Environ Monit Assess (2026) 198:42
https://doi.org/10.1007/s10661-025-14871-0

RESEARCH

Poisson, negative binomial, and zero-inflated negative binomial regression models for predicting daily airborne pollen concentration levels in Sinop (Türkiye)

Ayten Yiğiter · Cemile Cansı Demir · Canan Hamurkaroğlu · Hülya Özler · Ayşe Kaplan · Nazan Danacıoğlu · Sümeyra Sezer Kaplan

Received: 24 April 2025 / Accepted: 29 November 2025 / Published online: 12 December 2025

Abstract Pollen, produced during the flowering period of plants, especially anemogamous plants that produce high volumes of pollen, poses a risk to individuals with pollen allergies when it is present in the atmosphere. Meteorological factors are known to affect the duration, distribution, and amount of pollen in the air. The remarkable increase in allergic cases in recent years has led to many studies investigating the relationship between pollen and spores that cause allergies and meteorological factors in Türkiye as well as in the world. In this study, meteorological factors and their influence on pollen concentrations in the air were examined for the Sinop region in northern Türkiye. First, descriptive statistics for pollen obtained from plant taxa were obtained and interpreted. Precipitation, humidity, temperature, and wind speed were considered as meteorological parameters, and the effects of these variables on pollen counts and their annual changes were modelled using Poisson, negative binomial, and zero-inflated negative binomial (ZINB) regression models. The estimation results for all pollen taxa were then discussed. In the models obtained for each pollen type, the statistical significance of the independent variables such as temperature, precipitation,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10661-025-14871-0.

Present Address:
A. Yiğiter
Assosiate Professor, Department of Statistics, Faculty of Science, Hacettepe University, Ankara, Türkiye
e-mail: yigiter@hacettepe.edu.tr

C. C. Demir
Biology Teacher, Science High School, Sinop, Türkiye
e-mail: cemiledemir50@gmail.com

C. Hamurkaroğlu
Professor, Department of Actuarial Sciences, Faculty of Business, Karabuk University, Karabük, Türkiye
e-mail: cananhamurkaroglu@karabuk.edu.tr

H. Özler
Professor, Department of Biology, Faculty of Arts and Sciences, Sinop University, Sinop, Türkiye
e-mail: hozler@sinop.edu.tr

A. Kaplan
Assosiate Professor, Department of Biology, Faculty of Arts and Sciences, Bülent Ecevit University, Zonguldak, Türkiye
e-mail: ayse.kaplan@beun.edu.tr

N. Danacıoğlu
Assistant Professor, Department of Statistics, Faculty of Arts and Sciences, Sinop University, Sinop, Türkiye
e-mail: nazand@sinop.edu.tr

S. S. Kaplan (✉)
Research Assistant, Department of Actuarial Sciences, Faculty of Business, Karabük University, Karabük, Türkiye
e-mail: sskaplan@karabuk.edu.tr

relative humidity, wind speed, time, and lag 1 was found to be different according to the pollen type.

Keywords Poisson regression · Negative binomial regression · Zero-inflated negative binomial regression · Pollen · Meteorological conditions · Sinop (Türkiye)

Introduction

Pollens are the most important aeroallergens found in the atmosphere. The magnitude of allergenic effects depends on the allergens contained in the pollen. Climate change, one of the most important problems of today, affects the flowering periods of plants, allowing pollen to remain in the air for a longer time. As a result, it may lead to higher pollen production and increased allergic diseases (Bielory et al., 2012; Damialis et al., 2019). In this context, investigating the effects of climate change on pollen and performing statistical modeling has become important (Picornell et al., 2024; Janati et al., 2017; Zhang et al., 2013; 2015a, 2015b; Zhang & Steiner, 2022; Adams–Groom et al., 2022). The number of people susceptible to pollen allergy is expected to more than double by 2060 due to climate change. (O'Connor et al., 2017).

Aeropalynological studies are scientific analyses of pollen in the atmosphere of a region, encompassing seasonal variations in pollen and the identification of the respective botanical sources. Pollen calendars based on these data have been developed to raise awareness among individuals with pollen sensitivities living in these regions. These calendars also assist physicians in the diagnosis and treatment of allergic diseases. Many studies from Türkiye and other countries have focused on understanding regional pollen diversity, preparing pollen calendars, and exploring how weather conditions influence pollen concentrations in the air. The first aeropalynological study in Türkiye was carried out by Özkaragöz and Karamaoğlu (1967) in Ankara area and Aytuğ (1973) around İstanbul. Later, studies were carried out in many areas on the detection of allergenic pollens and the creation of a pollen calendar (Canşı Demir et al., 2021; Akpınar&Altunoğlu, 2024; Akdoğan Karadağ &Altunoğlu, 2025).

Numerous aerobiological investigations aim to assess the diversity of atmospheric pollen within a given area, establish a pollen calendar, and model the abundance of different plant pollens by utilizing meteorological data along with mathematical and statistical approaches. These studies, which cover various pollen taxa in diverse geographical locations worldwide, typically begin with straightforward linear regression models and progress to more intricate general linear models. Some of the authors authors who have implemented linear models in their research include Piotrowska-Weryszko (2013), Bonini. et al. (2015), García-Mozo et al. (2015), Sabariego et al. (2011), Ocaña-Peinado et al. (2013), Oteros et al. (2013), De Weger et al. (2014), Ritenberga et al. (2016), Sicard et al. (2012), Aboulaich et al. (2013), and Myszkowska (2013).

Similarly, Malkiewicz et al. (2014), Brighetti et al. (2014), Viney et al. (2021) and Aguilera et al. (2015), Canşı Demir et al. (2021), Lara et al. (2019), Vélez-Pereira et al. (2019) examined different pollens in different geographical regions using regression models and correlation analysis.

There are the following studies that utilize time series analysis to model the potential impacts of temporal factors on extensive datasets related to plant phenology. Puc and Wolski (2013) estimated pollen season probabilities for different pollens over a 13-year period (2000–2012 included. Garcia-Mozo et al. (2014) used time series methods to study how the timing of the pollen season varied from 1982 to 2011. Siniscalco et al. (2015) used a 26-year aerobiological dataset. Tassan-Mazzocco et al. (2015) made a modelling based on the time series of 2005–2010 for the pollen they were interested in. Nowosad et al. (2016), designed models to estimate atmospheric pollen levels.

Silva-Palacios et al. (2016) analyzed 21 years of pollen data using time series analysis. Similarly, Fernández-Rodríguez et al. (2016a, 2016b, & 2016c) modelled pollen concentrations using time series analysis. Additionally, Rojo et al. (2017) analyzed daily pollen concentration data for the years 2006 to 2014 through time series modelling.

In recent years, in addition to regression and time series analyses, numerous studies have been conducted utilizing machine learning (deep learning) and artificial neural networks (ANN) to ascertain the spatial distribution and concentration of pollen. Rodríguez-Rajo et al. (2010), Liu et al. (2017), Cordero et al. (2020), Lo et al. (2021) and Picornell et al.

(2024) developed computational models utilizing various machine learning algorithms. Valencia et al. (2019) applied ANN to forecast pollen concentrations using a comprehensive 19-year dataset (1999–2017). Klimczak et al. (2020) employed regression techniques and decision tree algorithms to classify airborne pollen samples collected during 2016 and 2017. Coric et al. (2023) used ANN models to estimate daily pollen levels, while Shokouhi et al. (2024) proposed a spatiotemporal machine learning model aimed at predicting daily pollen concentrations with enhanced accuracy.

Although Canşı Demir et al. (2021) investigated the relationship between weather conditions and pollen through descriptive statistics, our study builds on their findings by specifically modeling the relationship between pollen concentrations and meteorological factors in Sinop province.

Materials and methods

Aeropalynological data considered in this study were obtained from Canşı Demir et al. (2021). Their study involved a two-year pollen monitoring period, conducted from January 2016 to December 2017, in the central area of Sinop province (see Fig. 1) and pollen data were measured as the daily average amount of grains per m^3 of air. For air sampling and pollen analysis, Canşı Demir et al. (2021) used the method described by the Spanish Aerobiology Network (REA: Red Española de Aerobiología).

The coexistence of an oceanic climate and a Mediterranean microclimate in Sinop province enhances its floristic richness. Vegetation studies have identified various types, including forests, degraded forests, maquis, phrygana, and dune vegetation within the Sinop peninsula. Approximately 60% of the province is forested (Karaer & Kılınç, 1993; Kılınç & Karaer, 1995; Özen & Kılınç, 1995). Sinop province has rich biodiversity with a humid and warm climate. Therefore, Sinop province contains many allergenic pollen-producing plant flora. Twelve very important taxa (*Alnus, Ambrosia, Artemisia, Betula,* Chenopodiaceae (Amaranthaceae), *Corylus, Cupresssaceae, Olea, Platanus,* Poaceae, *Quercus*, Urticaceae (*Urtica, Parietaria*) were determined by COST (Europaean Cooperation in Science and Technology) Action EO603. The pollen taxa considered in this study for Sinop province are the same as those described by Canşı Demir et al. (2021). The atmospheric frequency of these pollen taxa was used as the dependent variable, while temperature, precipitation, humidity, wind speed, and time (year) were utilized as independent variables. Daily pollen counts for the 15 most allergenic pollen taxa in Sinop province were modeled using Poisson, negative binomial, and zero-inflated negative binomial

Fig. 1 Map showing Sinop, where pollen data were collected using Burkard traps

(ZINB) regression, with meteorological variables as predictors. The dispersion test proposed by Cameron and Trivedi (2013) was applied to identify overdispersion. Additionally, the dispersion parameter (α) was estimated, and since overdispersion was observed, the analysis proceeded with the negative binomial regression model. Normality was tested for the negative binomial residuals using the Kolmogorov–Smirnov (K–S) test.

Subsequently, the validity of all models was assessed, leading to the creation of Autocorrelation Function (ACF) graphs. Based on the evaluation of these graphs, a lag variable (first-order lag: lag 1) was added to the models. With the inclusion of the lag 1 variable, Q–Q plots were generated to evaluate the normality of the residuals, and their autocorrelations were examined through the Ljung-Box test (Box & Pierce, 1970; Ljung & Box, 1978).

The estimation results were discussed, and the best-fit model was determined using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

In this study, all calculations were made using the R program version 4.5.0, the packages used in computation and visualization were: readxl (Wickham & Bryan, 2023), MASS (Venables & Ripley, 2002), ggplot2 (Wickham, 2016), dplyr (Wickham et al., 2019), patchwork (Pedersen, 2024), glmmTMB (Brooks et al., 2017), DHARMa (Hartig, 2024), and AER (Kleiber & Zeileis, 2008).

Statistical analysis: regression model, Poisson regression, negative binomial regression, zero-inflated negative binomial regression, and model comparison criteria

Regression model

Researchers in many disciplines commonly use regression modelling to examine independent variables that are thought to explain a response (dependent) variable. It expresses the nature of the relationship between these variables using a mathematical equation and, if the model is valid, allows for reliable and consistent predictions for the future. In the regression model, it is assumed that the response variable is continuous with "constant variance and suitable for normal distribution". If the response variable does not meet these assumptions, an alternative approach is the generalized linear model (GLM). GLMs integrate both linear and nonlinear regression techniques, with the response variable following a distribution from the exponential family, which encompasses distributions like normal, Poisson, binomial, exponential, and gamma.

Poisson and negative binomial regression models, both of which are part of the GLM framework, are commonly applied when the response variable is count data and discrete. The Poisson model is used when the mean and variance of the response variable are equal. If the variance exceeds the mean, in which this case is known as overdispersion, the negative binomial regression model is preferred (Agresti, 2015). Because real-world count data often exhibit a variance greater than the mean, a negative binomial regression model is commonly used. Additionally, the ZINB regression model, which has been used in recent years to analyze data where the dominance of pollen counts is zero, was also considered.

Poisson regression model

When the response variable consists of count data, a suitable probability distribution to model this type of data is the Poisson distribution, which is given below for $\mu > 0$:

$$f(y) = \frac{e^{-\mu}\mu^{y}}{y!},\ y = 0,1,2,\ldots \tag{1}$$

The mean and variance of the Poisson distribution are equal, that is, $E(Y) = V(Y) = \mu$. In this case, the Poisson regression model can be expressed as $y_i = E(y_i) + \varepsilon_i i = 1,2,\ldots,n$. The i^{th} row of the $n \times (p+1)$ dimensional matrix, X containing the explanatory variables has the form$x_i{\prime} = (1,\ x_{i1}',\ x_{i2}\prime,,\ldots,\ x_{ip}',)$, the $1 \times (p+1)$ dimensional vector of unknown regression parameters corresponding to the explanatory variables is of the form $\beta' = (\beta_0,\ \beta_1,\ldots,\ \beta_p)$, The link function between $E(Y_i) = \mu_i$ and explanatory variables is of the form $g(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p = \mathbf{x}_i'\boldsymbol{\beta}$. In this case, the relationship between the mean of the response variable and the linear response function can be described by an identical link function $\mu_i = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta})$. Therefore, $E(Y_i) = \mu_i = \mathbf{x}_i'\boldsymbol{\beta}$.

Another link function frequently used for the Poisson distribution is called the log link function, defined as $g(\mu_i) = ln(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$. For the log-link function, the relationship between the mean of the response variable and the linear response function η is expressed as $E(Y_i) = \mu_i = g^{-1}(\mathbf{x}_i\prime\boldsymbol{\beta}) = e^{\mathbf{x}_i\prime\boldsymbol{\beta}}$.

In this regression model, the maximum likelihood estimation (MLE) method is used to estimate the β parameters vector, and the likelihood function for n random samples is,

$$L(\boldsymbol{y}, \boldsymbol{\beta}) = \prod_{i=1}^{n} f_i(y_i) = \frac{\prod_{i=1}^{n} \mu_i^{y_i} exp\left(-\sum_{i=1}^{n} \mu_i\right)}{\prod_{i=1}^{n} y_i!}$$

ln likelihood function,

$$lnL(\boldsymbol{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i ln(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} ln(y_i!)$$

is of the form.

Parameter estimates were obtained iteratively using the weighted least squares method (WLSM). For the identical link function, the estimated Poisson regression equation is $\hat{y}_i = g^{-1}\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\right) = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$ and for the log link function, the estimated Poisson regression equation is $\hat{y}_i = g^{-1}\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\right) = e^{\boldsymbol{x}_i\prime\hat{\boldsymbol{\beta}}}$ (Hilbe, 2011).

Negative binomial regression model

The unconditional distribution of Y is defined in the form of a negative binomial distribution,

$$f(y) = \frac{\Gamma(\frac{1}{\alpha}+y)}{\Gamma(\frac{1}{\alpha})\Gamma(y+1)}\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y},\ y = 0,1,2,\ldots\ \mu > 0, \alpha > 0 \tag{2}$$

In Eq. (2), α is the dispersion parameter, the mean and variance of the distribution are $E(Y) = \mu$ and $V(Y) = \mu + \frac{\mu^2}{\alpha}$, respectively. In a negative binomial regression model, the relationship between the mean of Y and the linear response function is usually given by the log link function $E(y_i) = \mu_i = e^{x_i\prime\beta}$.

In this case, if Eq. (2) can be rewritten,

$$f(y_i) = \frac{\Gamma(\frac{1}{\alpha} + y_i)}{\Gamma(\frac{1}{\alpha})\Gamma(y_i + 1)}\left(\frac{1}{1 + \alpha e^{x_i'\beta}}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha e^{x_i'\beta}}{1 + \alpha e^{x_i'\beta}}\right)^{y_i}$$

In the negative binomial regression model, the likelihood function for the MLE of parameters α and β is of the form,

$$L(y, \alpha, \beta) = \prod_{i=1}^{n} f_i(y_i) = \prod_{i=1}^{n} \frac{\Gamma(\frac{1}{\alpha} + y_i)}{\Gamma(\frac{1}{\alpha})\Gamma(y_i + 1)}\left(\frac{1}{1 + \alpha e^{x_i'\beta}}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha e^{x_i'\beta}}{1 + \alpha e^{x_i'\beta}}\right)^{y_i}$$

and the ln likelihood function,

$$lnL(y, \alpha, \beta) = \begin{pmatrix} \sum_{i=1}^{n} y_i \ln\alpha + y_i(\mathbf{x}_i'\beta) - (y_i + \frac{1}{\alpha})\ln(1 + \alpha e^{x_i'\beta}) + \ln\Gamma(\frac{1}{\alpha} + y_i) \\ - \ln\Gamma(\frac{1}{\alpha}) - \ln\Gamma(y_i + 1) \end{pmatrix}$$

is of the form. The values of α and β that maximize $lnL\ (\boldsymbol{y}, \alpha, \boldsymbol{\beta})$' were obtained as MLE (Hilbe, 2011).

Zero-inflated negative binomial regression

Zero-inflated non-negative count data have properties such as a large proportion of zero values and overdispersion. To account for these properties of zero-inflation data, the use of zero inflated Poisson (ZIP) and ZINB regression models is increasingly common. Some studies showing that the ZINB model is recommended instead of Poisson, negative binomial and ZIP regression models (Yıldırım et.al., 2022). In this study, data on pollen counts were also modelled according to the ZINB regression model.

The ZINB distribution is a mixture distribution that assigns mass π to zeros ($0 \le \pi \le 1$) and mass $(1-\pi)$ to a negative binomial distribution. The probability density of the ZINB distribution is of the form:

$$f(y) = \begin{cases} \pi + (1-\pi)\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}, for y = 0 \\ (1-\pi)\frac{\Gamma\left(\frac{1}{\alpha}+y\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y+1)}\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y}, for y \ge 1 \end{cases} \tag{3}$$

In Eq. (3), π is the probability of a structural zero. The mean and variance of the ZINB are $E(Y) = (1-\pi)\mu$ and $V(Y) = (1-\pi)\mu(1 + \pi\mu + \frac{\mu}{\alpha})$. ZINB distribution approaches the negative binomial distribution as $\pi \to 0$.

The ZINB regression model relates π and μ to covariates, that is, $E(y_i) = \mu_i = e^{x_i'\beta}$ and $logit(\pi_i) = z_i'\boldsymbol{\gamma}$ for $i = 1, \ldots, n$. Where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the corresponding vectors of regression coefficients (Mwalili et al., 2007). Parameter estimates are obtained iteratively using the method of maximum likelihood.

In the ZINB model, the purpose of the zero component is to model the probability of excess zeros observed in the data. However, it has some limitations. First, having two components in the model may both complicate interpretation and overfit the model to the data. On the other hand, if there are no zeros in the data, the zero-component causes unnecessary complexity and may reduce model's reliability. In this case, the Negative Binomial model may yield more appropriate results. The ZINB model increases the number of parameters to be estimated. This results in wider ranges of parameter estimates, especially in small samples (Fernandez & Vatcheva, 2022).

Model comparison criteria

AIC and the BIC were considered as model comparison criteria. The AIC is calculated as $AIC = 2p - 2\ln L$, where p represents the number of estimated parameters (Cameron & Trivedi, 2013). On the other hand, BIC is calculated as $BIC = 2lnn - 2\ln L$. The best model is determined by the model with the lowest BIC and AIC values (Cameron & Trivedi, 2013).

Results and discussion

Atmospheric pollen concentration is influenced by meteorological factors such as temperature, precipitation, and wind conditions. For some pollen taxa, increasing temperatures have led to higher pollen concentrations. For others, lower temperatures and wind velocity, along with precipitation and relative humidity have been shown to reduce pollen levels. Rain causes pollen to settle on the ground, preventing it from dispersing into the atmosphere. Hence, the variation in pollen counts between 2016 and 2017 for different pollen taxa in Sinop can be explained by meteorological factors.

An analysis of the meteorological data indicated that the mean temperature in 2016 (14.5 °C) was comparable to that in 2017 (14.0 °C). Furthermore, the mean wind speeds in 2016 and 2017 were 2.8 m/s and 2.7 m/s, respectively. A comparison of precipitation parameters indicated that the total precipitation in 2016 (846 mm) was marginally higher than that in 2017 (586 mm). The relative humidity levels in 2016 and 2017 were 84.5% and 80.9%. The decline in pollen production in 2017 can be explained by the combination of low temperatures and humidity, despite low total precipitation. Relative humidity, in particular, hitting of 91.7% in March, which created favorable conditions for the flowering of trees and woody plants that are typically associated with substantial pollen release.

In 2017, the average wind speed was reduced compared to 2016 (Cançı Demir et al., 2021). In this context, meteorological conditions in 2016 caused plants to bloom earlier, resulting in increased pollen production and a longer retention of pollen load in the atmosphere. This phenomenon poses significant risks to both local individuals with allergies as well as for tourists visiting.

Cupressaceae/Taxaceae, cypress; Pinaceae, pine; *Morus* sp., mulberry; *Quercus* sp., oak; *Alnus* sp., alder; *Corylus* sp., hazel; *Carpinus* sp., hornbeam; *Juglans* sp., walnut; *Fraxinus* sp., ash; *Olea europaea*, olive; Poaceae, grass; Urticaceae, nettle; *Ambrosia* sp., ragweed; Amaranthaceae, goosefoot; and *Mercurialis* sp., mercury. The monthly frequency of these pollen types from 2016 to 2017 is illustrated in Fig. 2. For instance, the *Alnus* pollen was observed mostly in February, March, and April, and was seen more in 2017 than in 2016. For additional details regarding the pollen taxa findings in Sinop province, see Cançı Demir et al. (2021).

Table 1 shows the descriptive statistics of *Alnus* pollen for 2016 and 2017. According to Table 1, the average *Alnus* pollen count in per m^3 of air was higher in 2017 than in 2016. The highest amount of *Alnus* pollen in 2016 was 209, whereas the highest amount of *Alnus* pollen in 2017 was 457. The average temperature for the months of February, March and April, when Alnus pollen is most common, was 9.72 °C in 2016, while the average temperature for the same months in 2017 was 6.94 °C. In 2016, average precipitation was 0.74 mm, and the peak precipitation was 10.4 mm. In 2017, the average precipitation was 0.1 mm, the standard deviation was 2.6 mm, and the peak precipitation was 16.2 mm, exceeding the levels observed in 2016. Similarly, the average humidity and wind speed in 2016 were lower than those in 2017. The descriptive statistics for the remaining pollen taxa were presented in Supplementary A.

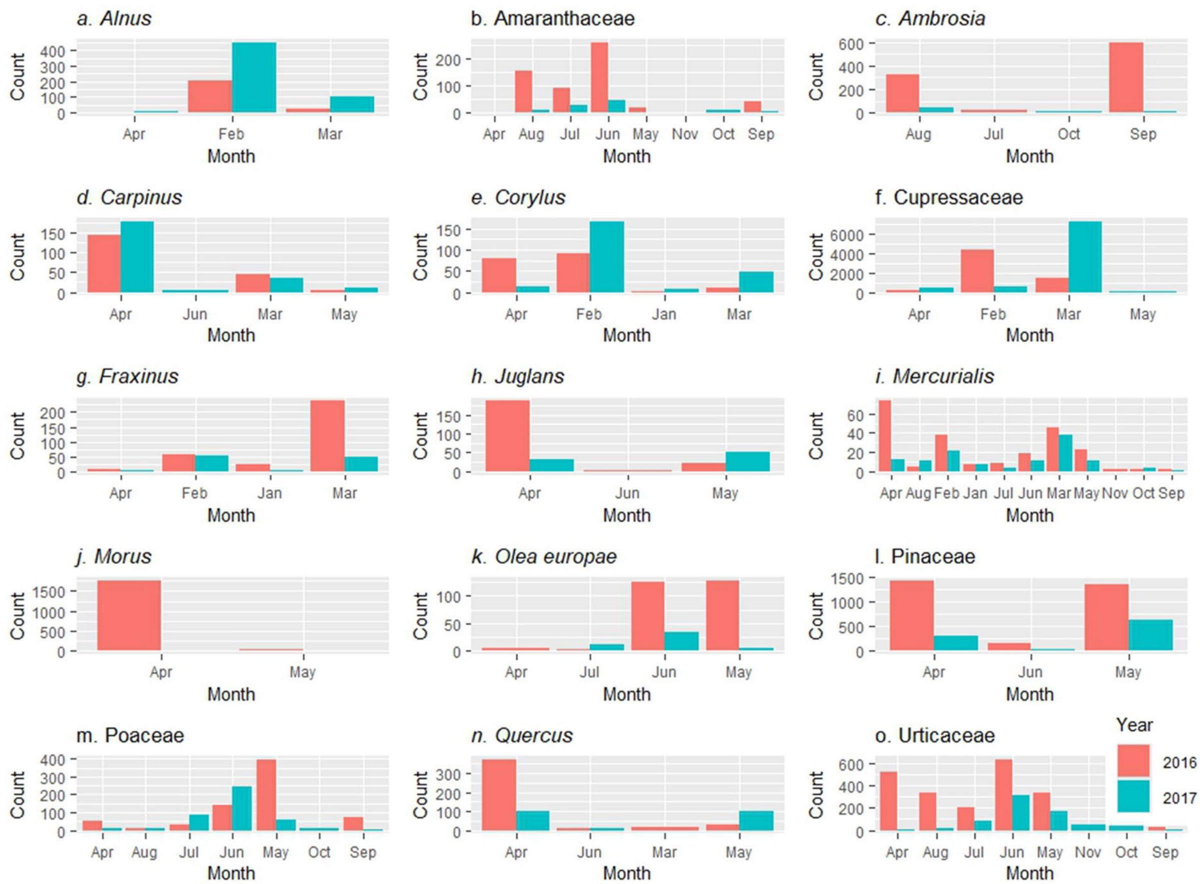

Fig. 2 Distribution of dominant pollen taxa identified in the atmosphere of Sinop province for the years 2016–2017

Table 1 Descriptive statistics of *Alnus* pollen

Year	Variable	Min	Q1	Median	Mean	Q3	Max	SD	Sum
2016	*Alnus* pollen	0	0	2	17.2	11	209	42.4	982
	Temperature (°C)	4.4	7.2	9.3	9.7	11.1	19.0	3.2	
	Precipitation(mm)	0	0	0	0.7	0.2	10.4	1.8	78.2
	Humidity (%)	53.8	78.8	86.0	84.9	95.6	99	11.9	
	Wind velocity(m/s)	1.6	2.2	2.7	2.8	3.5	4.0	0.7	
2017	*Alnus* pollen	0	1	3	26.0	13.5	452	68.3	1613
	Temperature (°C)	0.5	5.8	7.1	6.9	9.0	12.5	2.7	
	Precipitation(mm)	0	0	0	1.0	0.4	16.2	2.6	105.8
	Humidity (%)	49.1	81.5	95.4	89.0	99	99.5	12.9	
	Wind velocity(m/s)	1.4	2.3	2.8	3.1	3.475	6.6	1.2	

Tables 2 and 3 below show the results of the Poisson, negative binomial and ZINB models for Alnus pollen.

As shown in Tables 2 and 3, all p-values in the Poisson regression model are below 0.05, indicating that all predictor variables have statistically significant contributions to the model. In contrast, in the negative binomial regression model, the temperature, precipitation, and wind speed are not statistically significant predictors for *Alnus* pollen count.

Table 2 Parameter estimates ($\widehat{\boldsymbol{\beta}}$), standard error ($s_{\widehat{\beta}}$), *p*-value, AIC and BIC values for the *Alnus* pollen Poisson and negative binomial models

Pollen	Parameter	Poisson				Negative Binomial			
		Estimate	Std. error	*p*-value	Sig	Estimate	Std. error	*p*-value	Sig
Alnus	Intercept	4.090	0.240	2.49E-65	***	3.4178	1.7075	0.0453	*
	Temperature (°C)	0.135	0.009	2.70E-55	***	0.1388	0.0634	0.0286	*
	Precipitation(mm)	0.070	0.009	2.72E-14	***	−0.0547	0.0747	0.4639	
	Humidity(%)	−0.045	0.002	2.32E-117	***	−0.0423	0.0149	0.0045	**
	Wind velocity(m/s)	0.242	0.025	1.15E-22	***	0.2602	0.1737	0.1342	
	Year(ref:2016)	0.967	0.057	6.61E-64	***	1.1812	0.3606	0.0011	**
	Lag 1	0.001	0.000	2.14E-08	***	0.0112	0.0030	0.0002	***
	AIC	5310.265				777.932			
	BIC	5329.660				800.097			
	$\widehat{\alpha}$(Std. Error)	-				0.356 (0.0482)			

Sig.: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

Table 3 Parameter estimates ($\widehat{\boldsymbol{\beta}}$), standard error ($s_{\widehat{\beta}}$), *p*-value, AIC and BIC values for the *Alnus* pollen ZINB model

Pollen	Parameter	ZINB						
		Count				Zero		
		Estimate	Std. error	*p*-value	Sig	Estimate	Std. error	*p*-value
Alnus	Intercept	3.7972	1.0554	0.0003	***	−20.25	4285.29	1.00
	Temperature (°C)	0.0470	0.0387	0.2246		-		
	Precipitation(mm)	−0.0849	0.0677	0.2103				
	Humidity (%)	−0.0183	0.0094	0.0523				
	Wind velocity(m/s)	0.0181	0.1161	0.8760				
	Year(ref:2016)	0.2674	0.2379	0.2610				
	Lag 1	0.0046	0.0011	1.93E-05	***			
	AIC	784.805						
	BIC	809.741						

Sig.: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

Finally, in the ZINB model, only humidity shows a statistically significant effect on *Alnus* pollen levels.

According to the parameter estimates in the Poisson model, a 1 °C increase in temperature is associated with an approximately 1.14-fold increase in the average Alnus pollen count. Similarly, the average *Alnus* pollen count increases by 1.07 times when the amount of precipitation increases by 1 mm, while it decreased by approximately 0.96 times when the relative humidity increases by 1% one percent. The average *Alnus* pollen count increases by approximately 1.27 times when the wind speed increases by 1 m/s, and by approximately 2.63 times in 2017 compared to 2016. The estimated coefficient for the lag-1 variable indicates that a one-unit increase in the previous day's pollen count is associated with an approximate 1.13% increase in the expected pollen count on the following day. Coefficient estimates in other models can be interpreted in a similar manner.

The negative binomial regression model was found as the best-fitting model according to both the AIC and BIC criteria. As seen in Table 3, because the zero-inflation component of the ZINB model was not statistically significant, the negative binomial regression model was selected as the most appropriate model.

The results of the Poisson, Negative Binomial and the ZINB model for other pollen taxa were given in Supplementary B. When AIC and BIC values were examined for all pollen taxa, the best model was found as the negative binomial regression model except for *Morus, Mercurialis* and Poaceae.

Furthermore, in the dispersion test, the z-statistic corresponding to the Poisson model for the *Alnus* pollen was 2.3823, with a corresponding p-value of 0.0086 (see Supplementary C-a). The overdispersion parameter estimate ($\widehat{\alpha}$) was found to be 0.356. The fact that ($\widehat{\alpha}$) was close to zero supported the selection of the negative binomial regression model. The residuals from the negative binomial regression model for Alnus pollen showed no significant deviation from normality based on the K-S test (p-value = 0.4811) (see Supplementary D-a). The results of the dispersion and normality tests for other pollen taxa are provided in Supplementary C-a and D-a, respectively. Supplementary D-a displays K-S tests and Q-Q plots (see Figure D1-D15) generated through the DHARMa package for all pollen taxa which utilizes simulated data derived from estimated model parameters. For the Alnus pollen, the residuals followed a normal distribution. For the negative binomial regression model without the lag 1 variable, the ACF graphs (see Figure C1-C15) were provided for all pollen taxa in Supplementary C-b, and for the models including the lag 1 variable, the Q-Q plots (see Fig. 1a-15b) assessing residual normality based on deviance statistics were presented in Supplementary D-c. Examination of the Q–Q plots for *Alnus* pollen indicated that, in the negative binomial regression model, the residuals exhibited a better fit to normality compared with the Poisson regression model (see Fig. 1a-b in Supplementary D-c).

In the Ljung-Box test conducted to assess the autocorrelation of residuals for *Alnus* pollen, the p-value was 0.2048. Thus, it can be concluded that there is no evidence of autocorrelation among the residuals. Autocorrelation test results for residuals of other pollens are presented in the Supplementary D-b.

As shown in Tables 4 and 5, the statistical significance of the independent variables-temperature, precipitation, relative humidity, wind speed, and year varied among the pollen taxa, depending on their respective models.For example, in the negative binomial model for Alnus pollen, precipitation and wind speed variables were not significant, while

Table 4 Effect of independent variables on pollen quantity according to the negative binomial regression model

Pollen	Intercept	Temperature (°C)	Precipitation (mm)	Humidity (%)	Wind velocity (m/s)	Year	Lag 1
Alnus	■	■	○	▲	○	■	■
Amaranthaceae	■	■	▲	▲	○	▲	■
Ambrosia	○	■	○	▲	○	▲	■
Carpinus	■	○	○	○	■	○	■
Corylus	■	■	▲	▲	○	■	■
Cupressaceae	■	○	▲	▲	▲	■	■
Fraxinus	○	■	▲	○	○	○	■
Juglans	■	○	○	▲	○	○	■
Mercurialis	■	▲	▲	○	○	▲	■
*Morus**	○	■	○	○	○	○	■
Olea-europaea	■	○	■	▲	○	○	■
Pinaceae	■	▲	▲	▲	○	▲	■
Poaceae	■	▲	○	○	○	▲	■
Quercus	■	▲	○	▲	■	▲	■
Urticaceae	■	▲	○	○	○	▲	■

■, significant positive effect; ○, insignificant; ▲, significant negative effect

*For *Morus* pollen, due to the lack of convergence in the negative binomial regression model, the effects of coefficients are provided based on the ZINB regression model

Table 5 Coefficient of the negative binomial regression model

Pollen	Intercept	Temperature (°C)	Precipitation (mm)	Humidity (%)	Wind velocity (m/s)	Year	Lag 1
Alnus	3.4178	0.1388	-0.0547	-0.0423	0.2602	1.1812	0.0112
Amaranthaceae	1.9464	0.0903	-0.0417	-0.0209	0.1427	-2.1191	0.0231
Ambrosia	2.0703	0.1547	0.0565	-0.0303	0.1885	-2.6448	0.0083
Carpinus	3.9633	-0.0661	-0.0429	-0.0265	0.3736	-0.0910	0.0202
Corylus	2.7917	0.1108	-0.0964	-0.0274	0.0033	0.5366	0.0300
Cupressaceae	9.6573	-0.0302	-0.1425	-0.0388	-0.4896	0.7537	0.0008
Fraxinus	-0.1797	0.2550	-0.1476	0.0002	-0.1314	-0.1373	0.0199
Juglans	4.6135	-0.0074	-0.0299	-0.0391	0.1378	0.1208	0.0347
Mercurialis	3.1745	-0.0899	-0.0655	-0.0060	-0.0353	-0.6918	0.0603
*Morus**	4.1649	0.1340	-0.0188	-0.0132	-0.1666	-0.5154	0.0009
Olea-europaea	7.5228	0.0375	0.0977	-0.0827	-0.2209	-0.3122	0.0763
Pinaceae	14.2887	-0.1431	-0.0773	-0.0902	0.1124	-1.1122	0.0028
Poaceae	3.4144	-0.0336	-0.0320	-0.0066	0.1019	-0.5169	0.0193
Quercus	8.3909	-0.1187	-0.0316	-0.0601	0.2674	-0.6347	0.0226
Urticaceae	4.4651	-0.0746	0.0099	0.0014	0.0043	-0.7544	0.0120

*Since convergence was not achieved in the negative binomial regression model for *Morus* pollen, the effects of the coefficients were presented using the ZINB regression model (green, positive effect; red, negative effect; yellow, insignificant)

temperature, year, and lag 1 variables had significant positive effects, and humidity variable had a significant negative effect. For Cupressaceae pollen, temperature was only independent variable that was not statistically significant. Similarly, precipitation had significant negative effect for Amaranthaceae pollen and significant positive effect for *Olea-europaea* pollen. In almost all pollen types, the average pollen count was lower in 2017 than in 2016.

Conclusions

This study addressed the overdispersion in pollen taxa data by applying negative binomial regression in addition to the Poisson regression. Given the high frequency of zero counts in the dataset, ZINB regression was also employed to better model the data distribution. Model coefficients for each pollen taxon were estimated for all regression models and performances of models were evaluated using the AIC and BIC values. Based on the AIC and BIC values, the negative binomial regression model demonstrated a superior fit compared to the Poisson regression for the majority of pollen taxa. The only exception was Morus pollen, for which the negative binomial model failed to converge; consequently, the ZINB regression model was selected as the best-fitting alternative for this taxon. For *Mercurialis* and Poaceae pollens, the AIC and BIC values of the negative binomial and ZINB models were very similar, with the ZINB model showing a marginally better fit according to the AIC and BIC values. Therefore, for the purpose of consistency and model comparison across taxa, the negative binomial model was identified as the best overall model for the dataset. Eventually, the residuals in the regression models for the pollen taxa taken into consideration in the study were analyzed for normality and autocorrelation.

The Poisson model becomes inadequate in the presence of overdispersion. While the negative binomial model addresses this issue, it may produce wider uncertainty ranges. The ZINB regression model, though useful for handling excess zeros, has limitations such as insufficient parameter estimation for non-zero observations. Despite these known drawbacks, these models were applied to pollen taxa count data, and the relationship between pollen counts and meteorological factors was best captured using the negative binomial regression model.

The temperature dependent variable is insignificant for *Carpinus*, Cupressaceae, *Juglans* and *Olea-europaea*, while wind speed is insignificant except for *Carpinus*, Cupressaceae and *Quercus* pollen. The precipitation variable has a negative significance for Amaranthaceae, *Corylus*, Cupressaceae, *Mercurialis*, and Pinaceae, while the humidity variable has a negative significance for *Alnus*, Amaranthaceae, *Ambrosia*, *Corylus*, Cupressaceae, *Juglans*, *Olea-europaea*, Pinaceae, and *Quercus*. The lag 1 variable has a positive effect on all pollen counts. Therefore, the previous day's pollen count increases the next day's pollen count. The year variable has a negative effect for Amaranthaceae, *Ambrosia*, *Mercurialis*, Pinaceae, Poaceae, *Quercus*, and Urticaceae, while it has a positive effect for *Alnus*, *Corylus*, and Cupressaceae. Therefore, compared to 2016, the expected pollen counts for Amaranthaceae, *Ambrosia*, *Mercurialis*, Pinaceae, Poaceae, *Quercus*, and Urticaceae decreased in 2017, while they increased for *Alnus*, *Corylus*, and Cupressaceae.

This study on allergenic pollen-producing plants in the flora of Sinop province and its surroundings is important in terms of creating a pollen calendar and evaluation of allergenic risks. The results of this study suggest that aeropalynological research should be continued in various regions, including Sinop province and its districts. Further studies on allergenic pollen-producing plants would be beneficial not only for allergy sufferers and physicians but also for biologists, environmental scientists, and climatologists working in various botanical and ecological fields.

Author contributions A.Y. and C.H.: methodology, formal analysis, writing—original draft; (%40) H.Ö.: data sourcing, editing; C.C.D. and A.K.: data curation, investigation (%40). N.D. and S.S.K.: writing, review and editing (%20). All authors have read and agreed to the published version of the manuscript.

Funding The authors declare that no funding was received for this study.

Data availability The data that support the findings of this study are not publicly available but were used to perform the analyses presented in the manuscript.

Declarations

Ethics approval All authors confirm that they have reviewed, comprehended, and adhered to the ethical responsibilities outlined in the journal's Instructions for Authors, where applicable.

Competing interests The authors declare no competing interests.

References

Aboulaich, N., Achmakh, L., Bouziane, H., Trigo, M. M., Recio, M., Kadiri, M., Cabezudo, B., Riadi, H., & Kazzaz, M. (2013). Effect of meteorological parameters on *Poaceae* pollen in the atmosphere of Tetouan (NW Morocco). *International Journal of Biometeorology, 57*(2), 197–205. https://doi.org/10.1007/s00484-012-0566-2

Adams-Groom, B., Selby, K., Derrett, S., Frisk, C. A., Pashley, C. H., Satchwell, J., King, D., McKenzie, G., & Neilson, R. (2022). Pollen season trends as markers of climate change impact: *Betula*, *Quercus* and *Poaceae*. *Science of the Total Environment, 831*, Article 154882. https://doi.org/10.1016/j.scitotenv.2022.154882

Agresti A. (2015). Foundations of linear and generalized linear models. *JohnWiley & Sons*, 472p,Wiley Series in Probability and Statistics.

Aguilera, F., Fornaciari, M., Ruiz-Valenzuela, L., Galán, C., Msallem, M., Dhiab, A., la Guardia, C. D., del Mar Trigo, M., Bonofiglio, T., & Orlandi, F. (2015). Phenological models to predict the main flowering phases of olive (*Olea europaea* L.) along a latitudinal and longitudinal gradient across the mediterranean region. *International Journal of Biometeorology, 59*(5), 629–641. https://doi.org/10.1007/s00484-014-0876-7

Akdoğan Karadağ, G. E., & Altunoğlu, M. K. (2025). Airborne pollen seasonality of Kars province, a high – altitude region in NE Anatolia-Turkey. *Palynology*. https://doi.org/10.1080/01916122.2024.2382959

Akpınar, S., & Altunoğlu, M. K. (2024). Determination of atmospheric pollen grains by volumetric method in Sarıkamış district (Kars-Türkiye). *Biology, 13*(7), Article 475. https://doi.org/10.3390/biology13070475

Aytuğ, B. 91973). Calendrier pollinique de la region d'İstanbul. *U. I. Forest Fac. Rew*. XXIII, A1: 1-33.

Bielory, L., Lyons, K., & Goldberg, R. (2012). Climate change and allergic disease. *Current Allergy and Asthma Reports, 12*, 485–494. https://doi.org/10.1007/s11882-012-0314-z

Bonini, M., Šikoparija, B., Prentović, M., Cislaghi, G., Colombo, P., Testoni, C., Grewling, L., Lommen, S. T. E., Muller-Scharer, H., & Smith, M. (2015). Is the recent decrease in airborne *Ambrosia* pollen in the Milan area due to the accidental introduction of the ragweed leaf beetle *Ophraella communa*?

Aerobiologia, 31(4), 499–513. https://doi.org/10.1007/s10453-015-9380-8

Brighetti, M. A., Costa, C., Menesatti, P., Antonucci, F., Tripodi, S., & Travaglini, A. (2014). Multivariate statistical forecasting modeling to predict *Poaceae* pollen critical concentrations by meteoclimatic data. *Aerobiologia, 30*(1), 25–33. https://doi.org/10.1007/s10453-013-9305-3

Brooks, E. M., Kristensen, K., Benthem, J. K., Magnusson, A., Berg, W. C., Nielsen, A., Skaug, J. H., Maechler, M., & Bolker, M. B. (2017). GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association, 65*, 1509–1526. https://doi.org/10.2307/2284333

Cameron A.C., & Trivedi P.K. (2013). *Regression analysis of count data* (2nd ed.), Cambridge University Press

Canşı Demir, C., Özler, H., & Kaplan, A. (2021). Effects of meteorological factors on pollen flora in the atmosphere of Sinop (Türkiye). *Bangladesh Journal of Botany, 50*(1), 147–157. https://doi.org/10.3329/bjb.v50i1.52682

Cordero, J. M., Rojo, J., Gutiérrez-Bustillo, A. M., Narros, A., & Borge, R. (2020). Predicting the *Olea* pollen concentration with a machine learning algorithm ensemble. *International Journal of Biometeorology, 65*, 541–554. https://doi.org/10.1007/s00484-020-02047-z

Čorić, R., Matijević, D., & Marković, D. (2023). Pollennet-a deep learning approach to predicting airborne pollen concentrations. *Croatian Operational Research Review, 14*(1), 1–13. https://doi.org/10.17535/crorr.2023.0001

Damialis, A., Traidl-Hoffmann, C., & Treudler, R. (2019) Climate change and pollen allergies. In: Marselle MR, Stadler, J., & Korn, H. et al. (eds) Biodiversity and health in the face of climate change. *Springer International Publishing*, Cham, pp 47–66. https://doi.org/10.1007/978-3-030-02318-8_3.

De Weger, L. A., Beerthuizen, T., Hiemstra, P. S., & Sont, J. K. (2014). Development and validation of a 5-day-ahead hay fever forecast for patients with grass-pollen-induced allergic rhinitis. *International Journal of Biometeorology, 58*(7), 1047–1055. https://doi.org/10.1007/s00484-013-0753-9

Fernandez, G. A., & Vatcheva, K. P. (2022). A comparison of statistical methods for modeling count data with an application to hospital length of stay. *BMC Medical Research Methodology, 22*, 211. https://doi.org/10.1186/s12874-022-01685-8

Fernandez-Rodriguez, S., Duran-Barroso, P., Silva-Palacios, I., Tormo-Molina, R., Maya-Manzano, J. M., & Gonzalo-Garijo, A. (2016a). Regional forecast model for the *Olea* pollen season in Extremadura (SW Spain). *International Journal of Biometeorology, 60*(10), 1509–1517. https://doi.org/10.1007/s00484-016-1141-z

Fernandez-Rodriguez, S., Duran-Barroso, P., Silva-Palacios, I., Tormo-Molina, R., Maya-Manzano, J. M., & Gonzalo-Garijo, A. (2016b). Forecast model of allergenic hazard using trends of *Poaceae* airborne pollen over an urban area in SW Iberian Peninsula (Europe). *Natural Hazards, 84*(1), 121–137. https://doi.org/10.1007/s11069-016-2411-0

Fernandez-Rodriguez, S., Duran-Barroso, P., Silva-Palacios, I., Tormo-Molina, R., Maya-Manzano, J. M., & Gonzalo-Garijo, A. (2016c). *Quercus* long-term pollen season trends in the southwest of the Iberian Peninsula. *Process Safety and Environmental Protection, 101*, 152–159. https://doi.org/10.1016/j.psep.2015.11.008

Garcia-Mozo, H., Oteros, J., & Galan, C. (2015). Phenological changes in olive (*Ola europaea* L.) reproductive cycle in southern Spain due to climate change. *Annals of Agricultural and Environmental Medicine, 22*(3), 421–428. https://doi.org/10.5604/12321966.1167706

Garcia-Mozo, H., Yaezel, L., Oteros, J., & Galan, C. (2014). Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. *Science of the Total Environment, 473*, 103–109. https://doi.org/10.1016/j.scitotenv.2013.11.142

Hartig, F. (2024). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. https://doi.org/10.32614/CRAN.package.DHARMa

Hilbe, J. (2011). Negative Binomial Regression, 2nd ed., *New York: Cambridge University Press*, 573p. https://doi.org/10.1017/CBO9780511973420.

Janati, A., Bouziane, H., del Mar Trigo, M., Kadiri, M., & Kazzaz, M. (2017). Poaceae pollen in the atmosphere of Tetouan (NW Morocco): Effect of meteorological parameters and forecast of daily pollen concentration. *Aerobiologia, 33*, 517–528. https://doi.org/10.1007/s10453-017-9487-1

Klimczak, L. J., von Eschenbach, C. E., Thompson, P. M., Buters, J. T. M., & Mueller, G. A. (2020). Mixture analysis of air-sampled pollen extracts can accurately differentiate pollen taxa. *Atmospheric Environment, 243*(15), Article 117746. https://doi.org/10.1016/j.atmosenv.2020.117746

Karaer, F., & Kılınç, M. (1993). Sinop yarımadasının florası. *Turk. J. of Botany, 17*, 5–20.

Kılınç, M., & Karaer, F. (1995). Sinop yarımadasının vejetasyonu. *Turkish Journal of Botany, 19*, 107–124.

Kleiber, C., & Zeileis, A. (2008). *Applied Econometrics with R*. Springer-Verlag. https://doi.org/10.1007/978-0-387-77318-6

Lara, B., Rojo, J., Fernandez-Gonzalez, F., & Perez-Badia, R. (2019). Prediction of airborne pollen concentrations for the plane tree as a tool for evaluating allergy risk in urban green areas. *Landscape and Urban Planning, 189*, 285–295. https://doi.org/10.1016/j.landurbplan.2019.05.002

Liu, X., Wu, D., Zewdie, G. K., Wijerante, L., Timms, C. I., Riley, A., & Lary, D. J. (2017). Using machine learning to estimate atmospheric *Ambrosia* pollen concentrations in Tulsa, OK. *Environmental Health Insights, 11*, Article 1178630217699399. https://doi.org/10.1177/1178630217699399

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*, 297–303. https://doi.org/10.2307/2335207

Lo, F., Bitz, C. M., & Hess, J. J. (2021). Development of a random forest model for forecasting allergenic pollen in North America. *Science of the Total Environment, 773*, Article 145590. https://doi.org/10.1016/j.scitotenv.2021.145590

Malkiewicz, M., Klaczak, K., Drzeniecka-Osiadacz, A., Krynicka, J., & Migała, K. (2014). Types of *Artemisia* pollen season depending on the weather conditions in Wrocław (Poland), 2002–2011. *Aerobiologia, 30*(1), 13–23. https://doi.org/10.1007/s10453-013-9304-4

Myszkowska, D. (2013). Prediction of the birch pollen season characteristics in Cracow, Poland using an 18-year data series. *Aerobiologia, 29*(1), 31–44. https://doi.org/10.1007/s10453-012-9260-4

Mwalili, S., Lesaffre, E., & Declerck, D. (2007). The zero-inflated negative binomial regression model with correction for misclassification: An example in caries research. *Statistical Methods in Medical Research, 17*(2), 123–139. https://doi.org/10.1177/0962280206071840

Nowosad, J., Stach, A., Kasprzyk, I., Weryszko-Chmielewska, E., Piotrowska-Weryszko, K., Puc, M., Grewling, Ł, Pędziszewska, A., Uruska, A., & Myszkowska, D. (2016). Forecasting model of *Corylus*, *Alnus*, and *Betula* pollen concentration levels using spatiotemporal correlation properties of pollen count. *Aerobiologia, 32*(3), 453–468. https://doi.org/10.1007/s10453-015-9418-y

Ocana-Peinado, F. M., Valderrama, M. J., & Bouzas, P. R. (2013). A principal component regression model to forecast airborne concentration of Cupressaceae pollen in the city of Granada (SE Spain), during 1995–2006. *International Journal of Biometeorology, 57*(3), 483–486. https://doi.org/10.1007/s00484-012-0527-9

O'Connor, D., Markey, E., Jose Maria, M.J., Dowding, P., Aoife Donnelly, A., & Sodeau, J .(2017). *Pollen Monitoring and Modelling* (POMMEL). Epa Research Programme 2021–2030.

Oteros, J., Garcia-Mozo, H., Hervas, C., & Galan, C. (2013). Biometeorological and autoregressive indices for predicting olive pollen intensity. *International Journal of Biometeorology, 57*(2), 307–316. https://doi.org/10.1007/s00484-012-0555-5

Özen, F., & Kılınç, M. (1995). Alaçam-Gerze ve Boyabat-Durağan arasında kalan bölgenin florası. *Tr. J. Of Botany* 19: 241-275.

Özkaragöz, K., & Karamanoğlu, K. (1967). Survey of allergenic pollen grains and mold spores in the Ankara area., Rev. Palaeobot. Pal., 4: 251-256. https://doi.org/10.1016/0034-6667(67)90193-5.

Picornell, A., Hurtado, S., Antequera-Gómez, M. L., Barba-González, C., Ruiz-Mata, R., de Gálvez-Montañez, E., & Navas-Delgado, I. (2024). A deep learning LSTM-based approach for forecasting annual pollen curves: *Olea* and *Urticaceae* pollen types as a case study. *Computers in Biology and Medicine, 168*, Article 107706. https://doi.org/10.1016/j.compbiomed.2023.107706

Pedersen, T. (2024). *patchwork: The Composer of Plots*. R package version 1.3.0.9000, https://github.com/thomasp85/patchwork, https://patchwork.data-imaginist.com.

Piotrowska-Weryszko, K. (2013). The effect of the meteorological factors on the *Alnus* pollen season in Lublin (Poland). *Grana, 52*(3), 221–228. https://doi.org/10.1080/00173134.2013.772653

Puc, M., & Wolski, T. (2013). Forecasting of the selected features of *Poaceae* (R. Br.) Barnh., *Artemisia* L. and *Ambrosia* L. pollen season in Szczecin, north-western Poland, using Gumbel's distribution. *Annals of Agricultural and Environmental Medicine, 20*(1), 36–47.

Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., & Genikhovich, E. (2016). Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: Example of birch pollen. *Agricultural and Forest Meteorology, 226*, 96–107. https://doi.org/10.1016/j.agrformet.2016.05.016

Rodríguez-Rajo, F. J., Astray, G., Ferreiro-Lage, J. A., Aira, M. J., Jato-Rodriguez, M. V., & Mejuto, J. C. (2010). Evaluation of atmospheric Poaceae pollen concentration using a neural network applied to a coastal Atlantic climate region. *Neural Networks, 23*(3), 419–425. https://doi.org/10.1016/j.neunet.2009.06.006

Rojo, J., Rivero, R., Romero-Morte, J., Fernandez-Gonzalez, F., & Perez-Badia, R. (2017). Modeling pollen time series using seasonal- trend decomposition procedure based on LOESS smoothing. *International Journal of Biometeorology, 61*(2), 335–348. https://doi.org/10.1007/s00484-016-1215-y

Sabariego, S., Cuesta, P., Fernandez-Gonzalez, F., & Perez-Badia, R. (2011). Models for forecasting airborne Cupressaceae pollen levels in central Spain. *International Journal of Biometeorology, 56*(2), 253–258. https://doi.org/10.1007/s00484-011-0423-8

Shokouhi, B. V., de Hoogh, K., Gehrig, R., & Eeftens, M. (2024). Estimation of historical daily airborne pollen concentrations across Switzerland using a spatio temporal random forest model. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2023.167286

Sicard, P., Thibaudon, M., Besancenot, J. P., & Mangin, A. (2012). Forecast models and trends for the main characteristics of the *Olea* pollen season in Nice (south-eastern France) over the 1990–2009 period. *Grana, 51*(1), 52–62. https://doi.org/10.1080/00173134.2011.637577

Silva-Palacios, I., Fernandez-Rodriguez, S., Duran-Barroso, P., Tormo-Molina, R., Maya-Manzano, J. M., & Gonzalo-Garijo, A. (2016). Temporal modelling and forecasting of the airborne pollen of Cupressaceae on the southwestern Iberian Peninsula. *International Journal of Biometeorology, 60*(2), 297–306. https://doi.org/10.1007/s00484-015-1026-6

Siniscalco, C., Caramiello, R., Migliavacca, M., Busetto, L., Mercalli, L., Colombo, R., & Richardson, D. (2015). Models to predict the start of the airborne pollen season. *International Journal of Biometeorology, 59*(7), 837–848. https://doi.org/10.1007/s00484-014-0901-x

Tassan-Mazzocco, F., Felluga, A., & Verardo, P. (2015). Prediction of wind-carried Gramineae and Urticaceae pollen occurrence in the Friuli Venezia Giulia region (Italy). *Aerobiologia, 31*(4), 559–574. https://doi.org/10.1007/s10453-015-9386-2

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. *International Journal of Biometeorology, 63*, 735–745. https://doi.org/10.1007/s00484-019-01688-z

Vélez-Pereira, A. M., De Linares, C., Canela, M., & Belmonte, J. (2019). Logistic regression models for predicting daily airborne *Alternaria* and *Cladosporium* concentration levels in Catalonia (NE Spain). *International Journal of Biometeorology, 63*, 1541–1553. https://doi.org/10.1007/s00484-019-01767-1

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed). Springer. https://doi.org/10.1007/978-0-387-21706-2

Viney, A., Nicolás, J. F., Galindo, N., Fernández, J., Soriano-Gomis, V., & Varea, M. (2021). Assessment of the external contribution to *Olea* pollen levels in southeastern Spain. *Atmospheric Environment*. https://doi.org/10.1016/j.atmosenv.2021.118481

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis. *Springer-Verlag*. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2023). *readxl: Read Excel files* (Version 1.4.3) [R package]. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=readxl.

Yıldırım, G., Kaçıranlar, S., &Yıldırım, H. (2022). Poisson and negative binomial regression models for zero-inflated data: an experimental study. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics* 71 (2) 601–615. https://doi.org/10.31801/cfsuasmas.988880.

Zhang, Y., & Steiner, A. L. (2022). Projected climate-driven changes in pollen emission season length and magnitude over the continental United States. *Nature Communications, 13*, Article 1234. https://doi.org/10.1038/s41467-022-28764-0

Zhang, Y., Bielory, L., Mi, Z., Cai, T., Roboc, A., & Georgopoulos, P. (2015a). Allergenic pollen season variations in the past two decades under changing climate in the United States. *Global Change Biology, 21*(4), 1581–1589. https://doi.org/10.1111/gcb.12755

Zhang, Y., Isukapalli, S. S., Bielory, L., & Georgopoulos, P. G. (2013). Bayesian analysis of climate change effects on observed and projected airborne levels of birch pollen. *Atmospheric Environment, 68*, 64–73. https://doi.org/10.1016/j.atmosenv.2012.11.028

Zhang, Y., Bielory, L., Cai, T., Mi, Z., & Georgopoulos, P. (2015b). Predicting onset and duration of airborne allergenic pollen season in the United States. *Atmospheric Environment, 103*, 297–306. https://doi.org/10.1016/j.atmosenv.2014.12.019