Combining geostatistical and point process modelling

Janine B. Illian [1]

[1] School of Biological Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, Scotland UK, j.illian@abdn.ac.uk

Abstract: The spatial pattern of plants contains embedded information about the mechanisms controlling plant distribution and maintenance of species biodiversity. We aim to assess the relationship of tree distribution to environmental conditions while taking local interaction structures into account. To this end we combine a spatial point process model with a geostatistical model.

Keywords: point process modelling, model based geostatistics, biodiversity

1 Introduction

1.1 Motivation

There is a need to model multi-type spatial point patterns taking both the interaction structure and local environmental conditions into account. Environmental variables such as soil nutrient levels are typically continuous in space exhibiting spatial dependence but are measured discretely.

1.2 Approach

We jointly model the environmental variables' influence on the local point intensity and interactions between points in the pattern. We use a model based geostatistical approach to model the environmental variables. The random field derived from this is then used to as the intensity of a spatial point process with intra- and inter-type interaction.

2 The model

2.1 Univariate case

Initially, we consider the pattern formed by a single species, i.e. a single type point process. Let X defined in $S \subset R^2$ be a point process with the following density function:

$$f(x) \sim \prod_{\xi \in X} \phi(\xi) \prod_{(\xi,\eta) \subset X} \phi(\{\xi, \eta\}), \tag{1}$$

where $\phi()$ is a function relating the intensity of the point process to environmental properties, more specifically $\phi(\xi) = \exp(a_0 + a_1 Z_1(\xi) + a_2 Z_2(\xi) + \ldots + a_n Z_n(\xi))$ and $\phi(,)$ is a smooth interaction function.
The $Z_i = \{Z_i(x) : \ x \in S\}, i = 1, \ldots, n$ are spatial random fields, typically describing n environmental variables that exist in each x in the continuous spatial region S. Conditional on the $Z_i(.)$ the point process X is a Markov point process with interaction function $\phi(.,.)$.
More specifically, the $Z_i(.)$ are non-stationary Gaussian random fields with the mean-value specified by a linear model. I.e.

$$E[Z_i(x)] = \mu_i(x) = \sum_{l=1}^{q} \beta_{il} f_{il}(x),$$

$$Var[Z_i(x)] = \sigma^2$$

and correlation function

$$\rho(u) = Corr[Z_i(x), Z_i(x)].$$

Here, $u = \|x - x'\|$ is the Euclidean distance between two points in R^2.
In applications, the $Z_i(.)$ are not directly observable but have only been measured at a finite number of discrete locations $x_k, k = 1, \ldots, m$ typically different from the $\xi \in X$. We thus take on a geostatistical view here and define a measurement process $Y_i(.) = Y(x) : x \in S$ with measurements $Y_{i1}, \ldots, Y_{im}$ taken at $x_k, k = 1, \ldots, m$. I.e.

$$Y_i k = Z_i(x_k) + E_k, k = 1, \ldots, m,$$

where $E_1, \ldots, E_m$ are mutually independent, identically distributed with $E_k \sim N(0, \tau^2), k = 1, \ldots, m$.
We consider the Matern correlation function

$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi), \tag{2}$$

where $\kappa > 0$ and $\phi > 0$ are parameters and K_κ denotes a Bessel function of order κ.

2.2 Multivariate case

The model defined in (1) can be canonically generalised to the multivariate case as follows. Let $X_1, \ldots, X_p$ be defined in $S \subset R^2$ be a point process with the following density function:

$$f(x_1,\ldots,x_p) \sim \prod_{\xi_1 \in X_1} \cdots \prod_{\xi_p \in X_p} \phi_1(\xi_1)\ldots\phi_p(\xi_p) \prod_{1\le s<t\le p} \prod_{\xi_s \in x_s} \prod_{\xi_t \in x_t} \phi_{st}(\{\xi_s,\xi_t\}), \tag{3}$$

where the $\phi_k(), k = 1, \ldots, p$ a functions relating the intensity of the univariate point processes to environmental properties, as above, with $\phi_k(\xi) = \exp(a_{0k} + a_{1k}Z_1(\xi) + a_{2k}Z_2(\xi) + \ldots + a_{nk}Z_n(\xi))$. The $Z_i(.)$ are defined as above and are the same for all X_k.

3 Application

In an attempt to gain an understanding of the mechanisms that maintain species richness, Stephen Hubbell and his colleagues established a 50 ha plot on Barro Colorado Island (BCI) in Panama in the early 1980s recording the locations and sizes (diameter at breast height) of 235,349 individuals of 304 (rainforest) tree species in 1982. In addition, a large number of soil variables was collected. Since then, similar data have been repeatedly collected at regular times . Similar plots have been established in a network of 16 forest plots in several countries coordinated through the Center for Tropical Forest Science (CTFS) of the Smithsonian Tropical Research Institute in Panama (http://www.ctfs.si.edu).

The above models will be applied to data derived from the CTFS network.