

Regression Models to Forecast Cupressaceae Pollen Concentration in the City of Granada (SE of Spain) Based on Climatic Variables

Ocana-Peinado FM*

Department of Statistics and Operations Research, University of Granada, Faculty of Pharmacy, Campus de Cartuja s/n, Granada, Spain

*Corresponding author: Ocana-Peinado FM, Department of Statistics and Operations Research, University of Granada, Faculty of Pharmacy, Campus de Cartuja s/n, Granada, Spain, Tel: 958 24 66 62; E-mail: fmocan@ugr.es

Rec date: March 17, 2016; Acc date: March 22, 2016; Pub date: March 29, 2016

Copyright: © 2016 Ocana-Peinado FM. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Cupressaceae pollen has one of the highest pollen incidences in the Mediterranean area and is present in the atmosphere practically all year round, although it is predominant in the winter period, when no other plants are flowering, making this particle a powerful allergen. In Europe, allergy to Cupressaceae pollen was considered a rarity until 1975, but is now a recognized clinical entity, Belmonte et al. [1].

The cypress is one of the most widely found anemophilous trees in the world, and the effects of its pollen have a very high incidence among the population living in noncoastal areas. Its pollination season in the northern hemisphere covers approximately the first quarter of the year, with several high peaks at the end of the winter.

One of the most important tools employed in health-care planning is the anticipation of peak pollen concentrations because of their repercussion on allergy episodes, especially among children and the elderly. During the winter, allergy to pollen from trees can be confused with a common cold, and inadequate treatment can provoke chronic asthma. Various climatic variables may influence pollen concentration including environmental humidity, temperature, wind speed, hours of sunshine, and others that have been considered by several authors. As an example, Stark et al. [2] concluded, by means of Poisson regression model for ragweed pollen, that rain, wind speed, a smooth function of temperature and its residuals, as well as the day in the pollen season and its logarithmic transform are the most significant variables.

Due to the prevalence of this pollen in Granada during the winter, the relationship between the most important meteorological parameters on daily pollen counts was researched in order to investigate the conditions that influence the prevalence of Cupressaceae pollination. The present paper is focused on the presentation of three dynamic regression model used to forecast Cupressaceae pollen by means of climatic variables.

Forecasting methods for time series may be self-explicative or based on dynamic regression. The former only consider the past history of the series itself, whereas the latter methods include information from an input process, with the residuals also being represented as an autoregressive integrated moving average (ARIMA) model. Nevertheless, neither of the above techniques are suitable for forecasting peak pollen concentrations due to the very sparse nature of the data and especially because there are discontinuities in the atmospheric presence of pollen, which is limited to the first quarter of the year, reaching a peak by the second half of February.

Methodologies

The first strategy, Ocaña-Peinado et al. [3], is to derive a transfer function model with multiplicative intervention variable in order to

forecast air pollen concentration using the temperature as input series. The inertia process is at the same time modeled by means of a principal component analysis (PC) after a suitable time rescaling. This double modelization is known as TF-PC model and it is an alternative to the classical Box and Jenkins methodology. In this paper, we can observe that the TF-PC model provides more accurate forecasts than the classical Box and Jenkins method.

The second procedure is to apply functional data analysis, Valderrama et al. [4]. In this paper functional regression model based on functional principal component (FPC) analysis is proposed.

For FPC selection, we will apply a criterion based on the proportion of error reduction of each pair (past and current intervals) in the model, once they have been decreasingly ordered by their explained variance. We assume that the residual process has a dependence structure from the past history of pollen concentration in terms of another functional regression model, whereas existing methodologies assume either that the residuals constitute white noise or, at most, that they are autoregressive (for instance, transfer function models). Furthermore, the model performs a continuous-time prediction in a given (current) interval.

This regression model enabling us to forecast cypress pollen concentration in the interval spanning 15–28 February each year, on the basis of prior knowledge and of the air temperature during the previous month of January and the first half of February. Due to the seasonal behavior of cypress pollen, a long time series cannot be used; instead, it is truncated each year in an interval so that a set of sample curves are available.

In the third procedure, Ocaña-Peinado et al. [5], the problem of developing a 2-week-on ahead forecast of atmospheric cypress pollen levels is tackled by developing a principal component multiple regression model involving several climatic variables.

Based on the work of different authors, Díaz de la Guardia et al. [6]; Galán et al. [7]; Sabariego et al. [8]; Tortajada and Mateu [9], factors as temperature, humidity, and hours of sun and wind speed were incorporated in the model. This methodology explains approximately 75–80% of the variability in the airborne Cupressaceae pollen concentration.

The aim of this third paper was to select a set of variables suitable for modeling the stochastic process of Cupressaceae airborne pollen concentration during the pollination season. To do so, a dimensionality reduction on the basis of principal component analysis (PCA) was developed for both this process and for some climatic processes. The time predictive approach applied in this model means that the sample paths of the main processes were recorded one week in

advance of the others. Multiple linear regression among the principal components was then performed to obtain the predictive model.

In agreement with the results of other studies in the Mediterranean area (cited above), meteorological variables such as daily average temperature, daily humidity, daily hours of sun and daily maximum wind speed, were revealed as predictors to perform APC multiple regression model.

Conclusions

In our papers, we can observe that the proposed methodologies, captures the trends in an optimal way, and allows the anticipation of the appearance of peaks in the Cupressaceous airborne pollen process. However, pollen levels are related not only to meteorological variables; human activities such as pruning, watering, or introduction or elimination of plants can modify pollen values.

References

1. Belmonte J, Canela M, Guàrdia R, Guàrdia RA, Sbai L, et al. (1999) Aerobiological dynamics of the Cupressaceae pollen in Spain 1992–1998. *Polen* 10: 27-38.
2. Diaz de la Guardia C, Alba F, de Linares C, Nieto-Lugilde D, López Caballero J (2006) Aerobiological and allergenic analysis of cupressaceae pollen in Granada (Southern Spain). *J Investig Allergol Clin Immunol* 16: 24-33.
3. Galán C, Fuillerat MJ, Comtois P (1998) Bioclimatic factors affecting daily Cupressaceae flowering in southwest Spain. *Int J Biometeorol* 41: 95-100.
4. Ocaña-Peinado FM, Valderrama MJ, Aguilera AM (2008) A dynamic regression model for air pollen concentration. *Stoch Environ Res Risk Assess* 22: 59-63.
5. Ocaña-Peinado FM, Valderrama MJ, Bouzas PR (2013) A principal component regression model to forecast airborne concentration of Cupressaceae pollen in the city of Granada (SE Spain), during 1995-2006. *Int J Biometeorol* 57: 483-486.
6. Sabariego S, Cuesta P, Fernández-González F, Pérez-Badia R (2012) Models for forecasting airborne Cupressaceae pollen levels in central Spain. *Int J Biometeorol* 56: 253-258.
7. Stark PC, Ryan LM, McDonald JL (1997) Using meteorological data to predict daily ragweed pollen levels. *Aerobiologia* 13: 177-184.
8. Tortajada B, Mateu I (2008) Cupressaceae pollen in the atmosphere of Valencia (East of Spain) and relationship with meteorological parameters. *Polen* 18: 51-59.
9. Valderrama MJ, Ocaña FA, Aguilera AM, Ocaña-Peinado FM (2010) Forecasting pollen concentration by a two-step functional model. *Biometrics* 66: 578-585.