METMA V International Workshop on Spatio-Temporal Modelling
Santiago de Compostela
30 Jun - 2 Jul 2010

A mixed functional model for high peaks of Olea europaea l. airborne pollen

Escabias, M.[1*], Valderrama, M. J.[1], Aguilera, A. M[2] and Ocaña, F. A.[1]

[1*] *Departamento de Estadística e I.O., Facultad de Farmacia, Universidad de Granada. Campus de Cartuja S/N. 18071 Granada (Spain); escabias@ugr.es, valderra@ugr.es, focana@ugr.es*
[2] *Departamento de Estadística e I.O., Facultad de Ciencias, Universidad de Granada. Av. Severo Ochoa S/N. 18071 Granada (Spain); aaguiler@ugr.es*
[*] *Corresponding author*

Abstract. *Several regression schemes have been tested in order to model the Olea europaea l. airborne pollen considering several environmental variables as explicative ones. The novelty consists of considering a logit model with functional and non-functional variables as regressor recorded since 1992 to 2003, to predict the presence or absence of peaks of pollen concentration as binary response variable.*

Keywords. *Olea-europaea-l.-airborne-pollen, Functional-logit-regression. Environmental-data*

---

1 Introduction

High olive pollen concentration is one of those events that occur every year in late spring and early summer in provinces of the south of Spain (Jaén, Granada and Córdoba). This is because olive is the main crop in those provinces. High level of airborne olive pollen concentration is a problem for an important part of the population of those zones because of the amount of allergies that it causes, and make people to have to move their residences to coastal cities in that period. Many studies can be found about the factors that influence the amount of airborne pollen (see for example Alba *et al.* (2000), Galán *et al.* (2001) and Valderrama *et al.* (2010)). This paper offers a different related point of view in order to study the relationship between olive pollen concentration and different factors from the perspective of functional data analysis (FDA). This statistical methodology uses continuity and time-dependency of variables as tools for modeling and forecasting variables. In the olive pollen case it would be very interesting to know the probability of high (maybe extreme) level of pollen concentration in the future in order to minimize this effect on allergic people. So we analyze the relationship between the functional variables (time-dependent) that influence the olive pollen concentration. First we will find the functional variables that mostly influence pollen concentration in the future and second we will establish a model for prediction.

2 Variables and data

In order to address our objective, different variables have been registered daily from January the 1st, 1992 to June the 30th, 2003: Olea europaea l. airborne pollen, mean temperature, hours of sunshine, rainfall and wind speed among others.

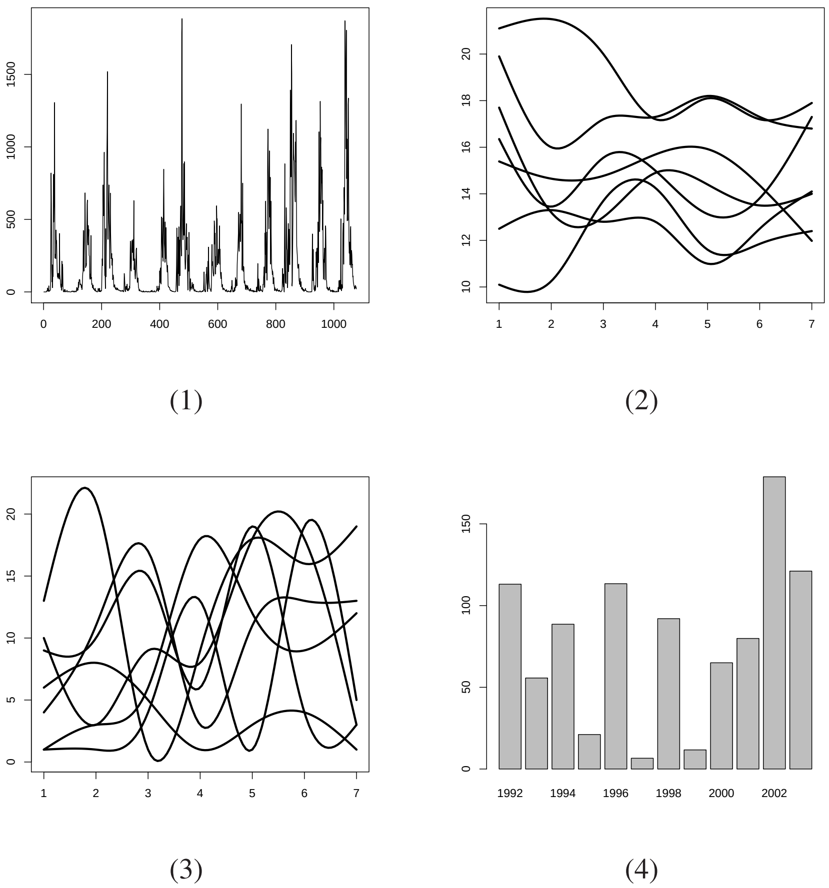

Figure 1: Olea europeae l. airborne pollen in the pollination period in all years (1), some curves of weekly mean temperature (2) and rainfall (3). Cumulated rainfall in the pollination season (4).

As addressed previously, the aim of the study is to predict high levels of pollen concentration from different functional variables. The period of the year when olive trees usually throw pollen into atmosphere is between 8 and 10 weeks in the months of May and June. Figure 1 shows these periods for the considered years. So we will use the functional data between April the 15th and July the 15 in each one of the years. The prediction of high levels of pollen concentration will be carried out from the time evolution of the different functional variables. We have considered after different proofs the week (7 days period) as the period in which the different functional variables influence the airborne olive pollen concentration. So the functional variables were considered as if they had sampled curves observations of

7 days. These curves for different variables can be seen in Figure 1

It is well known that the amount of precipitations cumulated before the pollination period can influence the amount of airborne olive pollen concentration so we will consider a nonfunctional variable in the model that cumulates in each year the precipitations registered from February 1st to April 15th (see Figure 1).

Finally the response is the binary variable that register peaks of pollen concentration so it takes value 1 for a week where we find at least one day with extreme concentration of olive pollen.

3 The model

Functional data are defined as variables whose observations are curves usually time dependent as the time evolution of temperatures, wind speed or rainfall. Different examples can be found in literature of these kind of data applied in Environmental Sciences as the examples of Ramsay and Silverman (2007), Escabias *et al.* (2005) or Aguilera *et al.* (2008) related to temperature and precipitations. The problem with functional data is that they can not be registered continuously and some methods to turn discrete data into functional data are necessary. One of the most used methods to do that is to consider that functional data belong to a space generated by a basis of functions $\{\phi_1(t),\ldots,\phi_p(t)\}$, so they are expressed in terms of these basic functions as $x(t)=\sum_{i=1}^{p} a_j\phi_j(t)$. The basis coefficients can be obtained from discrete observations $x(t_q)$ in different ways (smoothing, interpolation,...) depending on the origin and nature of data, that is, presence or absence of measurement error (see Escabias *et al.* (2004) for an explanation).

In order to predict events of extreme levels of olive pollen concentration, a mixed functional and non functional logit regression model is used to predict high level of pollen concentration. In order to formulate the model let Y be the binary response variable, $X(t)=(X_1(t),\ldots,X_k(t))'$ a vector of functional predictor variables and $U=(U_1,\ldots,U_l)'$ a vector of non-functional variables. The model we propose for predicting extreme levels of pollen concentration is

$$L=\alpha+\int_T X'(t)\beta(t)dt+U'\gamma \tag{1}$$

where $\beta(t)=(\beta_1(t),\ldots,\beta_k(t))'$ is a vector of functional parameters, $\gamma=(\gamma_1,\ldots,\gamma_l)'$ a vector of non functional parameter to estimate, and $L=(l_1,\ldots,l_n)'$ the vector of logit transformations $l_i=\ln\frac{\pi_i}{1-\pi_i}$, with π_i being the probability of a extreme levels of olive pollen concentration that for a specific observation of the functional variables $x_i(t)=(x_{i1}(t),\ldots,x_{ik}(t))'$ and nonfunctional variables $u_i=(u_{i1},\ldots,u_{il})'$ is expressed as

$$\pi_i=\frac{\exp\{\alpha+\int_T x_i'(t)\beta(t)dt+u_i'\gamma\}}{1+\exp\{\alpha+\int_T x_i'(t)\beta(t)dt+u_i'\gamma\}}$$

.

In order to model the relationship of airborne olive pollen concentration from functional and non functional variables, different models with different variable are compared. These comparatives let us find that we get better future predictions of extreme levels of pollen concentration by using both kind of variables in the model, functional and non functional, that by using them separately.

Acknowledgments. This research has been funded by project P06-FQM-01470 from *"Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía, Spain"* and project MTM2007-63793 from *Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain.*

References

[1] Aguilera, A. M., Escabias, M. and Valderrama, M. J. 2008. Forecasting binary longitudinal data by a functional PC-ARIMA model. *Computational Statistics and Data Analysis*, **52**(6), 3187-3197.

[2] Alba, F., Díaz de la Guardia, C. and Comtois, P. (2000). The effect of meteorological parameters on diurnal patterns of airborne olive pollen concentration. *Grana*, **39**, 200-208.

[3] Escabias, M., Aguilera, A. M. and Valderrama, M. J. 2004. Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, **16**(3-4), 365-384.

[4] Escabias, M., Aguilera, A. M. and Valderrama, M. J. 2005. Modelling environmental data by functional principal component logistic regression. *Environmetrics*, **16** (1), 95-107.

[5] Galán, C., Cariñanos, P., García-Mozo, H., Alcázar, P., and Domínguez-Vilches. (2001): Model for forecasting Olea europaea l. airborne pollen in south-west Andalusia, Spain. *International Journal of Biometeorology*, **45**, 5963.

[6] Ramsay, J. O. and Silverman, B. W. 2005. *Functional Data Analysis*. Second edition, Springer-Verlag: New York.

[7] Valderrama, M.J, Ocaña, F.A., Aguilera, A.M. and Ocaña-Peinado. F.M. (2010): Forecasting Pollen Concentration by a Two-Step Functional Model. *Biometrics*, in press