

Modelling spatial variation in risk by generalized additive models¹

Modellizzazione della variazione spaziale del rischio tramite modelli additivi generalizzati

Alessio Pollice², Massimo Bilancia
Dipartimento di Scienze Statistiche,

Università degli Studi di Bari, e-mail: [apollice, mabil]@dss.uniba.it

Marina Musti, Domenica Cavone

Dipartimento di Medicina Interna e Medicina Pubblica,

Università degli Studi di Bari, e-mail: [m.musti, d.cavone]@medlav.uniba.it

Riassunto: In questa nota si utilizzano i modelli additivi generalizzati per ottenere la stima della distribuzione spaziale del rischio nell'ambito di uno studio caso-controllo sul mesotelioma pleurico nell'area metropolitana della città di Bari.

Keywords: Binary regression, Epidemiology, Generalized additive models, Kernel smoothing

1. Introduction and materials

The present study analyzes the spatial distribution of deaths from malignant mesothelioma due to non occupational exposure in the urban area of Bari. The aim is to verify whether there is evidence of a greater concentration of cases around a factory, consisting of an industrial settlement for asbestos cement production, that was active from 1934 to 1989. The study includes 48 cases of malignant mesothelioma with no apparent occupational exposure among residents in the municipality of Bari over the period 1992/2003, taken from the mesothelioma national registry - Apulia regional operative center, and 273 controls randomly selected from the municipal list of deaths. Information about sex, place and date of birth, place and date of death, and the complete residential history were collected for cases and controls. Metrical coordinates of the spatial locations of both the nearest (to the asbestos-cement factory) and the longest lasting residential addresses were obtained.

2. Methods and preliminary results

It is assumed that the residential locations of cases and controls are realizations of two independent inhomogeneous Poisson processes with intensities $\lambda_0(x) = b(1 - \pi(x))g(x)$ and $\lambda_1(x) = a\pi(x)g(x)$ respectively, where x is a vector describing location within the

¹Partially funded by PRIN'06 grant "Analisi e valutazione di rischi ambientali mediante modelli temporali, spaziali e spazio-temporali".

²Address of correspondence: Dipartimento di Scienze Statistiche "Carlo Cecchi", Università degli Studi di Bari, Via C. Rosalba 53, 70124, Bari

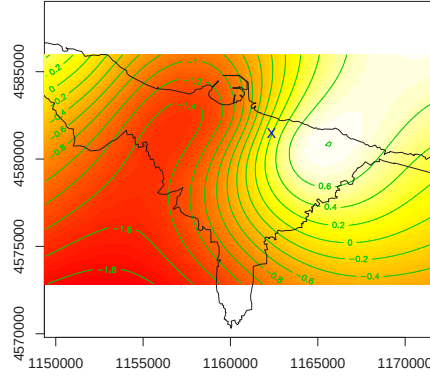


Figure 1: *Estimated log-odds surface adjusted for sex and age-class, obtained by using nearest addresses - the asbestos-cement factory location is marked by the special symbol ×.)*

study area, a and b are the sampling fractions, $\pi(x)$ is the probability that an individual located at x is a case and $g(x)$ is the background intensity function of a common generating inhomogeneous Poisson process (Diggle, 2003). Define a binary random variable Y_i to assume the value 1 or 0 according to whether the i -th event at x_i is a case or a control. Conditional on the superposition of the two Poisson processes the labels Y_i are independent with

$$P(Y_i = 1) = \lambda_1(x_i) / \{\lambda_1(x_i) + \lambda_0(x_i)\} \quad (1)$$

By substituting expressions for λ_1 and λ_0 into (1) we obtain

$$\log \{P(Y_i = 1)/P(Y_i = 0)\} \propto \log \{\pi(x_i) / [1 - \pi(x_i)]\} \quad (2)$$

Consideration of the smooth log-odds function $s(x) = \log \{\pi(x) / [1 - \pi(x)]\}$ describing the residual spatial variation after adjusting for covariate effects leads to the following semiparametric logistic regression model

$$\log \{P(Y_i = 1)/P(Y_i = 0)\} = \alpha + \beta z_i + s(x_i) \quad (3)$$

where α is an intercept term, z_i is a vector of known risk factors and β is a vector of regression parameters. The smooth function $s(x)$ can describe the dependence of the odds on the spatial location and on the distance from the putative source of risk, and it will be represented using penalized thin plate regression splines (Webster et al., 2006). The choice of the optimal value of the smoothing parameter is obtained by Generalized Cross Validation and Unbiased Risk Estimation (Wood, 2006; see Fig.1 for an example)

References

- Diggle P. (2003) *Statistical Analysis of Spatial Point Patterns*, Hodder Arnold, London.
 Webster T., Vieira V., Weinberg J. and Aschengrau A. (2006) Method for mapping population-based case-control studies: an application using generalized additive models, *International Journal of Health Geographics*, 5.
 Wood S. (2006) *Generalized Additive Models: an introduction with R*, CRC Press, London.