Archeologia e Calcolatori
8, 1997, 37-41

STATISTICAL MODELLING OF ARCHAEOLOGICAL SITES LOCATIONS AS NON-STATIONARY COX PROCESSES AND ITS USE IN FORECASTING

1. Introduction

The aim of this short paper is that of introducing a class of statistical instruments to modelling archaelogical sites locations (from now on simply ASL) in space and to predicting their probability of occurency. These models are based on collected statistical information and a priori knowledge.

In building up a statistical strategy to approach the problem, we encounter immediately a data problem. In fact ASL data are collected on a continuous space of coordinates whereas auxiliary information (such as information about the slope or exposure of the land) often derive from satellite data which are arranged on discrete, regular portions of space (the pixels of the computer image). When we implement forecasting models on a GIS (Arroyo-Bishop 1995), we technically distinguish between *vector* data (occurring in the first instance), and *raster* data (occurring in the second instance) (Arbia 1993).

For the reason described the problem of the forecasting ASL has been traditionally approached in two different ways. According to a first approach ASL is studied through data arranged on a continuous bidimensional (possibily three-dimensional) surface and discrete data related to auxiliary information insisting on the same area are smeared on the area itself. As opposed to this, a second approach is that of treating a study area as discrete one, and aggregate continuous information on ASL into small portions of land to match with auxiliary data.

In a previous paper (Arbia, Espa 1996) we discussed some extensions to models (proposed by Kvamme 1983; Kohler, Parker 1986; Warren 1990) on a discrete space. In the present paper we explore the other side of the moon by introducing a class of models aiming at producing probability maps of ASL defined on a continuous space.

2. A model for ASL forecasting on a continuous space

In this section we introduce a class of models for the description of ASL in the space. The formalism is derived from a model used by Rathbun, Cressie (1994) for the spatial diffusion of vegetation and by Arbia (1996) for the birth-growth-survival of firms in the spatial economy.

Let us start by defining A as the study area, $\mathbf{u} = [u_1, u_2]$ as the spatial co-ordinates of existing archaelogical sites, N(A) (or simply N) as the number of archaeological sites located in A, D_i (i=1,...,N) as a measure of the dimension of i-th archaelogical site (e. g. number of findings, physical extension or other variables), and d(**u**) as the infinitesimal region in the neighbourhood of point **u**.

Furthermore define the intensity of the process at point **u** as (DIGGLE 1983):

$$\lambda(\mathbf{u}) = \lim_{d(\mathbf{u})\to 0} E[N(d(\mathbf{u}))]/d(\mathbf{u})$$

which represents the expected number of points (archaelogical sites) in the infinitesimal area.

The model for the ASL is based on the assumption that the location of the existing archaeological sites is fixed and that the location of new archaelogical sites not yet found is a realization of a non-stationary Cox process (DIGGLE 1983) depending on the spatial interaction (positive or negative) with other existing sites, on the distance from important location or communication networks, and on a set of auxiliary variables.

More explicitly we can express the intensity of the process in the point of co-ordinates **u** as:

$$[1] \qquad \lambda(\mathbf{u}) = \exp\{\beta_0 + \beta_1\, d(\mathbf{u}) + \beta_2\, W(\mathbf{u}) + \beta_3\, X + \Phi(\mathbf{u})\}$$

with $\beta = \{\beta_0, \beta_1, \beta_2, \beta_3\}$ a set of parameters to be estimated, d(**u**) the distance of the archaeological site of coordinates **u** from main roads, communication networks or other significant points in space, and W(**u**) a term measuring the intensity of the interaction between the archaelogical site located in point **u** and the sites already found. In particular we can assume that:

$$[2] \qquad W(\mathbf{u}) = \Sigma_i \exp\{-r_i(\mathbf{u})\}$$

where the index i=1,..., N and $r_i(\mathbf{u})$ represents the distance between the archaelogical site of coordinates **u** and the i-th existing archaeological sites.

In Formula [1] X represents a vector of independent variables that are spatially heterogeneous (such as, for instance, slope of the ground, gradient, exposure, presence/absence of location conditions, etc.).

Finally $\Phi(\mathbf{u})$ is a random field which is assumed to be spatially stationary, Gaussian with zero mean. While the process $\Phi(\mathbf{u})$ is originally defined on a continuous grid of co-ordinates, in order to estimate the model we need to discretize it on a regular square grid of dimension h-by-k such that:

$$\Phi(u_{ij}),\ i = 1,\dots, h;\ j=1,\dots,k.$$

We can express the distribution of such a discrete set as being conditionally Gaussian (see HAINING 1990) such that:

$$D(\Phi(u_{ij} \mid \text{neighbours}) \sim N(\lambda\Sigma_p\Sigma_q w_{pq}\, \Phi(u_{pq});\ \tau^2).$$

The component $\Phi(u_{ij})$ is unobservable so that the model needs to be estimated through the EM algorithm for incomplete data (DEMPSTER, LAIRD, RUBIN 1977; ARBIA, ESPA 1996).

Once the model [1] is estimated on a test area it can be used to produce maps of the intensity $\lambda(\mathbf{u})$ in other areas provided we have available the vector of auxiliary information X, and we are able to specify a plausible form for the interaction between sites as expressed in Formula [2].

3. INTRODUCTION OF ARCHAEOLOGICAL SITE DIMENSION

It is also possible to introduce into discussion the information we have about the archaelogical site dimension. In fact it is plausible that the interaction between existing archaeological sites and candidate locations for new archaelogical sites is dependent on their dimension (D_i).

To introduce this let us divide the N existing archaeological sites into k classes. These can represent dimensional classes (e. g. k=3, small, medium size, large) as we will assume in the remainder of the paper, or, alternatively, definitorial classes (e. g. different categories, different historical period and so on). Define further $\mathbf{u}_{ik}$ as the coordinates of the i-th archaeological site belonging to the dimensional class k). Finally let D_{jk} be the dimension of the i-th archaeological site belonging to the dimensional class k (variously measured).

In this case the spatial interaction term β_2 W(u) in Formula [1] can be replaced by

$$\Sigma_j = {}_{1,3}\, \gamma_{ij}\, W_{ij}\, (u_{ik})$$

with $\{\gamma_{i1}, \gamma_{i2}, \gamma_{i3}\}$ a set of additional parameters to be estimated.

In this last expression $W_{ij}\,(u_{ik})$ is a measure of the intensity of spatial interaction (usually positive) of archaelogical site i belonging to the dimensional class k with its geographical neighbours belonging to the dimensional class j and can be expressed as (ARBIA 1996):

$$W_{ij}\,(u_{ik}) = \Sigma_{u_{jl} \in N(i,j,k)}\, |(u_{ik} - u_{jl}|^{-1}\, D_{ik}$$

where N(i,j,k) represents the set of neighbours of the i-th archaelogical site belonging to the dimensional class k and $|u_{ik} - u_{jl}|$ the distance between the i-th archaelogical site belonging to the dimensional class k and the j-th archaelogical site belonging to the dimensional class l.

4. SUMMARY AND CONCLUSIONS

This paper aims at introducing a class of testable statistical models to estimate the intensity of ASL viewed as realizations of non-stationary point processes.

The model relates the intensity of ASL in a point to the geographical distance with remarkable points or objects (such as water pool, rivers, communication networks etc.), to auxiliary information (such as for instance slope or exposure), and (since sites tend to cluster in space) to a positive interaction with existing archaeological sites.

These models can prove precious indeed in that they formalize some "common sense" knowledge and can help in producing automated probability maps on a continuous space to support digging decisions.

GIUSEPPE ARBIA
University "G. D'Annunzio" - Pescara
Department of Quantitative Methods and Economic Theory

GIUSEPPE ESPA
University of Trento
Institute of Statistics and Operational Research

BIBLIOGRAPHY

ARBIA G. 1993, *The use of GIS spatial statistical surveys,* «International Statistical Review», 61, 2, 339-359.

ARBIA G. 1996, *Modelling individual behaviour of firms in the spatial economy*, Quaderno "Ricerca di base", 1/96, Milano, Università Commerciale Bocconi.

ARBIA G., ESPA G. 1996, *Forecasting statistical models of archaeological site location*, in P. MOSCATI (ed.), *III Convegno Internazionale di Archeologia e Informatica (Roma, 22-25 novembre 1995)*, «Archeologia e Calcolatori», 7, 365-372.

ARROYO-BISHOP D. 1995, *MEDAIS: the potential contribution of GIS to the identification, study and conservation of the cultural heritage of the Mediterranean basin,* in *Proceedings of the First International Congress on: Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin (Catania-Siracusa, Italy, November 27 - December 2)*, 103.

DEMPSTER A.P., LAIRD N.M., RUBIN D.B. 1977, *Maximum likelihood from incomplete data via the EM algorithm,* «Journal of the Royal Statistical Society», B, 39, 1-38.

DIGGLE P.J. 1983, *Statistical Analysis of Spatial Point Patterns*, London, Academic Press.

HAINING R.P. 1990, *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, Cambridge University Press.

KOHLER T.A., PARKER S.C. 1986, *Predictive models for archaeological resource location*, in M. B. SCHIFFER (ed.), *Advances in Archaeological Method and Theory*, vol. 9, New York, Academic Press, 397-452.

KVAMME K.L. 1983, *Computer processing techniques for regional modeling of archaeological locations*, «Advances in Computer Archaeology», 1, 26-52.

RATHBUN S.L., CRESSIE N. 1994, *A space-time survival point process for a longleaf pine forest in Southern Georgia*, «Journal of the American Statistical Association», 89, 428, 1164-1174.

WARREN R.E. 1990, *Predictive modelling of archaeological site location: a case of study in the Midwest*, in K.M.S. ALLEN, S.W. GREEN, E.B.W. ZUBROW (eds.), *Interpreting Space: GIS and Archaeology*, New York, Taylor & Francis, 90-111.

ABSTRACT

The aim of this paper is to introduce a class of testable statistical models aiming at modelling archaeological sites locations (ASL) on a continuous space and at producing probability maps of ASL. These models are based on collected statistical information and

auxiliary information (such as information about the slope or exposure of the land, the topography, the hydrology, the topology, etc.). More explicitly the model for the ASL is based on the assumption that the location of the existing archaeological sites is fixed and that the location of new archaeological sites not yet found is a realization of a non-stationary point process depending on the spatial interaction with other existing sites, on the distance from important location or communication networks, and on a set of auxiliary variables. It is also possible to introduce into discussion the information we have about the archaeological site dimension. In fact it is plausible that the interaction between existing archaeological sites and candidate locations for new archaeological sites is dependent on their dimension. These models can prove precious indeed in that they formalize some "common sense" knowledge and can help in producing automated probability maps on a continuous space to support digging decisions.