appl* 32:1218–1225
- Belmonte J, Canela M (2002) Modelling aerobiological time series. Application to Urticaceae. *Aerobiologia* 18:287–295
- Bourjea J, Dalleau M, Derville S, Beudard F, Marmoeux C, M'Soili A, Roos D, Ciccione S, Frazier J (2015) Seasonality, abundance, and fifteen-year trend in green turtle nesting activity at Itsamia, Moheli, Comoros. *Endang Species Res* 27:265–276
- Bousquet J, Anto J, Auffray C, Akdis M, Cambon-Thomsen A, Keil T, Haahela T, Lambrecht BN, Postma DS, Sunyer J, Valenta R, Akdis CA, Annesi-Maesano I, Arno A, Bachert C, Ballester F, Basagana X, Baumgartner U, Bindslev-Jensen C, Brunekreef B, Carlsen KH, Chatzi L, Cramer R, Eveno E, Forastiere F, Garcia-Aymerich J, Guerra S, Hammad H, Heinrich J, Hirsch D, Jacquemin B, Kauffmann F, Kerkhof M, Kogevinas M, Koppelman GH, Kowalski ML, Lau S, Lodrup-Carlsen KC, Lopez-Botet M, Lotvall J, Lupinek C, Maier D, Makela MJ, Martinez FD, Mestres J, Momas I, Nawijn MC, Neubauer A, Oddie S, Palkonen S, Pin I, Pison C, Rancé F, Reitamo S, Rial-Sebbag E, Salapatas M, Siroux V, Smaghe D, Torrent M, Toskala E, van Cauwenberge P, van Oosterhout AJ, Varraso R, von Hertzen L, Wickman M, Wijmenga C, Worm M, Wright J, Zuberbier T (2011) MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. *Allergy* 66(5):596–604
- Brighetti MA, Costa C, Menesatti P, Antonucci F, Tripodi S, Travaglini A (2014) Multivariate statistical forecasting modeling to predict Poaceae pollen critical concentrations by meteorological data. *Aerobiologia* 30:25–33
- Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting, 2nd edn. Springer-Verlag, New York, USA
- Cleland EE, Chuine I, Menzel A, Mooney HA, Schwartz MD (2007) Shifting plant phenology in response to global change. *Trends Ecol Evol* 22(7):357–365
- Cleveland RB, Cleveland WS, McRae JE, Terpenning I (1990) STL: a seasonal-trend decomposition procedure based on loess. *J Off Stat* 6(1):3–33
- Cryer JD, Chan KS (2008) Time series analysis with applications in R, 2nd edn. Springer, New York, USA
- Currie KI, Brailsford G, Nichol S, Gomez A, Sparks R, Lassey KR, Riedel K (2011) Tropospheric $^{14}\text{CO}_2$ at Wellington, New Zealand: the world's longest record. *Biogeochemistry* 104:5–22
- D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, Liccardi G, Popov T, van Cauwenberge P (2007) Allergenic pollen and pollen allergy in Europe. *Allergy* 62:976–990
- Dagum EB, Luati A (2003) Global and local statistical properties of fixed-length nonparametric smoothers. *Stat Method Appl* 11: 313–333
- Estrella N, Menzel A, Krämer U, Behrendt H (2006) Integration of flowering dates in phenology and pollen counts in aerobiology: analysis of their spatial and temporal coherence in Germany (1992–1999). *Int J Biometeorol* 51:49–59
- Fernández-Llamazares A, Belmonte J, Boada M, Fraixedas S (2014) Airborne pollen records and their potential applications to the conservation of biodiversity. *Aerobiologia* 30:111–122
- Fernández-Rodríguez S, Adams-Groom B, Silva-Palacios I, Caeiro E, Brandao R, Ferro R, Gonzalo-Garijo A, Smith M, Tormo R (2015) Comparison of Poaceae pollen counts recorded at sites in Portugal, Spain and the UK. *Aerobiologia* 31:1–10
- Frenguelli G, Passalacqua G, Bonini S, Fiocchi A, Incorvaia C, Marcucci F, Tedeschini E, Canonica GW, Frati F (2010) Bridging allergologic and botanical knowledge in seasonal allergy: a role for phenology. *Ann Allerg Asthma Im* 105(3):223–227
- Galán C, Cariñanos P, Alcázar P, Domínguez-Vilches E (2007) Spanish aerobiology network (REA): management and quality manual. Servicio de Publicaciones, Universidad de Córdoba, Córdoba, Spain
- Galán C, Cariñanos P, García-Mozo H, Alcázar P, Domínguez-Vilches E (2001) Model for forecasting *Olea europaea* L. airborne pollen in South-West Andalusia, Spain. *Int J Biometeorol* 45:59–63
- García-Mozo H (2011) The use of aerobiological data on agronomical studies. *Ann Agric Environ Med* 18:159–164
- García-Mozo H, Mestre A, Galán C (2010) Phenological trends in southern Spain: a response to climate change. *Agric For Meteorol* 150: 575–580
- García-Mozo H, Oteros JA, Galán C (2016) Impact of land cover changes and climate on the main airborne pollen types in Southern Spain. *Sci Total Environ* 548–549:221–228
- 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
- Gordo O, Sanz JJ (2009) Long-term temporal changes of plant phenology in the Western Mediterranean. *Global Change Biol* 15:1930–1948
- Graham MH (2003) Confronting multicollinearity in ecological multiple regression. *Ecology* 84:2809–2815

- Gucel S, Guvensen A, Ozturk M, Celik A (2013) Analysis of airborne pollen fall in Nicosia (Cyprus). *Environ Monit Assess* 185:157–169
- Hamid N, Ali SM, Talib F, Sadiq I, Ghufra MA (2015) Spatial and temporal variations of pollen concentrations in Islamabad (Pakistan): effect of meteorological parameters and impact on human health. *Grana* 54(1):53–67
- Harvey AC, Peters S (1990) Estimation procedures for structural time series model. *J Forecast* 9:89–108
- Hirst JM (1952) An automatic volumetric spore trap. *Ann Appl Biol* 36:257–344
- Jato V, Rodríguez-Rajo FJ, Alcázar P, De Nuntiis P, Galán C, Mandrioli P (2006) May the definition of pollen season influence aerobiological results? *Aerobiologia* 22:13–25
- Kasprzyk I (2006) Comparative study of seasonal and intradialurnal variation of airborne herbaceous pollen in urban and rural areas. *Aerobiologia* 22:185–195
- Kasprzyk I, Walanus A (2010) Description of the main Poaceae pollen season using bi-Gaussian curves, and forecasting methods for the start and peak dates for this type of season in Rzeszów and Ostrowiec Św. (SE Poland). *J Environ Monit* 12:906–916
- Kasprzyk I, Walanus A (2014) Gamma, Gaussian and logistic distribution models for airborne pollen grains and fungal spore season dynamics. *Aerobiologia* 30:369–383
- León-Ruiz E, Alcázar P, Domínguez-Vilches E, Galán C (2011) Study of Poaceae phenology in a Mediterranean climate. Which species contribute most to airborne pollen counts? *Aerobiologia* 27:37–50
- Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. *Biometrika* 65:297–303
- Mabberley DJ (1987) *The plant book*. Cambridge University press, Cambridge
- Makra L, Matyasovszky I, Deák AJ (2011) Trends in the characteristics of allergenic pollen circulation in central Europe based on the example of Szeged, Hungary. *Atmos Environ* 45:6010–6018
- Meier U (2001) Growth stages of mono- and dicotyledonous plants. BBCH monograph. 2nd ed. Federal Biological Research Centre for Agriculture and Forestry
- Mevik BH, Wehrens R (2007) The pls package: principal component and partial least squares regressions in R. *J Stat Softw* 18(2):1–24
- Moseholm L, Weeke ER, Petersen BN (1987) Forecast of pollen concentrations of Poaceae (grasses) in the air by time series analysis. *Pollen Spores* 2-3:305–322
- Ocaña-Peinado F, Valderrama MJ, Aguilera AM (2008) A dynamic regression model for air pollen concentration. *Stoch Environ Res Risk Assess* 22(Suppl 1):S59–S63
- Oteros J, García-Mozo H, Hervás-Martínez C, Galán C (2013a) Year clustering analysis for modelling olive flowering phenology. *Int J Biometeorol* 57(4):545–555
- Oteros J, García-Mozo H, Vázquez L, Mestre A, Domínguez-Vilches E, Galán C (2013b) Modelling olive phenological response to weather and topography. *Agric Ecosyst Environ* 179:62–68
- Pauling A, Gehrig R, Clot B (2014) Toward optimized temperature sum parameterizations for forecasting the start of the pollen season. *Aerobiologia* 30:45–57
- Peel RG, Ørby PV, Skjøth CA, Kennedy R, Schlünsen V, Smith M, Sommer J, Hertel O (2014) Seasonal variation in diurnal atmospheric grass pollen concentration profiles. *Biogeosciences* 11:821–832
- Peñuelas J, Filella I, Comas P (2002) Changed plant and animal life cycles from 1952 to 2000 in the Mediterranean region. *Global Change Biol* 8(6):531–544
- Pérez-Badía R, Bouso V, Rojo J, Vaquero C, Sabariego S (2013) Dynamics and behaviour of airborne *Quercus* pollen in central Iberian Peninsula. *Aerobiologia* 29:419–428
- Pérez-Badía R, Rapp A, Morales C, Sardinero S, Galán C, García-Mozo H (2010) Pollen spectrum and risk of pollen allergy in central Spain. *Ann Agric Environ Med* 17:139–151
- Petrovskikh I, Evans R, McConville G, Manney GL, Rieder HE (2015) The influence of the North Atlantic Oscillation and El Niño–Southern Oscillation on mean and extreme values of column ozone over the United States. *Atmos Chem Phys* 15:1585–1598
- Preda C, Saporta G (2005) PLS regression on a stochastic process. *Comput Stat Data An* 48:149–158
- Prieto-Baena JC, Hidalgo PJ, Domínguez-Vilches E, Galán C (2003) Pollen production in the Poaceae family. *Grana* 42:153–160
- R Core Team (2015) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (accessed 11 Sept 2015)
- Ranzi A, Lauriola P, Marletto V, Zinoni F (2003) Forecasting airborne pollen concentrations: development of local models. *Aerobiologia* 19:39–45
- Recio M, Docampo S, García-Sánchez J, Trigo MM, Melgar M, Cabezero B (2010) Influence of temperature, rainfall and wind trends on grass pollination in Malaga (western Mediterranean coast). *Agric For Meteorol* 150(7-8):931–940
- Reyment RA, Jvreskog KG (1996) *Applied factor analysis in the natural sciences*. Cambridge University Press
- Rodríguez-Rajo FJ, Jato V, Aira MJ (2003) Pollen content in the atmosphere of Lugo (NW Spain) with reference to meteorological factors (1999–2001). *Aerobiologia* 19:213–225
- Rodríguez-Rajo F, Valencia-Barrera RM, Vega-Maray AM, Suárez FJ, Fernández-González D, Jato V (2006) Prediction of airborne *Alnus* pollen concentration by using ARIMA models. *Ann Agric Environ Med* 13:25–32
- Rojo J, Pérez-Badía R (2015) Models for forecasting the flowering of Cornicabra olive groves. *Int J Biometeorol* 59:1547–1556
- Rojo J, Rapp A, Lara B, Fernández-González F, Pérez-Badía R (2015) Effect of land uses and wind direction on the contribution of local sources to airborne pollen. *Sci Total Environ* 538:672–682
- Rummel RJ (1988) *Applied factor analysis*. Northwestern University Press, Evanston
- Sabariego S, Cuesta P, Fernández-González F, Pérez-Badía R (2012) Models for forecasting airborne Cupressaceae pollen levels in central Spain. *Int J Biometeorol* 56(2):253–258
- Sánchez-Mesa JA, Galán C, Martínez-Heras JA, Hervás-Martínez C (2005) The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with typical Mediterranean climate. *Int J Biometeorol* 49:355–362
- Sánchez-Mesa JA, Galán C, Martínez-Heras JA, Hervás-Martínez C (2002) The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. *Clin Exp Allergy* 32:1606–1612
- Sawa T (1978) Criteria for discriminating among alternative regression models. *Econometrica* 46(6):1273–1291
- Seasholtz MB, Kowalski B (1993) The parsimony principle applied to multivariate calibration. *Anal Chim Acta* 277:165–177
- Silva-Palacios I, Fernández-Rodríguez S, Durán-Barroso P, Tormo-Molina R, Maya-Manzano JM, Gonzalo-Garijo A (2015) Temporal modelling and forecasting of the airborne pollen of Cupressaceae on the southwestern Iberian Peninsula. *Int J Biometeorol*. doi:10.1007/s00484-015-1026-6
- Stach A, Smith M, Prieto-Baena JC, Emberlin J (2008) Long-term and short-term forecast models for Poaceae (grass) pollen in Poznań, Poland, constructed using regression analysis. *Environ Exp Bot* 62:323–332
- Subiza J (2003) Gramíneas: Aerobiología y polinosis en España. *Alergol Immunol Clin* 18(3):7–11
- Valderrama MJ, Ocaña FA, Aguilera AM, Ocaña-Peinado FM (2010) Forecasting pollen concentration by a two-step functional model. *Biometrics* 66:578–585
- Vallejos RO, Fabré NN, Batista VS, Acosta J (2013) The application of a general time series model to floodplain fisheries in the Amazon. *Environ Modell Softw* 48:202–212

- Veriankaitė L, Šaulienė I, Bukantis A (2011) Evaluation of meteorological parameters influence upon pollen spread in the atmosphere. *J Environ Eng Landsc* 19(1):5–11
- Voukantis D, Niska H, Karatzas K, Riga M, Damialis A, Vokou D (2010) Forecasting daily pollen concentrations using data-driven modeling methods in Thessaloniki, Greece. *Atmos Environ* 44: 5101–5111
- Winters PR (1960) Forecasting sales by exponentially weighted moving averages. *Manage Sci* 6:324–342
- Wold S, Ruhe A, Wold H, Dunn WJ (1984) The collinearity problem in linear regression. The Partial Least Squares (PLS) approach to generalized inverses. *SIAM J Sci Comput* 5(3):735–743
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab* 58:109–130
- Zhou J, Liang Z, Liu Y, Guo H, He D, Zhao L (2015) Six-decade temporal change and seasonal decomposition of climate variables in Lake Dianchi watershed (China): stable trend or abrupt shift? *Theor Appl Climatol* 119:181–191