Sensitization to allergens and environmental features: a preliminary analysis to study their relation.

Silvia Fabris[1*], Giovanna Jona Lasinio[2] and Mario Santoro[3]

[1*] *University of Bergamo (IT); silvia.fabris@unibg.it*
[2] *DSS, University of Rome "Sapienza"; giovanna.jonalasinio@uniroma1.it*
[3] IAC-CNR, Rome; m.santoro@iac.cnr.it
[*] *Corresponding author*

Abstract. *Prevalence of allergic disease, in the last decades, increases in all countries, as much that The World Health Organization consider allergy a non-transmittable disease which is out of control. The most common allergies are determined by production of Immunoglobulin E (IgE), that can cause different disorders. Diagnosis started at the end of the nineteenth century by in vivo test and during the seventies IgE detection in blood had been introduced and it allowed to identify allergenic molecules; FABER test combines these two sources of information. The purpose of this study is to explore 16 allergens behavior on the province of Rome and to find the most appropriate model to define a possible relationship between sensitizations' occurrences and environmental features. In this case of study, we will take into account rainfall and minimum, maximum and average temperature recorded by ARPA Lazio.*

Keywords. *Log-Gaussian processes; Epidemiology; Spatio-temporal models, Bayesian methods.*

1 Introduction

Prevalence of allergic disease, in the last decades, increases in all countries, as much that The World Health Organization consider allergy a non-transmittable disease which is out of control. The most common allergies are determined by production of Immunoglobulin E (IgE), that can cause different disorders. Allergens are protein contained in allergenic sources; sensitization occurs when specific IgE are produced by atopic individuals and bind the trigger molecules [1]. Diagnosis started at the end of the nineteenth century, with the introduction in medicine of the first clinical allergy basic test: the *Skin Test* (*ST*), which have some limitation and it is not riskless. During the seventies IgE detection in blood had been introduced and it allowed to identify allergenic molecules. FABER test combines these two sources of information; first patients are tested with skin prick test that indirectly shows the presence of specific IgE, then direct IgE detection is made on serum sample by several *in vitro* method [2]. The purpose of this study is to explore 16 allergens behavior on the province of Rome and to find the most appropriate model to define a possible relationship between sensitizations' occurrences and environmental features. We consider eleven allergens belonging to plants, Ambrosia (Amb a 1),

Artemisia vulgaris (Art v 1), Betula pendula (Bet v 1), Birch, Hazel and Oak species (Cor a 1), Arizona Cypress (Cup a 1), Olea europaea (Ole e 1), Parietaria (Par j 2), Grasses (Phl p 1, Phl p 2 and Phl p 5), American Sycamore (Pla a 1), two due to cat (Fel d 1) and dog (Can f 1), two regarding house dust mite (Der p 1, Der p 2) and one coming from the Alternaria fungi (Alt a 1).

2 Data

We focus on 5523 clinical tests, collected between 2012 and 2017, from patients who live in the province of Rome. IgE values on the continuous scale, as they are recorded by the test, are not comparable between different allergens, for this reason, in order to evaluate the impact of each allergen on the population, data has been recoded into presence and absence of sensitization. Blood analysis reveal that 2032 patients do not show any positive response, 751 subjects recorded one sensitization and just one reveals to be affected by every allergic source. The most observed molecules are Cup a 1, Phl p 1, Der p 2, Fel d 1, the remaining allergens occur in less than 20% of cases, Art v 1, Pla a 1 and Amb a 1 are detected in less than 5% of the tests. The phenomenon act in a similar way about male and female separately, since the number of sensitizations is proportionally similar distributed between sexes.
Weather data has been collected from ARPA Lazio web portal: we considered available data, recorded by 26 meteorological stations, spread all over the province, collected from Spring 2012 to 31th December 2017, about minimum, maximum and average temperature and rainfall. Daily time series of those variables suggest constant trend and annual seasonality for all temperature's measurements. Precipitations, analyzed on the log-scale, do not show the same seasonal behavior, but again we have a constant trend. Moreover, average, minimum and maximum temperatures and rainfall show a stationary trend that does not differ much between stations too. Observations have been quarterly summarized by seasons, not considering periods where data are partially missing. For each season of each year of observation, mean, minimum and maximum value have been displayed, except for rainfall data, for which only the average amount has been considered. Dividing the new data into quartiles, the spatial behavior of temperature variables can be discussed: despite moderate changes between years of observation, the area of Rome and the cost present the higher temperature values, on the other hand, Castelli Romani, lake of Bracciano and the Nord - Est area are the coldest one.

3 Methods

Once sensitization and weather have been explored, we can go on studying the relation between them, modelling the phenomenon and estimating the parameters of interest with an MCMC algorithm. Having the exact place of residence of each patient, the finite pattern of point of positive response is easily represented on the entire province of interest: Figure 1 maps positive sensitizations of Fel d 1, Ole e 1 and Pla a 1, these allergens have been chosen because of their different impact on the observed sample.

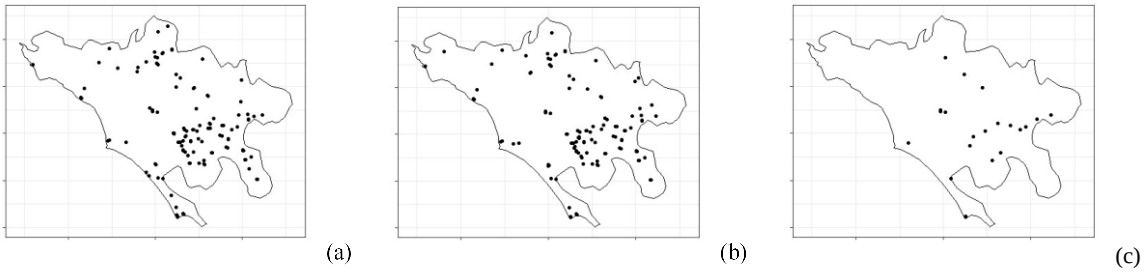

Fig. 1: Point Pattern of (a) Fel d 1, (b) Ole e 1, (c) Pla a 1 over the province of Rome.

Considering these locations as a random realization of a point pattern over a bounded window, fixed

and known, represented by the province of Rome, allergens phenomenon can be treated as a Poisson Process [3]. The phenomenon is clearly changing from one municipality to the other and many environmental phenomena may influence such occurrences; for this reason, we adopted an Inhomogeneous Poisson Process with intensity function varying in space. The class of models we are interested in is the Cox Process, in particular we used the log-Gaussian Cox Process [4]. In order to interpolate weather information all over the area of the Rome province, the time series needed to be predicted on every centroid of each municipality; we choose to model available spots using a Generalized Additive Model (GAM) [5]:

$$E(temperature) = \beta_0 + f(longitude, latitude) + f(time)$$
$$E(rain) = \beta_0 + f(longitude, latitude) + time\beta_1$$

For the purpose of the study, only prediction of the minimum of the minimum temperatures, the mean of the average temperatures and the maximum of the maximum temperatures and annually average rainfall have been used. Those have been summarized into one value for each centroid: first the median of average seasonal temperatures, minimum of minimum seasonal temperatures and maximum of maximum seasonal temperatures have been calculated, then these information have been transformed into their principal components. The first two components have been passed to the Cox process as covariates. Moreover, rainfall information have been summarized as the mean of the annual means. A grid, with square 2.3 X 2.3 km cells, overlaids the observation windows and covariates have been interpolated all over the grid by areal weight sum.

4 Results

Diagnostic results of the MCMC algorithm, concerning the considered allergenic, gave very good results, furthermore, inferential analysis of the estimated parameters finds out that meteorological features influence sensitization depending on molecules. Rain coefficient β is significantly different from 0 at 90% confidence interval just for Fel d 1: indeed the mean value of this coefficient shows a negative relation between allergens occurrence and rainfall, it means that each unit increase in rainfall lead to a reduction in relative risk with a mean of 3.373. Furthermore, the model can not explain any possible relation between those allergens that afflict less the population.

5 Further studies

The chosen protocol has several limitations, the main being the overly smoothed covariates added to the model. Furthermore, as far as vegetational allergens are concerned it would be of great interest to add information on the vegetation present in each municipality. Future developments will include a different interpolation of covariates and an ad hoc implementation of the Bayesian log-Gaussian Cox process.

References

[1] Alessandri C., Ferrara R., Bernardi M., Zennaro D., Tuppo L., Gianrieco I., Tamburrini M., Mari A., and Ciardiello M. (2017). Diagnosing allergic sensitization in the third millennium: why clinicians should know allergen molecule structures. *Clinical and translational allergy*, 7(1):21.

[2] Mari A. (2008). When does a protein become an allergen? Searching for a dynamic definition based on most advanced technology tools. *Clinical & Experimental Allergy*, 38(7):1089-1094

[3] Carlin, B. P., Gelfand, A. E., & Banerjee, S. (2014). *Hierarchical modeling and analysis for spatial data.*

Chapman and Hall/CRC.

[4] Taylor, B., Davies, T., Rowlingson, B., & Diggle, P. (2015). Bayesian inference and data augmentation schemes for spatial, spatiotemporal and multivariate log-Gaussian Cox processes in R. *Journal of Statistical Software*, 63, 1-48.

[5] Zuur, A. F. (2012). *A beginner's guide to generalized additive models with R*. Newburgh, NY, USA: Highland Statistics Limited.