



Research Report
Department of Statistics
Göteborg University
Sweden

Multitype spatial point patterns with hierarchical interactions

Aila Särkkä
Harri Högmänder

Research Report 1998:1
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Department of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se
Göteborg University	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
Box 660			
SE 405 30 Göteborg			
Sweden			

MULTITYPE SPATIAL POINT PATTERNS WITH HIERARCHICAL INTERACTIONS

Harri Högmander¹

University of Jyväskylä, Finland

and

Aila Särkkä²

University of Gothenburg, Sweden

December 19, 1997

Abstract

Multitype spatial point patterns with hierarchical interactions are considered. Here hierarchical interaction means directionality: Points on a higher level of hierarchy affect the locations of points on the lower levels, but not vice versa. Such relations are common, for example, in ecological communities.

Interacting point patterns are often modelled by Gibbs processes with pairwise interactions. However, these models are inherently symmetric, and the hierarchy can be acknowledged only when interpreting the results. We suggest the following trick allowing the inclusion of the hierarchical structure in the model. Instead of regarding the pattern as a realisation of a stationary multivariate point process, we build the pattern one type at a time according to the order of the hierarchy by using non-stationary univariate processes. As interactions connected to points \mathbf{x} on a certain level are considered, the effect of the higher levels is interpreted as heterogeneity of the pattern \mathbf{x} and the points on the lower levels are neglected owing to the hierarchical structure.

Key words: Hierarchical interaction; Multitype point patterns; Non-stationary Gibbs point processes; Maximum likelihood estimation; Maximum pseudolikelihood estimation.

1 Introduction

We consider multitype spatial point patterns with interactions both within and between the types. Furthermore, we assume that the interaction between

¹Address: University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FIN-40351 Jyväskylä, Finland. E-mail: hogmande@jyu.fi.

²E-mail: statasar@bones.stat.gu.se.

the types is hierarchical: The points on a higher level of hierarchy affect the locations of the points on the lower levels, but not vice versa. For example, let us think of nest sites of two territorial animal species, one large and another small. If the smaller ones avoid the individuals of the large species but the large choose their nest sites independently of the smaller ones, we have a hierarchical interaction described above. Such relationships are common, for instance, in ecological communities.

Some work on hierarchical interactions in spatial context does exist. Ford and Diggle (1981) model interaction between neighbouring plants as a spatial process taking the plant height into account. The interaction is regarded one-sided so that tall plants suppress small plants. Renshaw (1984) studies the same data by using two-dimensional spectral analysis. He has an example of a one-sided competition process. Mugglestone and Renshaw (1996) define cross-spectral functions, called gain spectra, which can be used if a causal relationship between the components of a bivariate process is suspected.

We would like to model interactions within and between types by applying point processes. A natural model for patterns with interacting objects is a Gibbs point process (see e.g. Diggle, 1983, or Stoyan *et al.*, 1995). Each type of association can be described using an interaction function, usually parametric, and in order to find out the strength and the range of interaction—which are the ordinary targets—the respective unknown parameters are estimated. Gibbs models, however, assume symmetric interactions: If a point x affects the location of another point y , then also y equally affects x . The hierarchy can be taken into account only when interpreting the obtained results, not already in the modelling phase. In practice this means, for example, that the observed repulsion between two species is thought being due to the tendency of the lower level points to avoid the dominating points.

Though we did not see how to include hierarchy into a Gibbs model, we can use Gibbs models as a basis to build a hierarchical system. Instead of regarding the pattern as a realisation of a stationary multivariate Gibbs process, we suggest to model the subpatterns type by type as realisations of non-stationary univariate processes. As interactions connected to points of a certain level are considered, the effect of the higher levels is interpreted as heterogeneity of this subpattern while the effect of the points on the lower level is neglected since they are not supposed to affect the higher levels. On the highest level the pattern is assumed to be a realisation of a stationary univariate Gibbs process.

The paper is organised as follows. In Chapter 2 we first recall the multivariate Gibbs point process model and then construct a new model for hierarchical systems. Interaction parameters for a hierarchical model can be estimated by using means developed for Gibbs point processes. The maximum likelihood (ML) method and the maximum pseudolikelihood (MPL) method are discussed in Chapter 3. We made a simulation study with hierarchical Strauss models to compare the performance of the two approaches, the traditional symmetric Gibbs model and our hierarchical model. Both the ML and the MPL estimation methods are applied. The results are presented in Chapter 4. In Chapter 5 we assess the interactions of a two species ant community (cf. Harkness and Isham, 1983) by modelling their nest site patterns according to a hierarchical model.

2 Two models for interaction

Let us assume that in a bivariate pattern the points of type A are completely indifferent to the points of type B , but that the locations of the A points affect the locations of the B points. We are interested in estimating the strength and range of each kind of interaction within and between the types.

2.1 Pairwise interaction model

Multivariate Gibbs point processes are a basic tool in modelling patterns which consist of several types of interacting points. Characteristic for Gibbs processes is that points interact mutually only if they are “neighbours”. The neighbourhood relation can be defined in several ways; we assume that the interaction is due to the locations and types of the points. The neighbourhood relation is symmetric: If a point x is a neighbour to another point y , then also y is a neighbour to x .

When introducing Gibbs processes it is necessary to make a distinction between a finite point process defined in a bounded subset of \mathbf{R}^2 and an infinite point process in \mathbf{R}^2 . Finite processes are simpler but they can not be stationary due to edge effects. We consider point processes defined on a torus W (a bounded rectangle in \mathbf{R}^2 with opposite edges identified). Hence we have both the simplicity of finite point processes and the stationarity of the infinite ones.

We restrict ourselves to stationary and isotropic bivariate point processes with pairwise interactions. Generalisation to k types of points is obvious. The density of a bivariate pairwise interaction process (with respect to the corresponding measure P) in W is given by

$$f(\mathbf{x}) = f(\mathbf{x}_A \cup \mathbf{x}_B) = \tag{1}$$

$$\frac{1}{Z_{AB}} b_A^{n(A)} b_B^{n(B)} \prod_{\substack{x_i, x_j \in \mathbf{x}_A \\ i < j}} h_A(x_i, x_j) \prod_{\substack{x_i, x_j \in \mathbf{x}_B \\ i < j}} h_B(x_i, x_j) \prod_{\substack{x_i \in \mathbf{x}_A \\ x_j \in \mathbf{x}_B}} h_{AB}(x_i, x_j),$$

where the simplified notation $\mathbf{x} = \mathbf{x}_A \cup \mathbf{x}_B$ stands for a realisation of the point process, \mathbf{x}_A and \mathbf{x}_B being the subconfigurations formed by the points of type A and B , respectively, and

$$Z_{AB} = \tag{2}$$

$$\int b_A^{n(A)} b_B^{n(B)} \prod_{\substack{x_i, x_j \in \mathbf{x}_A \\ i < j}} h_A(x_i, x_j) \prod_{\substack{x_i, x_j \in \mathbf{x}_B \\ i < j}} h_B(x_i, x_j) \prod_{\substack{x_i \in \mathbf{x}_A \\ x_j \in \mathbf{x}_B}} h_{AB}(x_i, x_j) P(d\mathbf{x})$$

is the scaling factor. Furthermore, the parameters b_A and b_B control the intensity of the subprocesses. The numbers of points in the subpatterns are denoted by $n(A)$ and $n(B)$. Last, h_A , h_B and h_{AB} are the non-negative interaction functions within types A and B and between the types. In a stationary and isotropic case the interaction function is a function of the distance between the points alone, $h_i(x_i, x_j) = h_i(d(x_i, x_j))$. The interaction functions are typically of a parametric form.

The conditional intensity is a useful tool when considering how adding a new point y to a pattern \mathbf{x} affects the probability density function. In case of the symmetric model the conditional intensity can be written as

$$\lambda(y; \mathbf{x}) = \frac{f(y \cup \mathbf{x})}{f(\mathbf{x})} \tag{3}$$

$$= \begin{cases} b_A \prod_{x_i \in \mathbf{x}_A} h_A(y, x_i) \prod_{x_i \in \mathbf{x}_B} h_{AB}(y, x_i), & \text{if } y \text{ is of type } A, \\ b_B \prod_{x_i \in \mathbf{x}_B} h_B(y, x_i) \prod_{x_i \in \mathbf{x}_A} h_{AB}(y, x_i), & \text{if } y \text{ is of type } B. \end{cases}$$

2.2 Hierarchical model

Instead of modelling all interactions h_A , h_B and h_{AB} simultaneously by having a bivariate pairwise interaction model we suggest modelling one point type in turn according to the hierarchy. More specifically, we advocate to use a homogeneous univariate model for the higher level of hierarchy, and for the lower level a heterogeneous univariate model, where the heterogeneity is due to the points on the dominating level. In case of three or more types the lower level patterns are heterogeneous owing to all the superior levels.

The density function corresponding to the hierarchical approach can be composed as follows. First, for the higher level process the density is

$$f(\mathbf{x}_A) = \frac{1}{Z_A} b_A^{n(A)} \prod_{\substack{x_i, x_j \in \mathbf{x}_A \\ i < j}} h_A(x_i, x_j), \quad (4)$$

where

$$Z_A = \int b_A^{n(A)} \prod_{\substack{x_i, x_j \in \mathbf{x}_A \\ i < j}} h_A(x_i, x_j) P(d\mathbf{x}_A).$$

The conditional intensity is

$$\lambda_A(y; \mathbf{x}) = b_A \prod_{x_i \in \mathbf{x}_A} h_A(y, x_i) \quad (5)$$

(cf. conditional intensity (3) above).

Second, given the dominating pattern \mathbf{x}_A , the density for the lower level process is

$$f_{\mathbf{x}_A}(\mathbf{x}_B) = \frac{1}{Z_B(\mathbf{x}_A)} b_B^{n(B)} \prod_{\substack{x_i, x_j \in \mathbf{x}_B \\ i < j}} h_B(x_i, x_j) \prod_{\substack{x_i \in \mathbf{x}_A \\ x_j \in \mathbf{x}_B}} h_{AB}(x_i, x_j), \quad (6)$$

where

$$Z_B(\mathbf{x}_A) = \int b_B^{n(B)} \prod_{\substack{x_i, x_j \in \mathbf{x}_B \\ i < j}} h_B(x_i, x_j) \prod_{\substack{x_i \in \mathbf{x}_A \\ x_j \in \mathbf{x}_B}} h_{AB}(x_i, x_j) P(d\mathbf{x}_B).$$

The conditional intensity is

$$\lambda_B(y; \mathbf{x}) = b_B \prod_{x_i \in \mathbf{x}_B} h_B(y, x_i) \prod_{x_i \in \mathbf{x}_A} h_{AB}(y, x_i),$$

which is equal to the corresponding conditional density of (3).

According to densities (4) and (6) the “joint” density of the whole process can be written as

$$\begin{aligned} f(\mathbf{x}_A, \mathbf{x}_B) &= f(\mathbf{x}_A) f_{\mathbf{x}_A}(\mathbf{x}_B) = \\ & \frac{b_A^{n(A)} b_B^{n(B)}}{Z_A Z_B(\mathbf{x}_A)} \prod_{\substack{x_i, x_j \in \mathbf{x}_A \\ i < j}} h_A(x_i, x_j) \prod_{\substack{x_i, x_j \in \mathbf{x}_B \\ i < j}} h_B(x_i, x_j) \prod_{\substack{x_i \in \mathbf{x}_A \\ x_j \in \mathbf{x}_B}} h_{AB}(x_i, x_j). \end{aligned} \quad (7)$$

Notice that the product $Z_A Z_B(\mathbf{x}_A)$ above is not equal to Z_{AB} in (2).

An extremely hierarchical model, that is, each point representing a different type, corresponds to a sequential inhibition process, where the order of generation of points follows the hierarchy of the model. Hence such a process can be thought as a special case of the present hierarchical model.

2.3 Comparison of the two models

The fundamental difference between the symmetric and hierarchical models for multitype point patterns is in the interpretation of the processes generating the pattern. In the hierarchical model the process consists of two (or k , in general) consecutive subprocesses, one for each type of points, while the symmetric approach is based on one bivariate (or multivariate) process, where the point types affect each other.

The difference between the models is not just philosophical, which can be seen by comparing the density functions (1) and (7), or the conditional intensities (3) and (5). The densities are equal up to the scaling factor. Notice that in the hierarchical case the integral with respect to \mathbf{x}_B is calculated by having one fixed realisation \mathbf{x}_A in the integrand. Hence, the two methods give, indeed, two different distributions.

We emphasise that the assumption of hierarchy is based on prior knowledge concerning the structure of the pattern. It may not be possible to deduce the hierarchy from a single point pattern *a posteriori*, and, even more confusing, essentially different models may show a good fit to the data.

3 Estimation of interactions

Our target in studying hierarchical multitype point patterns is to assess the strength and range of interaction within each type and between all pairs of types. In practice, this means estimation of the parameters of the respective interaction functions. We shall consider two estimation methods, maximum likelihood, which involves laborious approximation of ratios of scaling factors, and maximum pseudolikelihood, which is a common short-cut omitting the scaling factors.

3.1 Maximum likelihood

Maximum likelihood is a straightforward method to estimate the parameters involved. It requires approximation of the scaling factor, which may lead to heavy simulations (Geyer and Thompson, 1992, Geyer and Møller, 1994). Let us assume that the numbers of points of each type in the pattern are fixed. According to density (1) the log likelihood function for the symmetric model is

$$\ell_{\mathbf{x}}(\theta_A, \theta_B, \theta_{AB}) = \sum h_A + \sum h_B + \sum h_{AB} - \log \int \prod h_A \prod h_B \prod h_{AB} P(d\mathbf{x}),$$

where θ_A, θ_B and θ_{AB} are the parameters (possibly vectors) of the corresponding interaction functions h . Furthermore, we have used short-hand notations

$$\sum h_A = \sum_{\substack{x_i, x_j \in \mathbf{x}_A \\ i < j}} h_A(x_i, x_j) \quad (8)$$

and

$$\sum h_{AB} = \sum_{\substack{x_i \in \mathbf{x}_A \\ x_j \in \mathbf{x}_B}} h_{AB}(x_i, x_j) \quad (9)$$

($\sum h_B, \prod h_A, \prod h_B$ and $\prod h_{AB}$ are defined analogously).

Differentiation with respect to θ_A, θ_B and θ_{AB} yields the estimation equations

$$\sum \frac{\partial h_A}{\partial \theta_A} = \mathbf{E}_{\theta_A, \theta_B, \theta_{AB}} \sum \frac{\partial h_A}{\partial \theta_A},$$

$$\sum \frac{\partial h_B}{\partial \theta_B} = \mathbf{E}_{\theta_A, \theta_B, \theta_{AB}} \sum \frac{\partial h_B}{\partial \theta_B}$$

and

$$\sum \frac{\partial h_{AB}}{\partial \theta_{AB}} = \mathbf{E}_{\theta_A, \theta_B, \theta_{AB}} \sum \frac{\partial h_{AB}}{\partial \theta_{AB}},$$

all expectations being with respect to the density (1).

For the hierarchical model the estimation can be done type by type according to the hierarchy. First, for the dominating type A the log likelihood function can be written as

$$\ell_{\mathbf{x}_A}(\theta_A) = \sum h_A - \log \int \prod h_A P(d\mathbf{x}_A),$$

and the corresponding estimation equation is

$$\sum \frac{\partial h_A}{\partial \theta_A} = \mathbf{E}_{\theta_A} \sum \frac{\partial h_A}{\partial \theta_A},$$

the expectation being with respect to the distribution (4) of \mathbf{x}_A .

Given the A pattern, the log likelihood for the lower level B is

$$\ell_{\mathbf{x}}(\theta_B, \theta_{AB}) = \sum h_B + \sum h_{AB} - \log \int \prod h_B \prod h_{AB} P(d\mathbf{x}_B),$$

and the estimation equations are

$$\sum \frac{\partial h_B}{\partial \theta_B} = \mathbf{E}_{\theta_B, \theta_{AB}} \sum \frac{\partial h_B}{\partial \theta_B}$$

and

$$\sum \frac{\partial h_{AB}}{\partial \theta_{AB}} = \mathbf{E}_{\theta_B, \theta_{AB}} \sum \frac{\partial h_{AB}}{\partial \theta_{AB}},$$

where the expectations are with respect to the conditional distribution (6) of \mathbf{x}_B given \mathbf{x}_A .

Notice that the ML estimation equations in these two cases, the symmetric and the hierarchical one, are all different. The expectations are with respect to different distributions.

In the calculation of the expected values we encounter with the scaling factor $Z_{AB}(\theta_A, \theta_B, \theta_{AB})$ for the symmetric model and the factors $Z_A(\theta_A)$ and $Z_B(\theta_B, \theta_{AB}; \mathbf{x}_A)$ for the hierarchical model. To avoid direct estimation of the factor we apply the Markov chain Monte Carlo (MCMC) approximation method by Geyer and Thompson (1992).

We have three likelihood functions: one for the symmetric model and two for the hierarchical one. Let us denote the likelihood function in general by

$$L_{\mathbf{x}}(\theta) = \frac{1}{Z} g(\mathbf{x}), \quad (10)$$

where $g(\mathbf{x})$ is the unnormalised density of the process in question and Z is the corresponding scaling factor. In the symmetric case $L_{\mathbf{x}}(\theta)$ is of the form (1). In the hierarchical case we have two likelihood functions. For the higher level $L_{\mathbf{x}}(\theta)$ is as in (4), and for the lower level, given the dominating pattern, of the form (6).

Let P_{θ_o} be a probability distribution with density f_{θ_o} , where θ_o is fixed. The scaling factor $Z(\theta)$ can be written as

$$\begin{aligned} Z(\theta) &= \int g(\theta, \mathbf{x}) P(d\mathbf{x}) = Z(\theta_o) \int \frac{g(\theta, \mathbf{x})}{g(\theta_o, \mathbf{x})} \frac{1}{Z(\theta_o)} g(\theta_o, \mathbf{x}) P(d\mathbf{x}) \\ &= Z(\theta_o) \int \frac{g(\theta, \mathbf{x})}{g(\theta_o, \mathbf{x})} P_{\theta_o}(d\mathbf{x}), \end{aligned}$$

and hence

$$\frac{Z(\theta)}{Z(\theta_o)} = \mathbf{E}_{\theta_o} \frac{g(\theta, \mathbf{x})}{g(\theta_o, \mathbf{x})}. \quad (11)$$

Now the log likelihood function of (10) can be written as

$$\ell_{\mathbf{x}}(\theta) = \log g(\mathbf{x}) - \log \frac{Z(\theta)}{Z(\theta_o)},$$

since adding a constant (with respect to θ) $\log Z(\theta_o)$ does not affect the estimation. Using the MCMC approximation of the expectation in (11) yields

$$\ell_{\mathbf{x}}(\theta) \approx \log g(\mathbf{x}) - \log \frac{1}{m} \sum_{i=1}^m \frac{g(\theta, \mathbf{x}_i)}{g(\theta_o, \mathbf{x}_i)}, \quad (12)$$

where the \mathbf{x}_i , $i = 1, \dots, m$, are samples from P_{θ_o} .

One advantage of the method is that we need simulations only from one distribution instead of having to simulate from a new distribution every time the value of θ is updated. The method requires the initial value θ_o being close to the unknown true value of θ . The pseudolikelihood estimates are usually good choices as initial values.

3.2 Maximum pseudolikelihood

To avoid the laborious approximation of the scaling factor an estimation method called the maximum pseudolikelihood (MPL) has been developed (for marked point processes, see Jensen and Møller, 1991, and Goulard *et al.*, 1996).

The log pseudolikelihood function can be written as

$$\log \text{PL}_{\mathbf{x}}(\beta, \theta) = \sum_{y \in \mathbf{x}} \log \lambda_{\beta, \theta}(y; \mathbf{x} \setminus \{y\}) - \int_W \lambda_{\beta, \theta}(\xi; \mathbf{x}) d\xi,$$

where β is a vector of the parameters b , θ contains the parameters of the interaction functions and $\lambda_{\beta, \theta}(\cdot; \cdot)$ is the conditional intensity function.

For the symmetric model it holds that

$$\begin{aligned} \log \text{PL}_{\mathbf{x}}(\beta, \theta) &= n(A) \log b_A + n(B) \log b_B \\ &\quad + 2 \sum \log h_A + 2 \sum \log h_B + 4 \sum \log h_{AB} \\ &\quad - b_A \int_W \prod_{x_i \in \mathbf{x}_A} h_A(x_i, \xi) \prod_{x_i \in \mathbf{x}_B} h_{AB}(x_i, \xi) d\xi \\ &\quad - b_B \int_W \prod_{x_i \in \mathbf{x}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{x}_A} h_{AB}(x_i, \xi) d\xi, \end{aligned} \quad (13)$$

where the short-hand notations are analogous to those in (8) and (9).

Since we concentrate on cases where the numbers of points are fixed, we may get rid of b_A and b_B in (13) by maximising with respect to these terms. Furthermore, without loss of generality, the terms independent of θ can be omitted in the sum. After these modifications we may write

$$\begin{aligned} \log \text{PL}_{\mathbf{x}}(\beta, \theta) &= 2 \sum \log h_A + 2 \sum \log h_B + 4 \sum \log h_{AB} \\ &\quad - n(A) \log \int_W \prod_{x_i \in \mathbf{x}_A} h_A(x_i, \xi) \prod_{x_i \in \mathbf{x}_B} h_{AB}(x_i, \xi) d\xi \\ &\quad - n(B) \log \int_W \prod_{x_i \in \mathbf{x}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{x}_A} h_{AB}(x_i, \xi) d\xi. \end{aligned}$$

We have three parameters to be estimated, θ_A , θ_B and θ_{AB} . Differentiation with respect to θ_A gives

$$2 \sum \frac{\partial \log h_A}{\partial \theta_A} = \tag{14}$$

$$n(A) \frac{\int_W \sum_{x_i \in \mathbf{X}_A} \frac{\partial \log h_A(x_i, \xi)}{\partial \theta_A} \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) \prod_{x_i \in \mathbf{X}_B} h_{AB}(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) \prod_{x_i \in \mathbf{X}_B} h_{AB}(x_i, \xi) d\xi},$$

with respect to θ_B

$$2 \sum \frac{\partial \log h_B}{\partial \theta_B} = \tag{15}$$

$$n(B) \frac{\int_W \sum_{x_i \in \mathbf{X}_B} \frac{\partial \log h_B(x_i, \xi)}{\partial \theta_B} \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}$$

and with respect to θ_{AB}

$$4 \sum \frac{\partial \log h_{AB}}{\partial \theta_{AB}} = \tag{16}$$

$$n(A) \frac{\int_W \sum_{x_i \in \mathbf{X}_B} \frac{\partial \log h_{AB}(x_i, \xi)}{\partial \theta_{AB}} \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) \prod_{x_i \in \mathbf{X}_B} h_{AB}(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) \prod_{x_i \in \mathbf{X}_B} h_{AB}(x_i, \xi) d\xi}$$

$$+ n(B) \frac{\int_W \sum_{x_i \in \mathbf{X}_A} \frac{\partial \log h_{AB}(x_i, \xi)}{\partial \theta_{AB}} \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}.$$

The parameters for the hierarchical model can be estimated in two steps. First, we write the PL function for the dominating type A as

$$\log \text{PL}_{\mathbf{x}}(\theta_A) = 2 \sum \log h_A - n(A) \log \int_W \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) d\xi.$$

The estimate for θ_A is obtained as a solution of the equation

$$2 \sum \frac{\partial \log h_A}{\partial \theta_A} = n(A) \frac{\int_W \sum_{x_i \in \mathbf{X}_A} \frac{\partial \log h_A(x_i, \xi)}{\partial \theta_A} \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_A} h_A(x_i, \xi) d\xi}. \tag{17}$$

Second, the PL function for the type B can be written as

$$\begin{aligned} \log \text{PL}_{\mathbf{x}}(\theta_B, \theta_{AB}) &= 2 \sum \log h_B + 2 \sum \log h_{AB} \\ &\quad - n(B) \log \int_W \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi. \end{aligned}$$

By differentiating with respect to θ_B we obtain the estimation equation

$$\begin{aligned} 2 \sum \frac{\partial \log h_B}{\partial \theta_B} &= \tag{18} \\ n(B) \frac{\int_W \sum_{x_i \in \mathbf{X}_B} \frac{\partial \log h_B(x_i, \xi)}{\partial \theta_B} \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi} \end{aligned}$$

and with respect to θ_{AB}

$$\begin{aligned} 2 \sum_{AB} \frac{\partial \log h_{AB}}{\partial \theta_{AB}} &= \tag{19} \\ n(B) \frac{\int_W \sum_{x_i \in \mathbf{X}_A} \frac{\partial \log h_{AB}(x_i, \xi)}{\partial \theta_{AB}} \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}{\int_W \prod_{x_i \in \mathbf{X}_B} h_B(x_i, \xi) \prod_{x_i \in \mathbf{X}_A} h_{AB}(x_i, \xi) d\xi}. \end{aligned}$$

The equations (14) and (17) derived for the estimation of θ_B are equal for both models. Hence, if the estimates of θ_{AB} coincide for both models, then also the estimates of θ_B are equal.

The estimation equation for θ_A , interaction within the dominating type, is different in these two cases if the two types interact mutually. The interaction between the types affects the estimation of θ_A in the symmetric case but not in the hierarchical case.

The ML method, instead, preserves the scaling factor, and therefore the estimation equations for the hierarchical model differ from the corresponding ones for the symmetric model.

4 Simulation study

We carried out a simulation study to explore the performance both of the symmetric and the hierarchical model when true model is hierarchical.

Additionally, we were interested in comparing the ML and MPL estimation methods.

We restrict ourselves to the simplest case, the bivariate Strauss model

$$h_i(x, y; \theta_i) = \begin{cases} 0, & \text{if } d(x, y) = 0, \\ \theta_i, & \text{if } 0 < d(x, y) \leq R, \\ 1, & \text{if } d(x, y) > R, \end{cases}$$

where the θ_i , $0 \leq \theta_i \leq 1$, $i = A, B, AB$, are the interaction parameters within and between the types, and R is the interaction radius. We used two values for the interaction parameters θ_i : 0.7 representing weak repulsion and 0.2 representing strong repulsion. In addition, we had two values for the interaction radius R , 0.05 corresponding to a medium range and 0.1 to a large range interaction.

For the actual simulation of hierarchical patterns we chose to apply the spatial birth-and-death process with a fixed number of points (Ripley, 1977). The estimates of θ_A were calculated from 200 simulated A patterns consisting of 50 points, which were generated in the unit square mapped onto a torus. Before starting to record samples we ran 10 000 warm-up steps with a Poisson realisation as the initial pattern. To obtain approximate independence among the samples only every 200th pattern was sampled. To calculate the estimates for θ_B and θ_{AB} we first simulated a realisation of 50 A points as above. Then, by conditioning on this A pattern, we produced 200 B subpatterns of 50 points using a Poisson realisation as the initial pattern and having 10 000 warm-up steps before sampling. Again, every 200th pattern was sampled. Having the A pattern fixed it was sufficient to approximate only one ratio of scaling factors. To approximate the ratios of normalising constants needed in the ML estimation (see formulae (11) and (12)) 10 000 samples were collected. The integrals in the MPL estimation equations (14)–(18) are approximated by using a 100×100 grid.

The results presented in Tables 1 and 2 imply that both the ML and MPL methods give reliable and unbiased estimates if the model is correct, that is, hierarchical. Standard deviations are reasonable and, as expected, slightly smaller in the case of large range of interaction ($R = 0.1$).

If we assume a symmetric model when hierarchy is present we get different results. The MPL method gives almost the same estimates for θ_B and θ_{AB} under both the hierarchical and the symmetric model assumptions. This is plausible given the corresponding estimation equations (14), (17), (15)

and (18). The estimates of θ_A are different under different models. Symmetric model leads to underestimation of θ_A (or overestimation of the repulsion), especially, if the repulsion between the A points and B points is strong. The reason for different estimates of θ_A is that under the symmetric model the interaction between the types affects the estimation of θ_A while under the hierarchical model it does not. The ML method underestimates all interaction parameters if a symmetric model is assumed.

The ratios of scaling factors used in the ML estimation are approximated from the simulations of the symmetric model applying the original parameter values. If the ML estimates are far from the true values the approximation of the ratio of scaling factors may be poor or even the likelihood function may not have a maximum. The latter occurred several times in the estimation; the empty entries in Tables 1 and 2 indicate that the maximisation algorithm did not converge. Further seeking for correct ML estimates did not seem relevant since we already have evidence of the difference between the estimates under the symmetric model and the hierarchical model.

The simulation study reveals that if a symmetric model is applied when the truth is hierarchical, the estimation of the interaction becomes biased. This may lead to wrong conclusions of the intra- and intertype relationships. Therefore, it is important to recognise the hierarchical structure whenever it exists and to incorporate it into the model.

5 Example: A community of two ant species

Let us consider a bivariate point pattern of ants' nest sites studied first by Harkness and Isham (1983). The two species living in a 240ft \times 250ft study area (Figure 1) in Greece, are *Messor wasmanni* with 45 nests and *Cataglyphis bicolor* with 15 nests. The *Messors* (species M for short) collect seeds for nutrition while the *Cataglyphis* ants (species C) feed upon dead insects, mostly dead *Messor* ants. Therefore, Harkness and Isham, in their study of the intra- and interspecific interactions in this community, presumed that the *Cataglyphis* ants are attracted by the *Messor* ants and hence there should be a positive interaction between the nest locations of these two species. They concluded that the marginal pattern of *Messor* nests is inhibitive but the *Cataglyphis* nest pattern can be regarded as a realisation of a Poisson process. To estimate the interaction between the species they applied two methods, nearest neighbour measurements and the cross function

K_{MC} (Lotwick and Silverman, 1982). The nearest neighbour measurements support the hypothesis of attraction, that is, the observed data include more *Cataglyphis* nests with a *Messor* nest as the nearest neighbour than it would be expected if the patterns were mutually independent. Contradictory, the estimated K_{MC} function suggests independency between the species.

Takacs and Fiksel (1986) fitted a pairwise interaction model to the same data. Their result was that the *Cataglyphis* nests form a rather strongly repulsive pattern while the *Messor* nests are nearly Poisson distributed. The interaction between the two species is repulsive but not as strongly as among the *Cataglyphis* ants. The goodness-of-fit of the model was not tested.

A Strauss type model was fitted by Särkkä (1993) suggesting fairly strong repulsion among both species and weak repulsion between the species. However, an L function (1–1 linearisation of K) study reveals that the goodness-of-fit of the model is not satisfactory. Notice that both the model of Takacs and Fiksel and the one of Särkkä suggest weak repulsion between the species.

In the studies reviewed above the interaction between the species was of main interest. According to Harkness and Isham we can not make any conclusions owing to the contradictory results obtained by two different methods. Takacs and Fiksel did not test the fit of their model, and the model of Särkkä did not fit well to the data.

The ecological relationship of the two species, that is, the *Cataglyphis* ants feeding upon dead *Messor* ants but not hunting for them gives an idea for a hierarchical model: The *Messor* nests affect the locations of the *Cataglyphis* nests but not vice versa. Before fitting a hierarchical model we recall the study of Särkkä to make it easier to compare the goodness-of-fit of the symmetric model and the corresponding hierarchical model. In the estimation the MPL method was applied and the edge effects were taken into account by mapping the rectangular study region onto a torus.

We fitted both a symmetric and a hierarchical model applying Strauss type interactions to the data. Furthermore, to find out the goodness-of-fit of the models we calculated the empirical marginal L functions and the respective upper and lower envelopes by simulating the models identified by the estimated parameter values. This is a usual way to test the goodness-of-fit of such models. However, the L function is defined only for stationary and isotropic processes, and non-stationarity is an essential feature of our hierarchical approach. Since the dominating M points are assumed to be a realisation of a stationary process, the estimation and interpretation of

function L_M and the cross function L_{MC} estimated using only *Messor* nests as centre points are not problematic. Instead, the points on the lower level are regarded as a realisation of a non-stationary process, and thus L_C may not have an obvious interpretation. However, it can be used in testing the goodness-of-fit since it is estimated in the same way both from the data and from the simulated patterns so that it “measures” the same property in both cases.

5.1 Symmetric model

We fitted the symmetric model with Strauss type interactions presented in Section 2.1. The interaction function is of the form

$$h_i(d(x, y)) = \begin{cases} 0, & \text{if } d(x, y) \leq r_i, \\ \theta_i, & \text{if } r_i < d(x, y) \leq R_i, \\ 1, & \text{if } d(x, y) > R_i, \end{cases}$$

where the r_i are the fixed hard-core radii and the R_i the fixed interaction radii, $i = M, C, MC$.

We followed Takacs and Fiksel (1986) in choosing the minimum interpoint distances in the data as hard-core distances, that is, $r_M = 9.1\text{ft}$, $r_C = 2.5\text{ft}$ and $r_{MC} = 5.5\text{ft}$, and 45ft for all interaction distances R_M , R_C and R_{MC} . In general, taking the shortest observed interpoint distance as the hard-core radius of the process is maybe not recommendable. This may cause anomalies in simulations, since now our data, regarded as a realisation of the corresponding process, involves an outlier pair of points with an extremely short interpoint distance.

The maximum likelihood method gives estimates $\hat{\theta}_M = 0.31$, $\hat{\theta}_C = 0.16$ and $\hat{\theta}_{MC} = 0.65$, while the corresponding MPL estimates are $\hat{\theta}_M = 0.45$, $\hat{\theta}_C = 0.40$ and $\hat{\theta}_{MC} = 0.88$. The results indicate rather strong repulsion both within the *Cataglyphis* nests and within the *Messor* nests but only weak repulsion between the species.

To test the goodness-of-fit of the model we did the following. First the marginal empirical functions L_M and L_C , the cross L_{MC} and the joint L disregarding the types of the points were calculated from the data. Then the corresponding upper and lower envelopes were calculated from 99 realisations which were simulated by using the respective model, that is, the bivariate

symmetric model with the estimated (ML and MPL) parameter values. Figures 2 and 3 show that the symmetric models do not fit well to the data: The empirical functions L_M , L_C and the joint L do not lie between the envelopes.

5.2 Hierarchical model

Finally, we fitted the corresponding hierarchical model with Strauss type interaction functions. The values of the hard-core and interaction radii were the same as for the symmetric model.

According to the hierarchy the interaction within the *Messor* nests was estimated first. The ML method yields estimate $\hat{\theta}_M = 0.43$ and the MPL method $\hat{\theta}_M = 0.44$. Both indicate repulsion within the *Messors*. Second, the locations of *Cataglyphis* nests were considered given the *Messor* pattern. Repulsion within the *Cataglyphis* nests is strong according to both the ML and the MPL estimate $\hat{\theta}_C = 0.40$, but the repulsion between the species is weaker, both methods giving $\hat{\theta}_{MC} = 0.79$.

The goodness-of-fit of the model was tested by using the marginal functions L_M and L_C and the cross L_{MC} (see Figure 4). Concordant with the hierarchy, L_{MC} is estimated using the *Messor* nests as centre points only. First we calculated the empirical L function for the *Messor* nests and then simulated 99 configurations using the univariate models with the estimated parameter values to get the upper and lower envelopes. The model seems to fit well to the data. Second, for the *Cataglyphis* nests we calculated the empirical L_C function from the *Cataglyphis* data and estimated the envelopes by simulating 99 *Cataglyphis* nest patterns conditioned on the true configuration of the *Messor* nests. The goodness-of-fit of this part of the model is also good.

5.3 Comparison of the models

Both models, the symmetric and the hierarchical one, give very similar results if only the parameter estimates are considered: There is rather strong repulsion within both species and weaker repulsion between the species. Furthermore, the order of the strenghts of interaction is the same in both models.

The L function study shows that the hierarchical model fits well to the data, unlike the symmetric model. This supports the assumption of a hierarchical interaction. It is maybe surprising that, according to our results,

the *Cataglyphis* ants tend to avoid the vicinity of *Messor* nests when choosing their nest sites, as attraction has been anticipated. A possible ecological explanation might be that the *Cataglyphis* ants locate their nests reasonably close to many *Messor* nests instead of only one, and since the pattern of the *Cataglyphis* nests is repulsive itself, this leads to a repulsive inter-species configuration.

Acknowledgments

We wish to thank Antti Penttinen and Pavel Grabarnik, as well as a good number of other colleagues for fruitful discussions on the topic. The financial support of the foundations Adlerbertska forskningsfonden, Göteborgs handelshögskolefonder and Selma Andersons fond is highly appreciated. This work is part of the Projects 37206 and 54892 of the Academy of Finland.

References

- Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
- Ford, E. D. & Diggle, P. J. (1981). Competition for light in a plant monoculture modelled as a spatial stochastic process. *Annals of Botany* **48**, 481–500.
- Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.* **21**, 359–373.
- Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *J. Roy. Statist. Soc. B* **54**, 657–699.
- Goulard, M., Grabarnik, P. & Särkkä, A. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scand. J. Stat.* **23**, 365–379.
- Harkness, R. D. & Isham, V. (1983). A bivariate spatial point pattern of ants' nests. *Appl. Statist.* **32**, 293–303.
- Jensen, J. L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **3**, 445–461.
- Lotwick, H. W. & Silverman, B. W. (1982). Methods for analysing spatial processes of several types of points. *J. Roy. Statist. Soc. B* **44**, 406–413.
- Mugglestone, M. A. & Renshaw, E. (1996). The exploratory analysis of bivariate spatial point patterns using cross-spectra. *Environmetrics* **7**, 361–377.
- Renshaw, E. (1984). Competition experiments for light in a plant monoculture: an analysis based on two-dimensional spectra. *Biometrics* **40**, 717–728.
- Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. B* **39**, 172–212.

- Särkkä, A. (1993). Pseudo-likelihood approach for pair potential estimation of Gibbs processes. *Jyväskylä Studies in Computer Science, Economics and Statistics* 22, University of Jyväskylä.
- Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and Its Applications*. 2nd ed. Chichester: John Wiley & Sons.
- Takacs, R. & Fiksel, T. (1986). Interaction pair-potentials for a system of ants' nests. *Biom. J.* 28, 1007–1013.

Table 1. Empirical means and standard deviations of the ML and the MPL estimates when $R = 0.1$.

Parameter	True value	ML (hier)	ML (symm)	MPL (hier)	MPL (symm)
θ_A	0.70	0.70 (0.17)	0.45 (0.23)	0.68 (0.16)	0.63 (0.15)
θ_B	0.70	0.69 (0.14)	0.48 (0.21)	0.67 (0.14)	0.67 (0.14)
θ_{AB}	0.70	0.71 (0.11)	0.49 (0.11)	0.71 (0.11)	0.72 (0.11)
θ_A	0.70	0.70 (0.17)	0.78 (6.06)	0.67 (0.16)	0.33 (0.12)
θ_B	0.70	0.69 (0.12)		0.67 (0.14)	0.68 (0.13)
θ_{AB}	0.20	0.20 (0.05)	0.00 (0.00)	0.19 (0.05)	0.22 (0.06)
θ_A	0.20	0.20 (0.07)	0.00 (0.00)	0.20 (0.07)	0.19 (0.07)
θ_B	0.20	0.20 (0.07)	0.00 (0.00)	0.19 (0.07)	0.19 (0.07)
θ_{AB}	0.70	0.73 (0.17)	0.31 (0.16)	0.71 (0.17)	0.71 (0.18)
θ_A	0.20	0.21 (0.07)		0.20 (0.07)	0.10 (0.05)
θ_B	0.20	0.21 (0.08)		0.19 (0.07)	0.19 (0.07)
θ_{AB}	0.20	0.21 (0.06)		0.19 (0.07)	0.20 (0.07)

Table 2. Empirical means and standard deviations of the ML and the MPL estimates when $R = 0.05$.

Parameter	True value	ML (hier)	ML (symm)	MPL (hier)	MPL (symm)
θ_A	0.70	0.72 (0.30)	0.46 (0.24)	0.70 (0.28)	0.68 (0.27)
θ_B	0.70	0.67 (0.26)	0.48 (0.67)	0.65 (0.26)	0.65 (0.26)
θ_{AB}	0.70	0.72 (0.21)	0.45 (0.21)	0.71 (0.20)	0.71 (0.20)
θ_A	0.70	0.71 (0.28)	0.24 (1.40)	0.70 (0.27)	0.59 (0.23)
θ_B	0.70	0.73 (0.28)		0.70 (0.25)	0.70 (0.25)
θ_{AB}	0.20	0.20 (0.11)	0.00 (0.01)	0.20 (0.10)	0.20 (0.10)
θ_A	0.20	0.22 (0.17)	0.01 (0.02)	0.21 (0.15)	0.21 (0.15)
θ_B	0.20	0.21 (0.13)	0.00 (0.02)	0.20 (0.13)	0.20 (0.13)
θ_{AB}	0.70	0.71 (0.22)	0.13 (0.21)	0.71 (0.22)	0.71 (0.21)
θ_A	0.20	0.22 (0.19)	0.01 (0.06)	0.21 (0.14)	0.17 (0.12)
θ_B	0.20	0.19 (0.12)	0.00 (0.01)	0.19 (0.12)	0.19 (0.12)
θ_{AB}	0.20	0.21 (0.10)	0.00 (0.00)	0.21 (0.09)	0.21 (0.09)

Labels of figures

Figure 1. Locations of ants' nests: Nests of *Messor wasmanni* are denoted by crosses and nests of *Cataglyphis bicolor* by circles.

Figure 2. Estimated L functions (solid) and the upper and lower envelopes (dashed) from 99 simulations of the ML estimated symmetric model: a) marginal L for the *Messor* nests, b) marginal L for the *Cataglyphis* nests, c) joint L and d) cross L_{MC} .

Figure 3. Estimated L functions (solid) and the upper and lower envelopes (dashed) from 99 simulations of the MPL estimated symmetric model: a) marginal L for the *Messor* nests, b) marginal L for the *Cataglyphis* nests, c) joint L and d) cross L_{MC} .

Figure 4. Estimated L functions (solid) and the upper and lower envelopes (dashed) from 99 simulations of the estimated hierarchical model: a) marginal L for the *Messor* nests, b) marginal L for the *Cataglyphis* nests and c) cross L_{MC} .

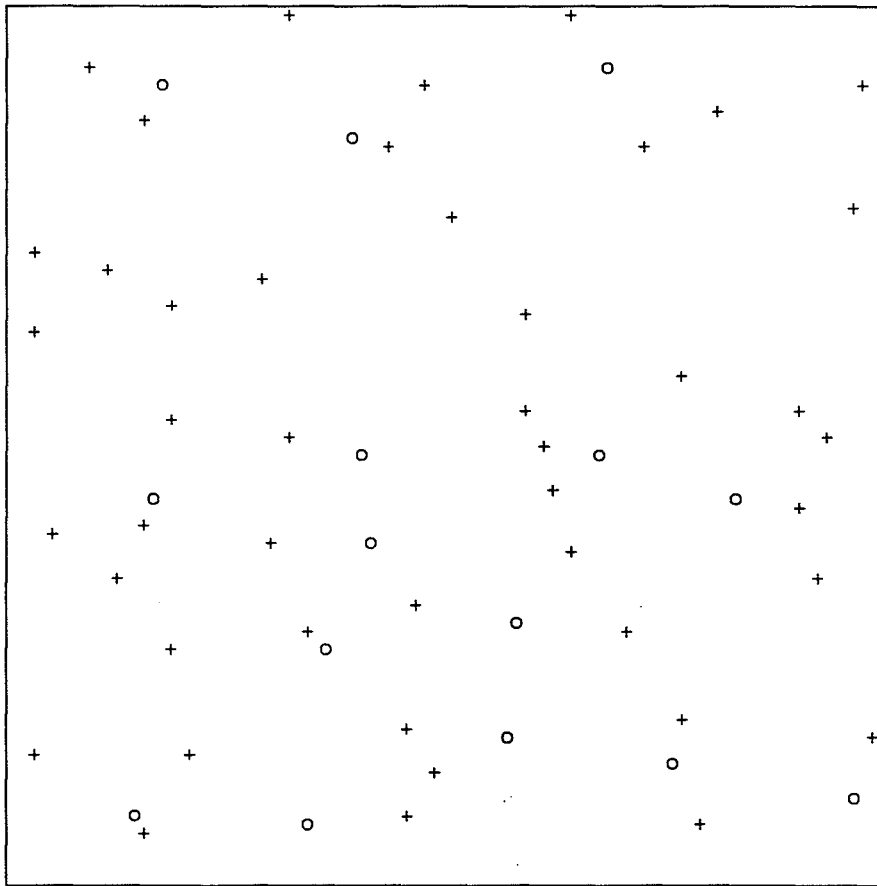


FIGURE 1

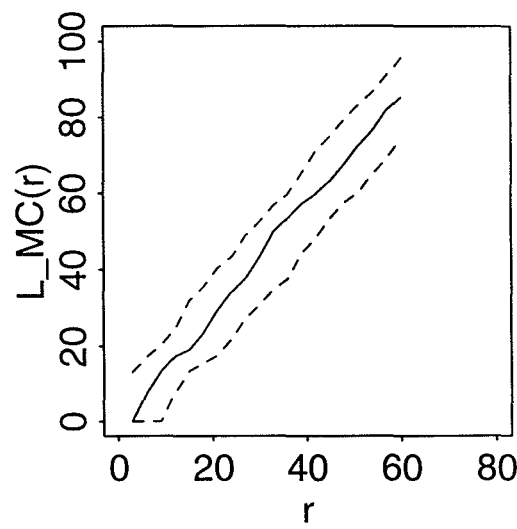
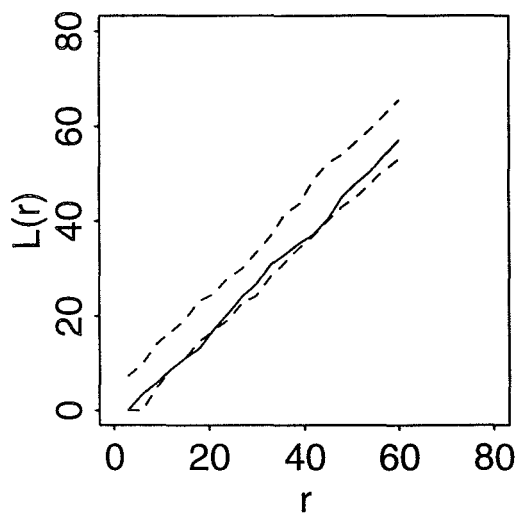
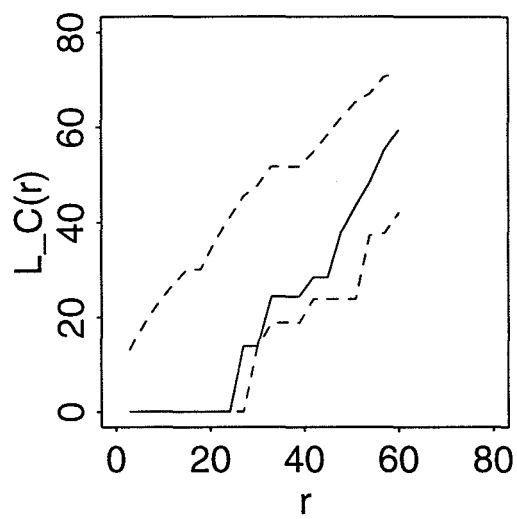
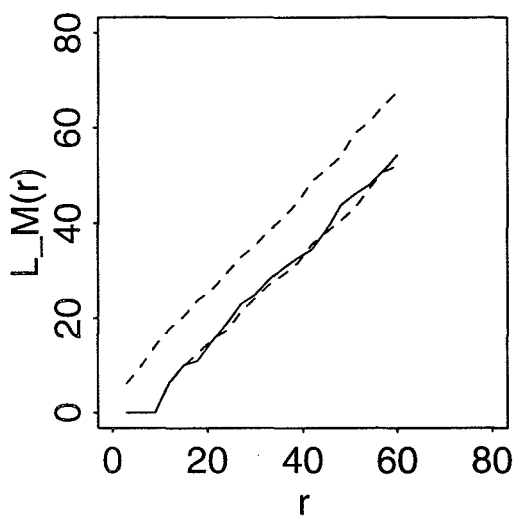


FIGURE 2

FIGURE 3

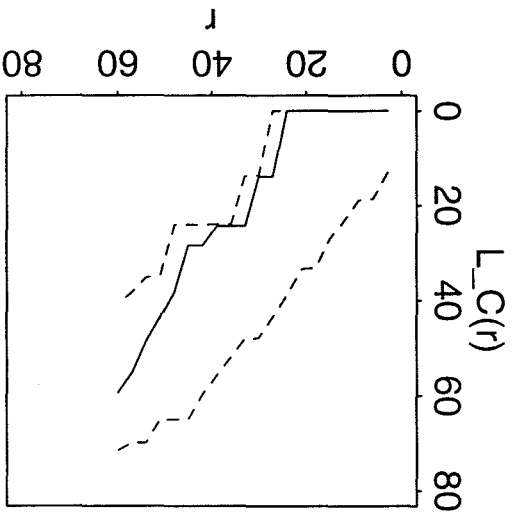
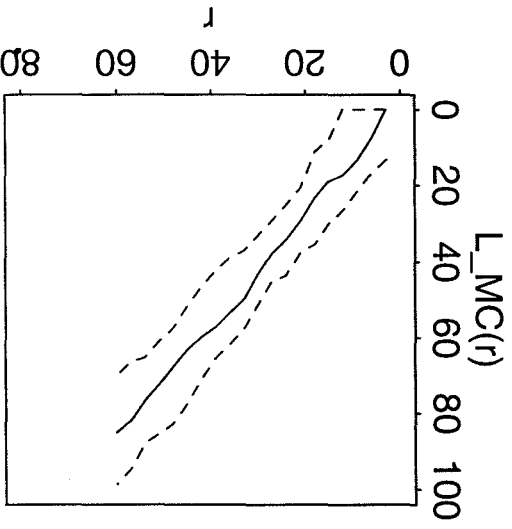
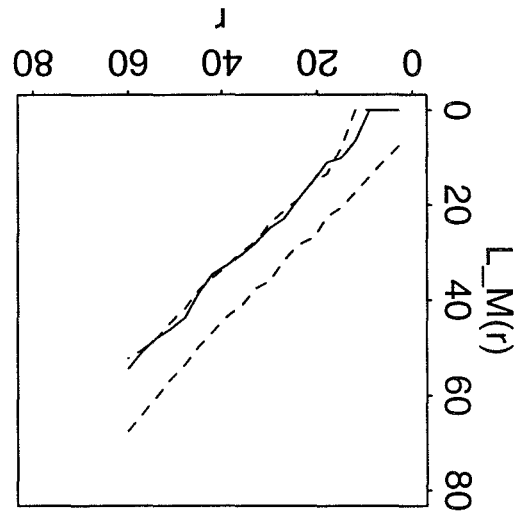
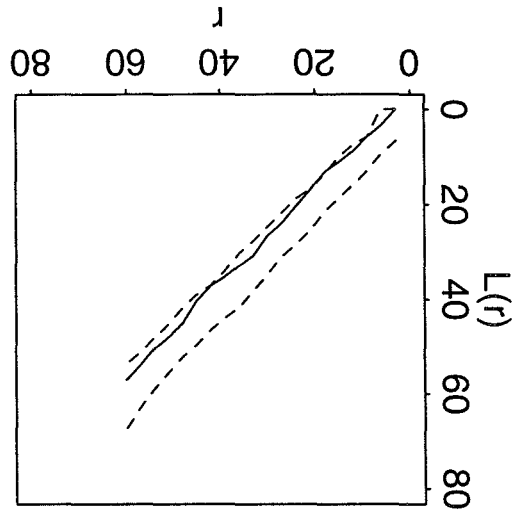
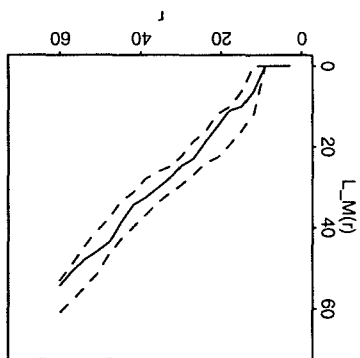
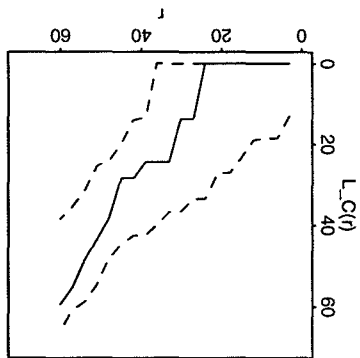
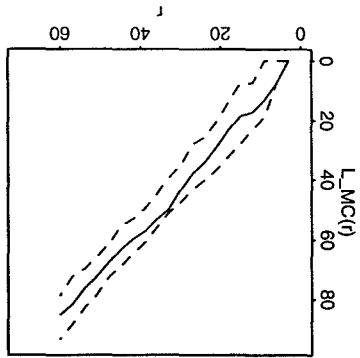


FIGURE 4



Research Report

- | | | |
|--------|-------------------------|---|
| 1995:1 | Arnkeldottir, H | Surveillance of rare events. On evaluations of the sets method. |
| 1995:2 | Sveréus, A | Detection of gradual changes. Statistical methods in post marketing surveillance. |
| 1995:3 | Ekman, C | On second order surfaces estimation and rotatability. |
| 1996:1 | Ekman, A | Sequential analysis of simple hypotheses when using play-the-winner allocation. |
| 1996:2 | Wessman, P | Some principles for surveillance adopted for multivariate processes with a common change point. |
| 1996:3 | Frisén, M. & Wessman, P | Evaluations of likelihood ratio methods for surveillance. |
| 1996:4 | Wessman, P. | Evaluation of univariate surveillance procedures for some multivariate problems. |
| 1996:5 | Särkkä, A. | Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes. |
| 1997:1 | Ekman, A. | Sequential probability ratio tests when using randomized play-the-winner allocation. |
| 1997:2 | Pettersson, M. | Monitoring a Freshwater Fishpopulation: Statistical Surveillance of Biodiversity. |
| 1997:3 | Jonsson, R. | Screening-related prevalence and incidence for non-recurrent diseases. |
| 1997:4 | Johnsson, T. | A stepwise regression procedure applied to an international production study - a multiple inference solution. |