Transactions on Ecology and the Environment vol 16, © 1997 WIT Press, www.witpress.com, ISSN 1743-3541

# A marked bivariate spatial model to detect interactions between Mediterranean shrubs

J. Mateu[1], J. L. Usó[1], F. Montes[2]
[1]*Departament de Matemàtiques. Universitat Jaume I. Campus Penyeta Roja, 12071 Castellón, Spain.*
[2]*Departament de Estadística e I.O. Facultad de Matemàtiques. Universitat de València, 46100 Burjassot, València, Spain.*

## Abstract

In many branches of science there is interest in the study and description of data which represent the locations of objects as points in an essentially planar region. In the present paper we assume that the objects are of two qualitatively distinguishable types and refer to the data as a bivariate spatial point pattern. A statistical analysis of such data should seek to provide some meaningful summary descriptions of the separate point patterns for each of the two types of objects, and of any possible inter-relationship between these two patterns. It is also considered a variable, an object characteristic, attached to each individual of both types. Such kind of variables are called marks. An illustration of above methodology is given using real data based on a two-type mediterranean shrub ecosystem in which the attached marks represent the total biomass of each shrub. The chosen shrub species characterize a post-fire mediterranean ecosystem.

## 1 Introduction

Statistical analysis of stands of trees as a whole need suitable methods of spatial statistics. Obviously, trees within a stand affect development and survival of their neighbours. Thus, it is not a priori acceptable to treat them and their measured characteristics as independent random variates; this would lead to considerable errors in further statistical reasoning if there are indeed correlations. A suitable mathematical model is a marked point process where the points are tree positions given with respect to a Cartesian coordinate system. The marks are qualitative or quantitative tree characteristics. The theory of marked point processes offers various characteristics for the quantitative description of the interaction of individuals (trees in a forest, say). One of them is the *pair-correlation function*, which is in some sense analogous to the correlation function of a stochastic process. It characterizes the variability of the system of tree locations. The correlation of the marks are described by the *mark correlation function.* This function characterizes relations between the marks of different points conditional on point distances.

These functions, pair correlation function and mark correlation function, are of value at least for three important problems of forest statistics: (a) For exploratory analysis of the interaction among trees (e.g., type of interaction, determination of the range of interaction and estimation of its strength); (b) For modelling and simulation of forests to ensure realistic assumptions about tree variates; (c) For determining the accuracy of estimators of global characteristics of tree stands in forest inventory (such as stand density or total basal area).

A bivariate spatial point pattern is one in which the events are of two distinguishable types. Fig.1 shows a bivariate spatial pattern consisting of the locations of two types of mediterranean shrubs, *Ulex Parviflorus* and *Cistus Clusii*, in a 10x10 $m^2$ square region. A quantitative description of such patterns is of interest to discriminate between two competing hypotheses: (H1) the two components are formed initially in two separate layers; (H2) trees are formed initially in a single layer, differentiation into *Ulex* and *Cistus* occurring at a later stage of development. One implication of H1 is that the two component patterns are statistically independent, whilst under H2 they are generally dependent, even if the differentiation is completely random. One possible benchmark hypothesis for the assessment of interactions between the two types is that the two component patterns are determined independently of one another, as would be appropriate under hypothesis H1. If, on the other hand, a bivariate pattern arises through some form of labelling mechanism, as under H2, the question of whether or not the component patterns are statistically independent is not relevant. A more natural benchmark hypothesis is that the two component patterns are formed by random labelling, by which we mean that events are labelled independently, each event being labelled type 1 with probability p, and type 2 with probability 1-p. Independence and random labelling are in general statistically distinct hypotheses: they coincide if and only if the superposition of the two types of event forms a completely random pattern, in which case both component patterns are also completely random.

Transactions on Ecology and the Environment vol 16, © 1997 WIT Press, www.witpress.com, ISSN 1743-3541



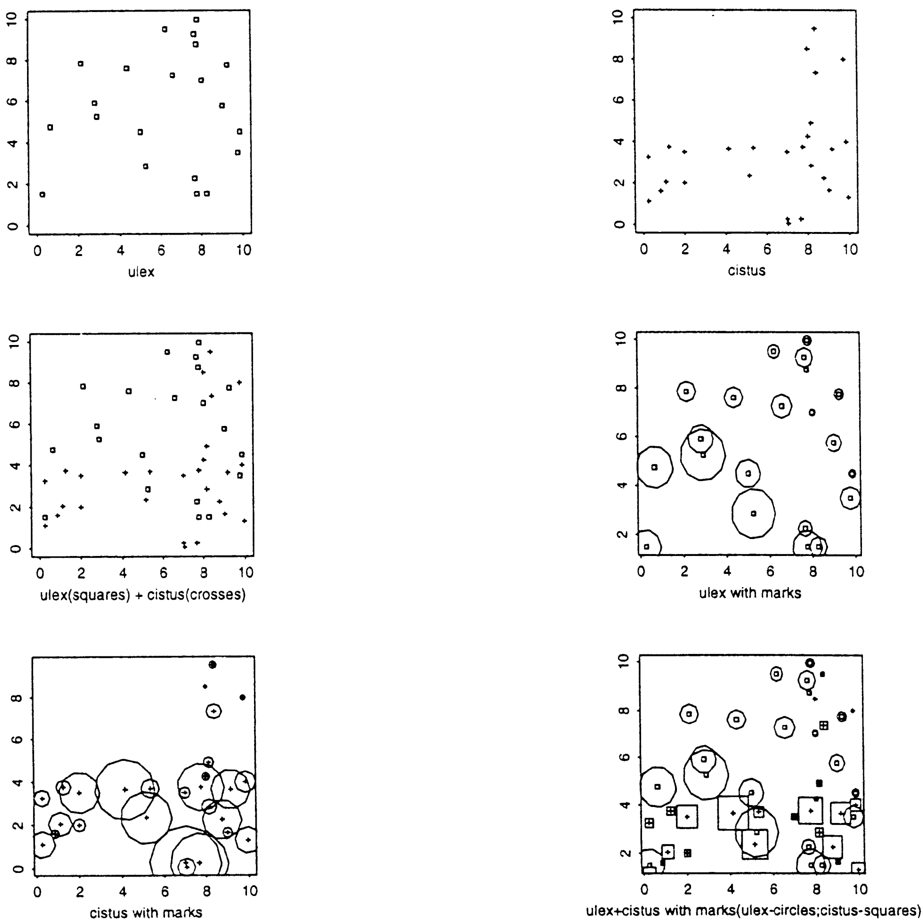


**Figure 1:** Mediterranean shrubs with biomass attached: *Ulex* (squares, n=21), *Cistus* (crosses, n=27) and biomass poltted with circles and/or squares.

## 2 Statistical methods

It is assumed here that the random system of marked points $[s, Z(s)]$ is homogeneous and isotropic. Roughly speaking, this means that there are no systematic fluctuations of point density and no preferred directions in the point pattern. Furthermore, it is assumed that the marks are positive numbers. Fundamental characteristics of a homogeneous marked point process are the density $\lambda_0$ and the mean mark $\mu_z$: $\lambda_0$ is the mean number of points per unit area without considering marks and $\mu_z$ is the mean of the marks Z(s). An example of mean mark is the mean diameter. Both of these characteristics are widely applied in theoretical and practical forestry. They are called *first-order characteristics*. This word is used since these characteristics regard one individual only. When describing variability and correlations in individual stands we have to consider pairs of individuals. The corresponding characteristics are therefore called *second-order characteristics*. Let us consider two infinitesimally small circles of areas ds and dt of intercenter distance. Let P(h) denote the probability that both circles each contain a point of the point process. We can write the following formula in terms of the product density $l_{N_0}(h)$, where $N_0$ is the point process without marks,

$$P(h) = \lambda^2{}_0 g(h) ds dt = l_{N_0}(h) ds dt.$$

The function g(h) is called *pair correlation function* and is a function of the interpoint distance h. For a completely random point process (i.e., a homogeneous Poisson process) g(h)=1. Values of the pair correlation function larger than 1 indicate that the interpoint distances around h are relatively more frequent compared to those in a complete random point process, and conversely, values of g(h) smaller than 1 indicate that the corresponding distances are rare. The pair correlation function can take all values between zero and infinity, for large h it tends to one.

A cumulative second-order characteristic is

$$K(h) = \int_0^h g(u) 2\pi u du, \ h > 0$$

called K-function (or Ripley's K) (Stoyan [4,5]). It has the following intuitive interpretation: $\lambda_0$K(h) is the mean number of further points within a distance no more than h from a randomly chosen point. For a homogeneous Poisson process $K(h) = \pi h^2$. A common practice is to use a transformed version of the K-function $L(h) = \sqrt{K(h)/\pi}$ . For a Poisson process L(h)=h, a straight line with a slope of 1. In spatial statistics, the pair correlation function is often used for exploratory data analysis and Ripley's K-function or L-function for statistical testing.

Let us now also take the marks into account. We consider again the two infinitesimal circles at a distance h from each other. Let MP(h) be the mean of

Transactions on Ecology and the Environment vol 16, © 1997 WIT Press, www.witpress.com, ISSN 1743-3541



the product of the marks of the points in them; this product vanishes unless each circle contains a point. MP(h) is written, in terms of the product density $l_m(.)$ in the form

$$l_m(h)dsdt = MP(h) = \lambda^2{}_0 g(h) K_{zz}(h) dsdt.$$

The new function $K_{zz}(h)$ describes relations between the marks. Values of $K_{zz}(h)$ larger than $\mu_z^2$ show that the marks of both individuals of a point pair at distance h tend to vary together in the same direction indicating positive correlation. The value of $K_{zz}(h)$ has the following interpretation: it is the expectation of Z(s)·Z(t) given that there are two points, one at some location s and another at some location t where the Euclidean distance $\|s-t\| = h$. The function $K_{zz}(h)$ can be scaled in order to make interpretation easier. In the special case when the marks are not correlated (it is then said that the marked point process is independently marked or randomly labelled) $K_{zz}(h)$ is identical to $\mu_z^2$ for all h. Therefore it is suggested the use of $MC_{zz}(h) = K_{zz}(h) / \mu^2{}_z$, the scaled counterpart of $K_{zz}(h)$. We call $MC_{zz}(h)$ the *mark correlation function*. The meaning of $MC_{zz}(h)$ is as follows: Small values of it (small compared to the value one) suggest negative correlation (mutual inhibition) between the marks at a distance of size h, large values indicate positive correlation (mutual attraction) at this distance. In the case where the mark variable is uncorrelated $MC_{zz}(h)=1$.

For the description of the interaction between two species the functions $g_{12}(h)$ and $L_{12}(h)$, analogues of the pair correlation function and the L-function, have to be applied. Now we have type 1 points and type 2 points with densities $\lambda_{01}$ and $\lambda_{02}$, respectively. Let us consider two infinitesimally small circles of areas ds and dt of intercenter distance h. If $P_{12}(h)$ is the probability that both circles each contain a point, one of them is of type 1 and the other one of type 2, then

$$P_{12}(h) = \lambda_{01}\lambda_{02} g_{12}(h) dsdt$$

The $L_{12}$(h), a transformed version of the $K_{12}(h)$ function, is obtained in a similar way as the L-function, the exceptions being that the g(h) function is replaced by the $g_{12}(h)$ function and $\lambda_{01}\lambda_{02}$ is used instead of $\lambda_0^2$ in scaling. A complete description of the second-order properties of the process requires us to define the bivariate K-functions $K_{ij}(h)$ by,

$\lambda_j K_{ij}(h)$ =mean number of type j events within distance h of an arbitrary type i event.

Note that $K_{11}(h)$ and $K_{22}(h)$ correspond to K(h) as previously defined, also that $K_{12}(h) = K_{21}(h)$.

The statistical testing of the independence hypothesis is done using a Monte Carlo test conditioned by the marginal point patterns, see Diggle [1,2]. The idea of the test is that, say, the points of type 1 are kept fixed and random shifts to the pattern of points of type 2 are performed. The $L_{12}$(h) functions are estimated from the shifted two-species patterns and the upper and lower envelopes are formed using repetitions.

Under independence, $K_{12}(h) = \pi h^2$, irrespective of the forms of $K_{11}(h)$ and $K_{22}(h)$. Finally, under random labelling, $K_{11}(h) = K_{22}(h) = K_{12}(h)$.

In order to study correlation between species with regard to some mark variable an analog of the mark correlation function is introduced. Again pairs of individuals are to be considered where one member belongs to type 1 points and the other one to type 2 points. To be more precise, let us consider two infinitesimally small circles of areas ds and dt with intercenter distance h and let $MP_{12}(h)$ be the mean of the product of the marks of the points in them: one of them is of type 1 and the other one of type 2. Then we define a new function ${K^{12}}_{zz}(h)$ through

$$MP_{12}(h) = \lambda_{01}\lambda_{02}\, g_{12}(h)\, {K^{12}}_{zz}(h)\, ds\, dt$$

The scaled version ${MC^{12}}_{zz}(h)$ is obtained in a similar way to the $MC_{zz}(h)$ function through dividing ${K^{12}}_{zz}(h)$ by $\mu_{z1}\mu_{z2}$, where $\mu_{z1}$ and $\mu_{z2}$ are the mean marks of type 1 and type 2 points respectively. ${MC^{12}}_{zz}(h)$ will be called *bivariate mark correlation function.*

## 2.1 Statistical estimation of characteristics from mapped data

Let us assume that we have a complete map of shrubs in a region A with shrub variable Z measured, and these observations are $[s_1, Z(s_1)], \ldots, [s_n, Z(s_n)]$ . Following Stoyan [4,5], the estimators for the pair correlation and the mark correlation function are respectively

$$\hat{g}(h) = \sum_{i \neq j} w(\|s_i - s_j\| - h) / (\hat{\lambda}^2 2\pi h s(h)), \quad r > 0$$

and

$$\hat{MC}_{zz}(h) = \sum_{i \neq j} w(\|s_i - s_j\| - h) Z(s_i) Z(s_j) / (\hat{\lambda}^2 \hat{\mu}^2 2\pi h s(h) \hat{g}(h)), \quad r > 0$$

where $\hat{\lambda}$ is the mean number of shrubs per unit area, $\hat{\mu}$ is the estimated mean mark, s(r) is an edge-correction factor (see Ohser [3]) and w(.) is the well-known Epanechnikov kernel.

The estimates of K-function and $L_{12}$(h) can be found in Diggle [1]. The estimates of the bivariate versions of the pair and mark correlation functions, $g_{12}(.)$ and ${MC^{12}}_{zz}(.)$ are the natural generalizations of the corresponding estimates for the univariate case shown above.

Transactions on Ecology and the Environment vol 16, © 1997 WIT Press, www.witpress.com, ISSN 1743-3541



# 3 Data

The data set consists of a sample area of a two-species of mediterranean shrubs: the species are *Ulex Parviflorus* and *Cistus Clusii*. The stand is situated at a height of 700 metres close to the see in Castellon (Spain). 21 *Ulex* shrubs and 27 *Cistus* shrubs have been measured in a planar squared region of 10x10 $m^2$. One shrub variable has been determined: the total (green and woody) biomass has been measured following Usó et al. [6]. These data allow us to give an example of studying spatial interspecies interaction of shrubs characteristics. The data are illustrated in Fig.1.

# 4 Results and conclusions

Let us first describe the point pattern of shrubs ignoring the biomass marks. Th estimates of g(.) and L(.)-h are shown in Fig.2 for both species, *Ulex* and *Cistus*. The pair correlation function for *Ulex* indicate that the point pattern is completely random as g(h) is around 1 for the considered distances from 0.1 to 4 metres. The fluctuations of the g(.) function around 1 are considered as random fluctuations. The estimate of the L(.)-h function lies between both envelopes indicating again that *Ulex* constitute a sample of a random pattern. The L(.)-h function for *Cistus* does not lie between the envelopes deducing that the observed point pattern cannot be considered as a sample from a Poisson process. The g(.9 function suggests that a strong small-scale tendency towards clustering (from 0 to 2 metres) exists.

Let us now look at the correlation of marks. In Fig.2 it is also plotted the mark correlation function $MC_{zz}(h)$. Correlation amongst *Ulex* with respect to biomass is strongly negative from 0.5 to 2 metres. Ulex shrubs close together appear to inhibit one another. On the other hand, correlation amongst *Cistus* shrubs with respect to biomass is strongly positive from 0 to 0.5 metres and from 2.5 to 4 metres. Cistus shrubs very close together attract one another and also those far appart.

Both traces for the data lie between envelopes from 99 simulations under random labelling and independence (Fig.3) concluding that the bivariate pattern of *Ulex* and *Cistus* is compatible with the hypothesis of random labelling and independence.

The estimates of the bivariate pair and mark correlation functions are shown in Fig.3. Both species attract each other when they are very close together (in distances from 0 to 0.5) and inhibit each other in distances over 3 metres. Taking the marks into account, the bivariate mark correlation fiınction is considered to fluctuate randomly around 1 indicating that there is no correlation between species with respect to biomass.

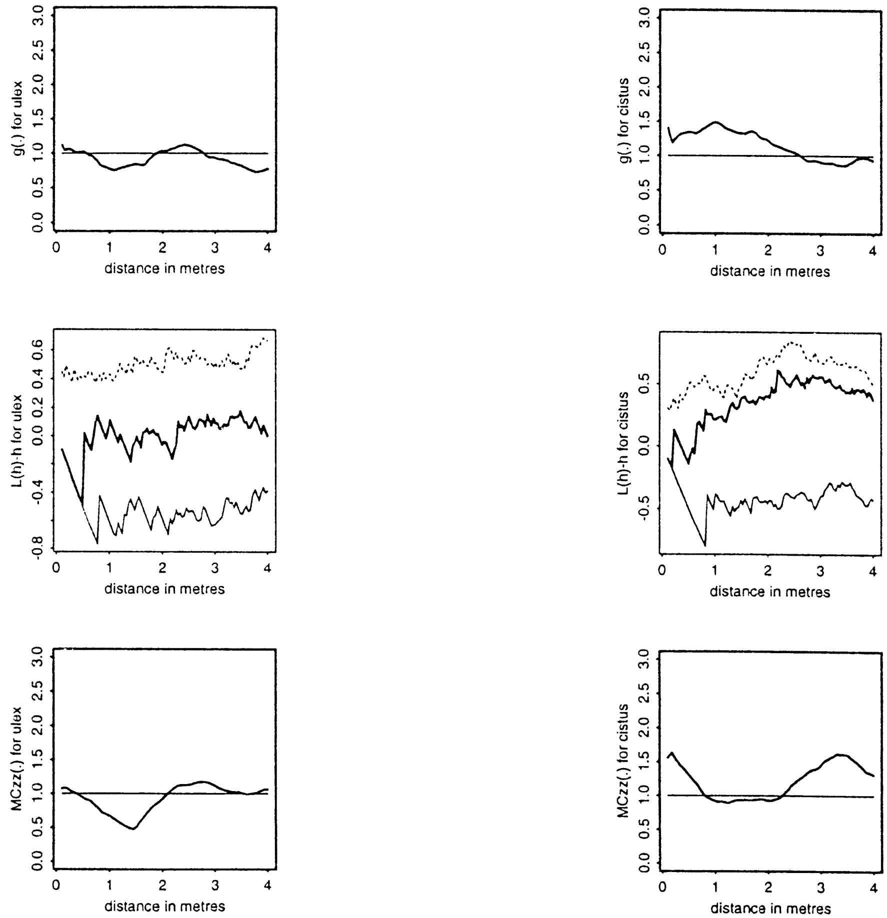


**Figure 2:** Kernel estimates of the pair and mark correlation functions for *Ulex* and *Cistus*. Estimates of the L function with envelopes from 99 simulations.

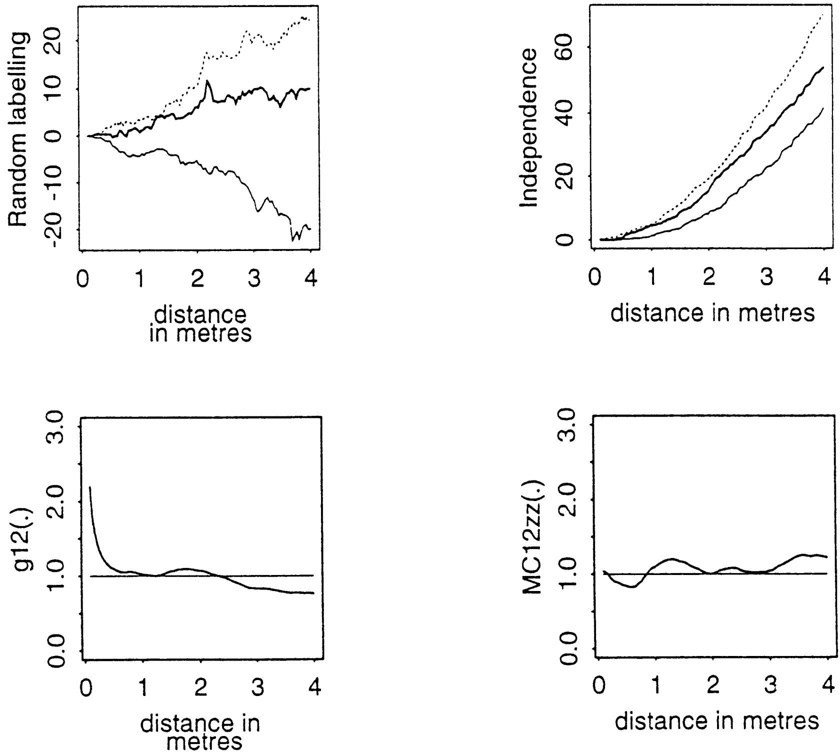


**Figure 3:** Trace of data together with envelopes from 99 simulations to study random lavelling and independence. Kernel estimates of the bivariate pair and mark correlation functions for *Ulex* and *Cistus*.

## References


1. Diggle, P.J., *Statistical Analysis of Spatial Point Patterns*. London, Academic Press, 1983.
2. Diggle, P.J., Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *Journal of the Neuroscience Methods*, 1986, **18**,115-25.
3. Osher, J., On estimators for the reduced second moment measure of point processes, Math. Oper. Statist., Ser. Statist., 1983, **14**, 63-71.
4. Stoyan, D., Statistical analysis of spatial point processes: A soft-core model and cross-correlations of marks, *Biometric Journal*, 1987, **29**, 971-980.
5. Stoyan, D., Kendall, W.S. & Mecke, J., *Stochastic Geometry and its Applications*, Wiley, New York, 1987.
6. Usó, J.L., Mateu, J., Karjalainen, T. & Salvador, P., Allometric Regression equations to determine aerial biomasses of Mediterranean shrubs, *Plant Ecology*, 1997 (in press).