Research Report
Department of Statistics
Göteborg University
Sweden

Some interaction models for clustered point patterns

Pavel Grabarnik
Aila Särkkä

Research Report 1998:8
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			

Some interaction models for clustered point patterns

Pavel Grabarnik
The Institute of Basic Problems of Biology, Pushchino, Russia

Aila Särkkä
Chalmers University of Technology, Göteborg, Sweden

November 2, 1998

Abstract

We introduce a class of spatial point processes, interacting neighbour processes, where the density of the process can be written by means of local interactions between a point and subsets of its neighbourhood but where the processes are not Markov processes with respect to this neighbourhood in the Ripley-Kelly sense. However, we show that the processes are nearest neighbour Markov processes as introduced by Baddeley and Møller (1989). Furthermore, we introduce a subclass of interacting neighbour processes, full neighbourhood interaction processes, where instead of subsets of the neighbourhood all neighbours of a point affect it simultaneously. A simulation study is presented to show that some simple full neighbourhood interaction models can produce clustered patterns of great variety. Finally, an empirical example is given.

Key words: Clustered point patterns, local interactions, static and dynamic neighbours, Ripley-Kelly Markov processes, nearest neighbour Markov processes

1 Introduction

Markov point processes are often used to analyse point patterns with interaction between points. Commonly used pairwise interaction processes are good models for repulsive point patterns but they do not seem to be able to produce clustered patterns in large enough variety. Some other local interaction processes have been suggested as models for clustered patterns. Baddeley and van Lieshout (1995) suggested area-interaction processes as an alternative for pairwise interaction

processes. Shot noise weighted processes were considered by van Lieshout and Molchanov (1997) and continuum random cluster processes by Møller (1998). Furthermore, Geyer (1998) introduced two point process models, triplet processes and saturation processes, which are both feasible models for clustering.

In order to model clustered point patterns we would like to introduce a class of interaction processes with the following properties: First, the models should have a natural distance-based neighbourhood relation, i.e. there is interaction between the points only if they are close to each other. The points of the continuum random cluster process may interact even if they are far away from each other; they only need to be in the same connected component. Secondly, the interaction function should have a natural interpretation. Furthermore, the number of parameters should be kept reasonably small.

Geyer's saturation process has the properties mentioned above. It resembles the Strauss process which is a pairwise interaction process, where the interaction is defined in terms of the number of r-close pairs. The Strauss process, however, is not a suitable model for clustering since the number of neighbours of a point can increase without bound if the number of points is not fixed and, therefore, the density cannot be normalized. Even if we condition on the number of points, the process is not an appropriate model for clustering. The saturation process, on the other hand, has an extra parameter which puts an upper bound on the influence of any single point, and which, therefore, overcomes the normalizing problem in the clustered case. Characteristic for the saturation process is that all neighbours of a point affect the point simultaneously; interaction cannot be factorized into pairwise, triple etc. interactions. Our aim is to find a class of processes with the same property.

First, we define a general class of point processes, interacting neighbour processes, where a point is affected by subsets of its neighbourhood. This class of processes includes all Ripley-Kelly Markov processes with the same fixed neighbourhood. Furthermore, we show that the interacting neighbour processes are nearest neighbour Markov processes. Then, we introduce a subclass of interacting neighbour processes, full neighbourhood interaction processes, where all neighbours of a point (instead of subsets of the neighbourhood) affect the point simultaneously. The saturation process is one example of full neighbourhood interaction processes. Some other examples are given in Section 4.

Furthermore, we present a simulation study to see how well some of the full neighbourhood interaction processes and the continuum random cluster process can model clustering. Finally, we fit a full neighbourhood interaction process model to the Scots pines data (Penttinen *et al.*, 1992).

2 Interacting neighbour processes

We consider a finite spatial point process X on a bounded set $A \subset \mathbf{R}^d$ (here, $d = 2$) which is defined by a positive density function f with respect to the Poisson process with intensity 1. A sample space Ω_A is a set of all possible point patterns $\mathbf{x} = \{x_1, ..., x_{n(\mathbf{x})}\}$ with no multiple points. The number of points in the realization $\mathbf{x}$ is denoted by $n(\mathbf{x})$.

To model point patterns with interaction between the points, it is convenient to use Markov point processes. Basis of the Markov point processes is the definition of a neighbourhood. Let $\sim$ be a symmetric, reflexive relation on A. If two points ξ and η in A are related under $\sim$, we write $\xi \sim \eta$, and say that ξ and η are neighbours with respect to the relation $\sim$. The set of all $\eta \in \mathbf{x} \setminus \{\xi\}$, $\eta \sim \xi$, is called the neighbourhood of ξ and denoted by $N(\xi)$ or $N_{\mathbf{x}}(\xi)$. Note that according to this definition of neighbours, two points are or are not neighbours independently of the realization. We refer to such neighbours as *static* neighbours. Baddeley and Møller (1989) define nearest neighbour Markov processes which allow the neighbourhood relation to depend on the realization. We say that such neighbourhood relation is *dynamic* and denote it by $\sim_{\mathbf{x}}$.

According to the factorization theorem by Ripley and Kelly (1977) the density of a Markov point process can be written as a product of interaction functions, where the product is over all cliques in $\mathbf{x}$. A *clique* is a set of points of $\mathbf{x}$, where all pairs of points are neighbours (does not have to be a maximal set). In particular, a one point set is a clique. Our aim is to construct models that allow interactions between a point and any set of its neighbours even though the set of neighbours is not a clique (with respect to $\sim$).

Definition 1: A point process X is an *interacting neighbour process* with a given neighbourhood relation $\sim$ if its density has the form

$$f(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} \prod_{\mathbf{z} \subseteq \mathbf{x} \setminus \{x_i\}} Q(x_i, \mathbf{z}),$$

where α is a normalizing constant and $Q : A \times \Omega_A \to (0, \infty)$ is a *neighbour set interaction function* for which $Q(x_i, \mathbf{z}) = 1$ if $\mathbf{z}$ is not a subset of $N(x_i)$. Thus, the density of the interacting neighbour process can be written as

$$f(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} \prod_{\mathbf{z} \subseteq N(x_i)} Q(x_i, \mathbf{z}). \tag{1}$$

Remark: Ripley-Kelly Markov processes with the given neighbourhood relation $\sim$ are interacting neighbour processes. The density of Ripley-Kelly Markov processes can be written as a product of interaction functions over cliques $\mathcal{C}$, i.e.

$$f_{RK}(\mathbf{x}) = \alpha \prod_{\mathbf{y} \subseteq \mathbf{x}} h(\mathbf{y}) = \alpha \prod_{\mathbf{y} \in \mathcal{C}} h(\mathbf{y}),$$

where $h : \Omega_A \to (0, \infty)$ is an interaction function for which $h(\mathbf{y}) = 1$ if $\mathbf{y}$ is not a clique. We can write

$$f_{RK}(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} \prod_{\mathbf{z} \cup \{x_i\} \in \mathcal{C}} h^{\frac{1}{\#\{\mathbf{z} \cup \{x_i\}\}}}(\mathbf{z} \cup \{x_i\}),$$

where $\#\{\cdot\}$ denotes the number of points in $\{\cdot\}$. A set $\mathbf{z} \cup \{x_i\}$ cannot be a clique if $\mathbf{z}$ is not a subset of $N(x_i)$, and therefore, we can write

$$f_{RK}(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} \prod_{\mathbf{z} \subseteq N(x_i)} Q(x_i, \mathbf{z}),$$

where

$$Q(x_i, \mathbf{z}) = \begin{cases} h^{\frac{1}{\#\{\mathbf{z} \cup \{x_i\}\}}}(\mathbf{z} \cup \{x_i\}) & \text{if } \mathbf{z} \cup \{x_i\} \text{ is a clique} \\ 1 & \text{otherwise.} \end{cases}$$

Since the class of interacting neighbour processes includes the whole class of Ripley-Kelly Markov processes (with a given neighbourhood $\sim$, which does not depend on the realization), the class contains processes with interactions of any order.

3 Markov property

According to Ripley and Kelly (1977) a point process is a Markov point process with respect to $\sim$ if the conditional intensity $\lambda(\xi; \mathbf{x}) = \frac{f(\mathbf{x} \cup \{\xi\})}{f(\mathbf{x})}$ depends only on ξ and its $\sim$-neighbours. Baddeley and Møller (1989) and Kendall (1990) consider nearest neighbour Markov processes which allow the neighbourhood relation to depend on the realization.

Let us now consider the Markov property of the interacting neighbour processes. First, we investigate whether the processes are Ripley-Kelly Markov processes with respect to a given neighbourhood relation $\sim$. The conditional intensity can be written as

$$\begin{aligned} \lambda(\xi; \mathbf{x}) &= \frac{\prod_{\mathbf{z} \subseteq N_{\mathbf{x} \cup \{\xi\}}(\xi)} Q(\xi, \mathbf{z}) \cdot \prod_{x_i \in \mathbf{x}} \prod_{\mathbf{z} \subseteq N_{\mathbf{x} \cup \{\xi\}}(x_i)} Q(x_i, \mathbf{z})}{\prod_{x_i \in \mathbf{x}} \prod_{\mathbf{z} \subseteq N_{\mathbf{x}}(x_i)} Q(x_i, \mathbf{z})} \qquad (2) \\ &= \prod_{\mathbf{z} \subseteq N_{\mathbf{x} \cup \{\xi\}}(\xi)} Q(\xi, \mathbf{z}) \cdot \prod_{x_i \sim \xi} \frac{\prod_{\mathbf{z} \subseteq N_{\mathbf{x} \cup \{\xi\}}(x_i)} Q(x_i, \mathbf{z})}{\prod_{\mathbf{z} \subseteq N_{\mathbf{x}}(x_i)} Q(x_i, \mathbf{z})} \\ &= \prod_{\mathbf{z} \subseteq N_{\mathbf{x} \cup \{\xi\}}(\xi)} Q(\xi, \mathbf{z}) \cdot \prod_{x_i \sim \xi} \prod_{\mathbf{z} \subseteq N_{\mathbf{x}}(x_i)} Q(x_i, \mathbf{z} \cup \{\xi\}). \end{aligned}$$

We can conclude that the conditional intensity depends not only on ξ and the neighbours of ξ, but also on the neighbours of neighbours of ξ. Therefore, the process is not a Markov process in the sense of Ripley and Kelly with respect to the neighbourhood relation $\sim$.

Let us then consider another neighbourhood relation. Two points ξ and η of $\mathbf{x}$ are said to be *iterated neighbours*, and write $\xi \sim^2_{\mathbf{x}} \eta$, whenever $\xi \sim \eta$ or whenever there is another point $\zeta \in \mathbf{x}$ such that $\zeta \sim \xi$ and $\zeta \sim \eta$. A similar neighbourhood relation, iterated Dirichlet neighbours, was introduced by Baddeley and Møller (1989). Note that two points can be iterated neighbours in one realization but not in another and are, therefore, dynamic neighbours.

Neighbours of neighbours are always iterated neighbours and therefore, we have proved the following:

Proposition: An interacting neighbour process with $\sim$-neighbours is a nearest neighbour Markov process with respect to the neighbourhood relation $\sim^2_{\mathbf{x}}$.

Interacting neighbour processes are not Ripley-Kelly Markov processes with respect to $\sim$ but they are Ripley-Kelly Markov with respect to a wider neighbourhood. For example, let two points ξ and η be neighbours if they are r-close, i.e. if $0 < d(\xi, \eta) \leq r$, where $d(\xi, \eta)$ is the distance between ξ and η. Then, an interacting neighbour process with r-close neighbours is not Ripley-Kelly Markov with r-close neighbours but is Ripley-Kelly Markov with $2r$-close neighbours. We would, however, have to make some restrictions on the interaction function since not all $2r$-close points are iterated neighbours. Therefore, it is more natural to consider the interacting neighbour processes as nearest neighbour processes instead of as Ripley-Kelly Markov processes with a wider neighbourhood.

The class of interacting neighbour processes is a subclass of the connected component Markov point processes (see e.g. Baddeley and Møller, 1989; Møller, 1998). The general form of the density of the connected component Markov process is

$$f_{CC}(\mathbf{x}) = \alpha \prod_{\mathbf{y} \in C(\mathbf{x})} \phi(\mathbf{y}),$$

where $C(\mathbf{x})$ is the set of all connected components in $\mathbf{x}$. If we choose

$$\phi(\mathbf{y}) = \prod_{x_i \in \mathbf{y}} \prod_{\mathbf{z} \subseteq N(x_i)} Q(x_i, z)$$

we obtain the interacting neighbour process.

There are reasons to consider this new class of processes which is somehow between Ripley-Kelly Markov processes and nearest neighbour Markov processes. First, we want to have a process where the basic neighbourhood with respect to $\sim$ is the most important set when considering interactions between the points. The connected component process allows iterated neighbours of infinite order and gives the same importance to all points in one component. Second, we want to introduce a class of processes where a set of neighbours of a point affects the point

simultaneously but where the set of points is not necessarily a clique. Therefore, Ripley-Kelly Markov processes (with respect to a given neighbourhood) are not wide enough a class. On the other hand, Ripley-Kelly Markov processes with a wider neighbourhood is a too wide class.

4 Full neighbourhood interaction processes

Whereas interaction due to a point of an interacting neighbour process can be factorized into interaction with subsets of its neighbourhood, we want to specify a class of processes which has the property that all neighbours of a point affect the point simultaneously. We call such processes full neighbourhood interaction processes.

Definition 2: A process is a *full neighbourhood interaction process* with a given neighbourhood relation $\sim$ if it has the density

$$f(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} \phi(x_i)\psi(x_i, N(x_i)), \tag{3}$$

where $\phi : A \to (0, \infty)$, $\psi : A \times \Omega_A \to (0, \infty)$, and where the *neighbourhood interaction function* ψ cannot be factorized.

Conditional intensity of the full neighbourhood interaction process is of the form

$$\lambda(\xi; \mathbf{x}) = \phi(\xi)\psi(\xi, N_{\mathbf{x}\cup\{\xi\}}(\xi)) \cdot \prod_{x_i \sim \xi} \frac{\psi(x_i, N_{\mathbf{x}\cup\{\xi\}}(x_i))}{\psi(x_i, N_{\mathbf{x}}(x_i))}. \tag{4}$$

It depends on ξ, neighbours of ξ and neighbours of neighbours of ξ. Therefore, the full neighbourhood processes are nearest neighbour Markov processes with respect to the neighbourhood relation $\sim^2_{\mathbf{x}}$.

Remark: It can be shown that the full neighbourhood interaction process is a specific case of the interacting neighbour process. For this it is enough to check that the following choice of the neighbour set interaction functions in the density (1),

$$Q(x_i) = \phi(x_i)$$

and

$$Q(x_i, \mathbf{z}) = \prod_{\emptyset \neq \mathbf{z}' \subseteq \mathbf{z}} \psi(x_i, \mathbf{z}')^{(-1)^{|\mathbf{z} \setminus \mathbf{z}'|}},$$

where $|\mathbf{z} \setminus \mathbf{z}'|$ denotes the number of points in $\mathbf{z} \setminus \mathbf{z}'$, leads to the density of the form (3).

Let us now give some examples of the full neighbourhood interaction processes.

Example 1: Let us first recall the saturation process introduced by Geyer (1998). For each point $x_i \in \mathbf{x}$, we define

$$t_{\mathbf{x}}(x_i) = \sum_{x_j \in \mathbf{x}} \mathbf{1}_{(0,r]}(d(x_i, x_j)),$$

where $d(x_i, x_j)$ is the distance between the points x_i and x_j and $\mathbf{1}(\cdot)$ is the indicator function. Therefore, $t_{\mathbf{x}}(x_i)$ is the number of points of $\mathbf{x}$ within distance r from the point x_i, and $\sum_{x_i \in \mathbf{x}} t_{\mathbf{x}}(x_i)$ is twice the number of r-close pairs which is the sufficient statistic used in defining the Strauss process. In the case of the saturation process, instead of adding up the $t_{\mathbf{x}}(x_i)$, $i = 1, ..., n(\mathbf{x})$, one puts an upper bound on the influence of any single point. Let $c > 0$ be an arbitrary constant, and define

$$u(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} \min\{c, t_{\mathbf{x}}(x_i)\}.$$

The density for the saturation process is then

$$f(\mathbf{x}) = \alpha b^{n(\mathbf{x})} \gamma^{u(\mathbf{x})}, \tag{5}$$

where $b > 0$ is connected to the intensity of the process, and $\gamma > 0$ is an interaction parameter. Values $\gamma < 1$ indicate regularity, $\gamma > 1$ clustering, and $\gamma = 1$ mutual independence between the points.

Given that two points are neighbours if they are r-close, the saturation process is a full neighbourhood interaction process since

$$f(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} b\gamma^{\min\{c, t_{\mathbf{x}}(x_i)\}} = \alpha \prod_{x_i \in \mathbf{x}} \phi(x_i)\psi(x_i, N(x_i)),$$

where $\phi(x_i) = b$ and $\psi(x_i, N(x_i)) = \gamma^{\min\{c, t_{\mathbf{x}}(x_i)\}}$, and $\psi(x_i, N(x_i))$ cannot be factorized.

It may be interesting to compare the saturation process and the Strauss process. The density of the Strauss process can be written as

$$f_{ST}(\mathbf{x}) = \alpha b^{n(\mathbf{x})} \gamma^{\frac{1}{2}\sum_{x_i \in \mathbf{x}} t_{\mathbf{x}}(x_i)}.$$

The neighbour statistic $\min\{c, t_{\mathbf{x}}(x_i)\}$ of the saturation process reaches and stays at constant level c while $\frac{1}{2}t_{\mathbf{x}}(x_i)$, the neighbour statistic of the Strauss process, increases (see Figure 1a). Therefore, if γ is greater than 1, the same is true for the interaction term $\gamma^{\frac{1}{2}\sum_{x_i \in \mathbf{x}} t_{\mathbf{x}}(x_i)}$. Here, only the clustered case is of interest because in the regular case both processes are most likely the same (if c is not too small) since the probability that the number of neighbours of a point being greater or equal to c is small, and therefore, saturation gives only a negligible effect.

Example 2: Let us now consider a process with the density

$$f(\mathbf{x}) = \alpha b^{n(\mathbf{x})} \gamma^{\sum_{x_i \in \mathbf{x}} \max\{0, t_{\mathbf{x}}(x_i)(c - t_{\mathbf{x}}(x_i))\}},$$

where b, $t_{\mathbf{x}}(x_i)$ and c are like in Example 1. This process is again a full neighbourhood interaction process with r-close neighbours: The density can be written as

$$f(\mathbf{x}) = \alpha \prod_{i=1}^{n(\mathbf{x})} b\,\gamma^{\max\{0, t_{\mathbf{x}}(x_i)(c - t_{\mathbf{x}}(x_i))\}} = \alpha \prod_{x_i \in \mathbf{x}} \phi(x_i)\psi(x_i, N(x_i)), \qquad (6)$$

where $\phi(x_i) = b$, $\psi(x_i, N(x_i)) = \gamma^{\max\{0, t_{\mathbf{x}}(x_i)(c - t_{\mathbf{x}}(x_i))\}}$ and $\psi(x_i, N(x_i))$ cannot be factorized.

Let us look at the process with density (6) in more detail. The neighbour statistic $\max\{0, t_{\mathbf{x}}(x_i)(c - t_{\mathbf{x}}(x_i))\}$ is plotted against $t_{\mathbf{x}}(x_i)$ in Figure 1b. In the case $\gamma > 1$ the interaction term $\prod_{x_i \in \mathbf{x}} \psi(x_i, N(x_i))$ gets its largest value as the number of neighbours equals $\frac{c}{2}$, i.e. the cluster size $\frac{c}{2} + 1$ is favoured. Therefore, we can control the size of clusters by choosing the constant c appropriately. Since for large values of the interaction parameter γ all clusters are of the same size, we call this process a *twin clusters process.*

The case $\gamma < 1$ is interesting as well. Then, the interaction term gets its smallest value when the number of neighbours of each point equals $\frac{c}{2}$ and its largest value if points have either no neighbours or c (or more) neighbours. Hence, instead of being a repulsive model, it is a specific cluster model: either isolated points (no neighbours) or clusters with c or more neighbours are favoured. Therefore, the process is a combination of a regular and a clustered process. We call it a *bipattern process* and believe that it is a suitable model when data looks heterogeneous consisting of some clusters superimposed on a regular pattern.

5 Simulation study

We carried out a simulation study to explore what type of clustering can be produced by the models in Examples 1 and 2. The bipattern model is included in the study even though it does not create purely clustered patterns but rather mixtures of regular and clustered patterns.

For comparison we considered also the continuum random cluster (CRC) process (Møller, 1998) which has the density function

$$f(\mathbf{x}) = \alpha b^{n(\mathbf{x})} \gamma^{-|C(\mathbf{x})|},$$

where $|C(\mathbf{x})|$ is the number of connected components in $\mathbf{x}$. The process is not a full neighbourhood interaction process but it is capable of generating clustered patterns.

In the simulation of the patterns from these four models, we used the Metropolis-Hastings algorithm with Metropolis updates and with a fixed number of points (Møller, 1998). Each simulated pattern consists of $n(\mathbf{x}) = 100$ points generated in the unit square mapped onto a torus. To obtain independence between the samples we used different runs for each sample. Before recording a sample we ran

1000-20000 (depending on the interaction parameter value) burn-in steps with a binomial realization as the initial pattern.

For all three full neighbourhood interaction processes we used the same values of the interaction radius $r = 0.08$, and the constant c was chosen to be 5 for the saturation and the bipattern models, and 12 for the twin clusters model. For the CRC model we fixed the radius of the discs to be equal to 0.02 which corresponds to the interaction radius 0.04. Only one set of values was used here but we think that even this simple experiment is able to show us something about the variety of clustered patterns that these models can produce as we let the interaction parameter γ vary (Figures 2-5).

The saturation model (Fig. 2) seems to be able to create clustered patterns of large variety: from slight to dense clustering. Unlike the saturation model the twin clusters model (Fig. 3) seems to create clustered patterns with clusters that are clearly separated from each other. For values of the interaction parameter γ close to 1 the difference between the saturation model and the twin clusters model is not significant. As was expected, the bipattern model (Fig. 4) enables us to generate point configurations combining some features of regular and clustered patterns. The CRC model (Fig. 5), instead, produces clearly clustered patterns only with high values of the interaction parameter. Moreover, clusters generated by the CRC model are not necessarily dense and they have a large variety of shapes.

Figures 2-5 show only one realization of each type and therefore they can not tell us anything about the variation of the different models. Properties of the models can be summarized in the distribution of the number of r-close neighbours. Figure 6 presents histograms averaging the empirical distributions obtained from 1000 generated patterns of each type. As the value of the interaction parameter of the saturation process increases, the more likely it is for a point to have more than c neighbours. The histogram of the twin clusters model, instead, is almost symmetrical with the highest frequency at $\frac{c}{2}$. Thus, this model seems really to be able to create clusters of any desired size. Furthermore, as the interaction of the bipattern process gets stronger the distribution of the number of neighbours becomes clearly bimodal with peaks at the values 0 and c. This shows empirically that this model is a mixture of regular and clustered models.

Whereas all three full neighbourhood interaction processes form clusters of a specific size, the CRC process (with fixed value of γ) seems to produce patterns with clusters of various size and shape. It may be due to the fact that all points in one component are neighbours even though they can be far away from each other.

Calculation of Ripley's K-function (see e.g. Diggle, 1983) shows the difference between the second-order properties of the models. Here, we use a modified version of the K-function, namely $L^*(r) = \sqrt{K(r)/\pi} - r$. The values close to zero indicate that the process is close to Poisson, positive values clustering and

negative values regularity. Figure 7 shows averages of L^*-functions calculated from 1000 realizations. The L^*-functions of the saturation and the twin clusters processes have very similar shape indicating clustering for the chosen set of parameters. The cusp points of the functions correspond to the value of the interaction radius. Note that the slope after the cusp point is steeper for the latter model. The L^*-function of the bipattern model does not have a clear cusp point and it decreases very slowly. An interesting feature of the CRC model is that the mean size of clusters, and hence, the effective radius of interaction increases as the interaction parameter γ increases.

6 Application

To show how the full neighbourhood interaction processes can be applied for the analysis of real data we chose to analyze a data set which has previously been studied in the literature by using other models. The data set consists of locations of 126 Scots pine saplings in a square plot of 10 x 10 m^2 (Figure 8a). The data have been analyzed by Penttinen *et al.* (1992) and by Stoyan and Stoyan (1994) who fitted Matern's cluster processes, and by Møller *et al.* (1998) who fitted the log Gaussian Cox process to the data.

First, we fitted the saturation model to the Scots pine data. The interaction radius was chosen to be 0.65 m based on the L^*-function estimated from the data (solid line in Figure 9a) and the saturation parameter $c = 8$ based on the histogram of the number of neighbours. The Monte Carlo maximum likelihood (MCML) method introduced by Geyer and Thompson (1992) suggests 1.19 as an estimate for the unknown parameter γ indicating that there is slight attraction between the points. Edge effects were taken into account by considering them as a missing data problem as was suggested by Geyer (1998).

To check how well the data can be modelled by the saturation process we plotted the empirical L^*-function of the data together with the envelopes (dashed and dotted) from 200 simulations of the estimated model (Figure 9a), and the empirical distribution $\hat{G}$ of the nearest neighbour distances (Figure 9b) together with the envelopes calculated also from 200 simulated patterns. The L^*-function study shows a reasonably good fit except in small distances, where the attraction between the pines is stronger than according to the fitted model. The same conclusion can be drawn from the plot of the G-functions. The empirical curve is above the upper envelope in small distances (up to 0.2-0.3 m).

In order to take the stronger attraction in small distances into account we modificated the saturation model by including another interaction parameter which describes short range interactions. The new model has the density

$$f(\mathbf{x}) = \alpha b^{n(\mathbf{x})} \gamma_1^{u_1(\mathbf{x})} \gamma_2^{u_2(\mathbf{x})},$$

where

$$u_1(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} \min\{c_1, \sum_{x_j \in \mathbf{x}} \mathbf{1}_{(0,r_1]}(d(x_i, x_j))\}.$$

and

$$u_2(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} \min\{c_2, \sum_{x_j \in \mathbf{x}} \mathbf{1}_{(r_1,r_2]}(d(x_i, x_j))\}.$$

The sufficient statistics u_1 and u_2 are built up by summing up the number of r_1-close neighbours of each point bounded by c_1 and the number of neighbours between distances r_1 and r_2 away of each point bounded by c_2, respectively. Therefore, neighbours that are further away (but still neighbours) contribute differently from the very close neighbours.

With the choice $r_1 = 0.2$, $r_2 = 0.65$ and constants $c_1 = 3$, $c_2 = 5$ we obtained MCML estimates $\hat{\gamma}_1 = 9.97$ and $\hat{\gamma}_2 = 1.26$. Therefore, clustering is very strong in small distances. The graph of the envelopes from 200 patterns simulated by the estimated model leads to the conclusion that the model fits to the data much better. A simulated realization of the fitted model is given in Figure 8b. Note that the plot of G-functions (Figure 9b) reveals that both estimated models overestimate the nearest neighbour distribution for the distance range 0.6-1.0 m which means that besides clustering the data exhibit slight regularity. Therefore, the bipattern process could be an appropriate model. However, according to our experiments, the bipattern process fits to the data poorly because it overestimates the number of isolated points.

7 Discussion

We have introduced a new class of point processes, interacting neighbour processes, which includes Ripley-Kelly Markov processes (with a given neighbourhood) and which itself is a subset of the class of nearest neighbour Markov processes. Characteristic for the processes is that subsets of the neighbourhood of a point, which are not necessarily cliques, affect the point simultaneously.

The full neighbourhood interaction processes are a subclass of interacting neighbour processes, where all neighbours (not subsets of the neighbourhood) affect a point simultaneously. The neighbourhood interaction function ψ depends on the realization but can be written by means of neighbour set interaction functions that do not depend on the realization. Some full neighbourhood interaction processes seem to be good models for clustering. The saturation process as well as the twin clusters process is able to produce clustered patterns in large variety. The bipattern process creates combinations of clustered and regular patterns which can be interesting from the point of view of applications.

Acknowledgements

We would like to thank Docent Antti Penttinen for his valuable comments on the manuscript. The financial support by Malmstensmedlen, which made it possible to invite Pavel Grabarnik to the Department of Statistics at Göteborg University, is greatly appreciated. This work was supported by the RFBR research grant 97-04-48684 as well as by the EU grant ERB-FMRX-CT96-0096.

References

Baddeley, A.J. and van Lieshout, M.N.M. (1995). Area-interaction point processes, *Ann. Inst. Statist. Math.*, **47**, 601–619.

Baddeley, A.J. and Møller, J. (1989). Nearest-neighbour Markov point processes and random sets, *Int. Statist. Rev.*, **57**, 89–121.

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns.* London: Academic Press.

Geyer, C.J. (1998). Likelihood Inference for Spatial Point Processes. In *Current Trends in Stochastic Geometry and Applications.* Ed. W.S. Kendall, M.N.M. van Lieshout and O.E. Barndorff-Nielsen. Chapman and Hall, London. To appear.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society B* **54**, 657–699.

Kendall, W.S. (1990). A spatial Markov property for nearest-neighbour Markov point processes. *J. Appl. Prob.* **28**, 767–778.

van Lieshout, M.N.M. and Molchanov, I.S. (1997). Shot noise weighted processes: a new family of spatial point processes. To appear in *Stochastic Models.*

Møller, J. (1998). Markov chain Monte Carlo and spatial point processes. In *Current Trends in Stochastic Geometry and its Applications.* Ed. W.S. Kendall, M.N.M. van Lieshout and O.E. Barndorff-Nielsen. Chapman and Hall, London. To appear.

Møller, J., Syversveen, A.R. and Waagepetersen, R.P. (1998) Log Gaussian Cox processes. To appear in *Scandinavian Journal of Statistics.*

Penttinen, A. and Stoyan, D. and Henttonen, H. (1992). Marked point processes in forest statistics, *Forest Sci.*, **38**, 806–824.

Ripley, B.D. and Kelly, F.P. (1977). Markov point processes. *Journal of the London Math. Soc.*, **15**, 188–192.

Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes and point fields.* John Wiley & Sons, Chichester.

Labels of Figures

Figure 1: The neighbour statistics a) $\min\{c, t_{\mathbf{x}}(x_i)\}$ connected to the saturation process (solid), $t_{\mathbf{x}}(x_i)$ connected to the Strauss process (dashed) and b) $\max\{0, t_{\mathbf{x}}(x_i)(c-t_{\mathbf{x}}(x_i))\}$ connected to the twin clusters and the bipattern process are plotted as a function of $t_{\mathbf{x}}(x_i)$ and with $c = 4$.

Figure 2: Simulated realizations of the saturation process with $\ln\gamma$ equal to a) -0.2 b) -0.3 c) -0.4 d) -0.7 e) -1.0 and f) -2.0.

Figure 3: Simulated realizations of the twin clusters process with $\ln\gamma$ equal to a) -0.04 b) -0.08 c) -0.10 d) -0.15 e) -0.20 and f) -0.30.

Figure 4: Simulated realizations of the bipattern process with $\ln\gamma$ equal to a) 0.15 b) 0.20 c) 0.30 d) 0.40 e) 0.70 and f) 1.0.

Figure 5: Simulated realizations of the CRC process with $\ln\gamma$ equal to a) 1.0 b) 1.5 c) 2.0 d) 2.5 e) 3.0 and f) 5.0.

Figure 6: Frequencies of the number of r-close neighbours averaged over 1000 simulated patterns of each process. The interaction radius equals 8.

Figure 7: L^*-functions calculated from 1000 simulations of the a) saturation process: $\ln\gamma$ = -0.2, -0.3, -0.4, -0.7, -1.0 and -2.0; b) twin clusters process: $\ln\gamma$ = -0.04, -0.08, -0.10, -0.15, -0.20 and -0.30; c) bipattern process: $\ln\gamma$ = 0.15, 0.20, 0.30, 0.40, 0.70 and 1.00; and d) CRC process: $\ln\gamma$ = 1.0, 1.5, 2.0, 2.5, 3.0 and 5.0. The values of the parameters $\ln\gamma$ correspond to the L^*-curves from the bottom to the top.

Figure 8: a) Locations of Scots pines. b) A realization of the fitted modified saturation model.

Figure 9: a) Empirical L^*-function (solid) together with the lower and upper envelopes calculated from 200 simulated patterns of the fitted saturation model (dashed and dotted) and the fitted modified saturation model (dashed). b) Empirical nearest neighbour distribution (solid) and the envelopes calculated from 200 simulated patterns of the fitted saturation model (dashed and dotted) and the fitted modified saturation model (dashed).

Fig. 1

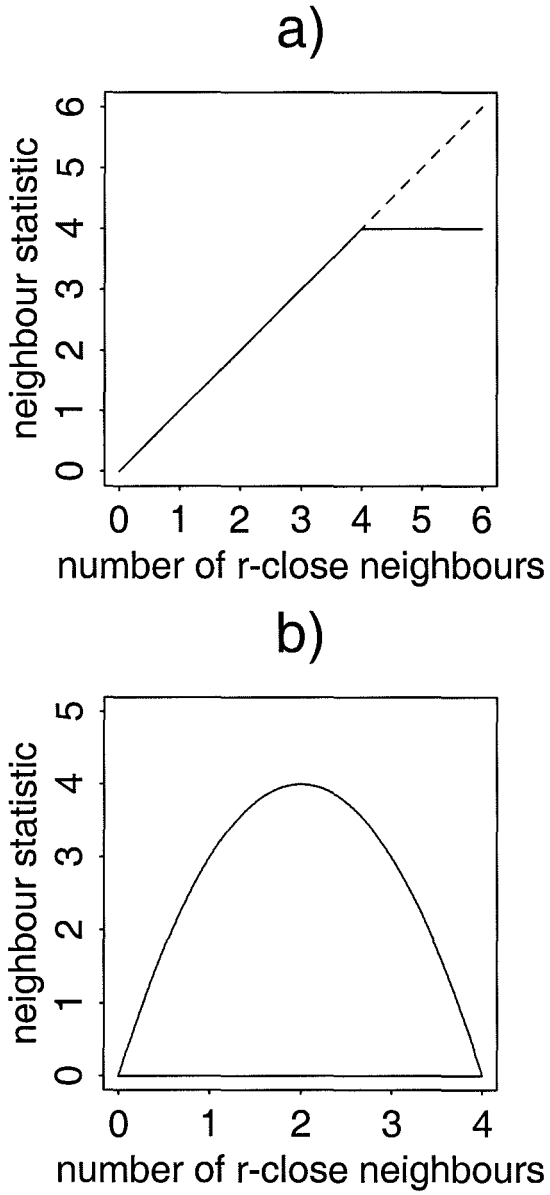

a)
neighbour statistic
0 1 2 3 4 5 6
0 1 2 3 4 5 6
number of r-close neighbours
b)
neighbour statistic
0 1 2 3 4 5
0 1 2 3 4
number of r-close neighbours

Fig. 2

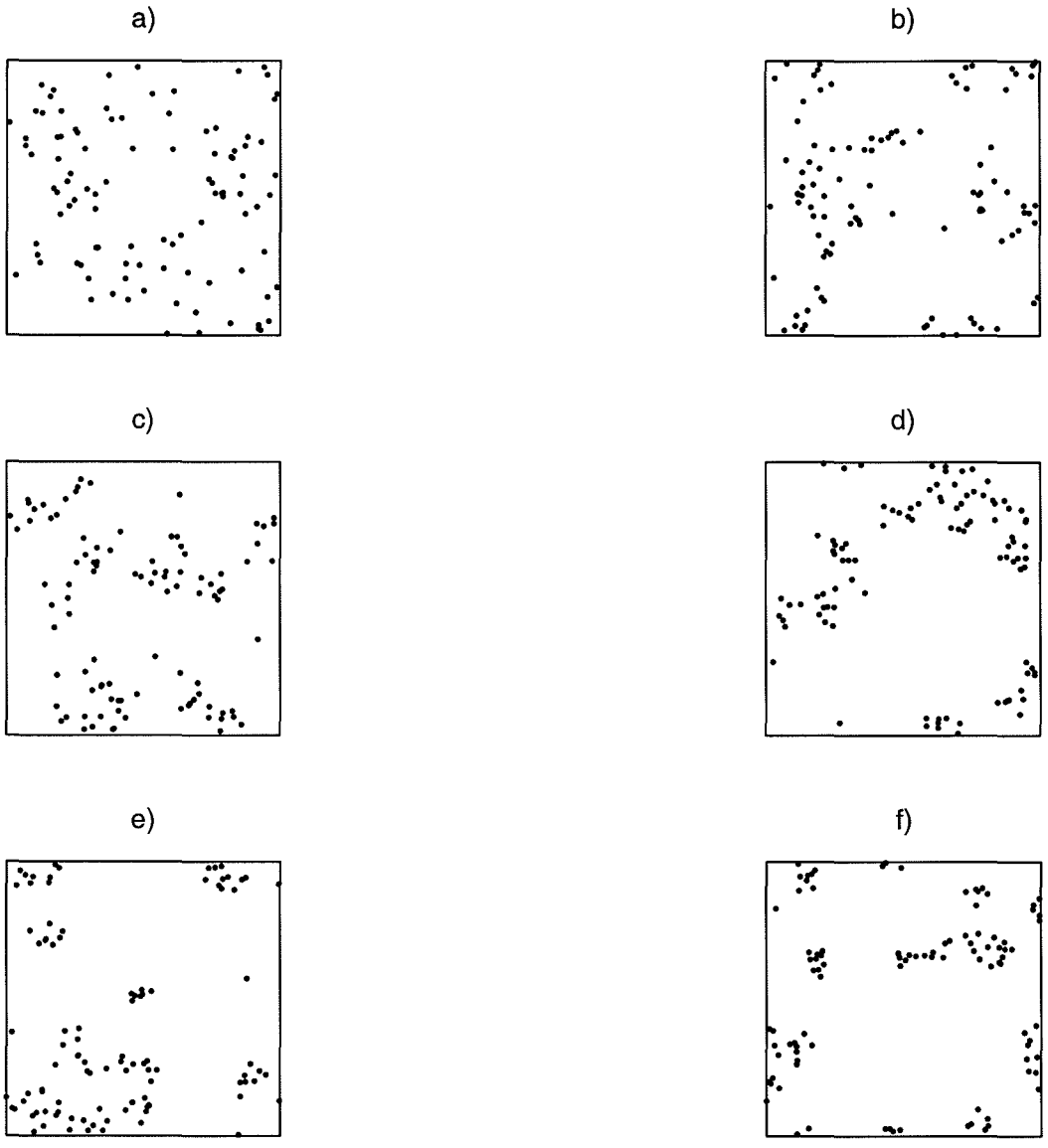
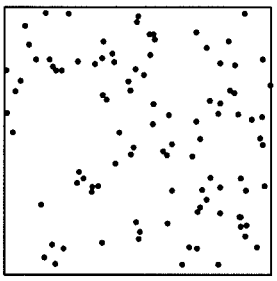
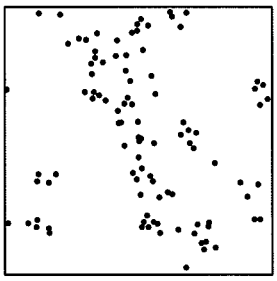


Fig. 3

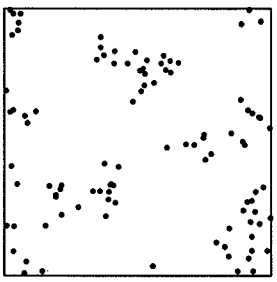
a)



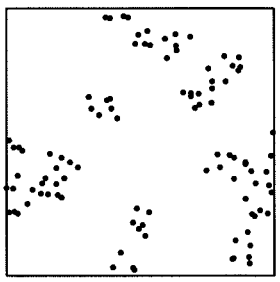
b)



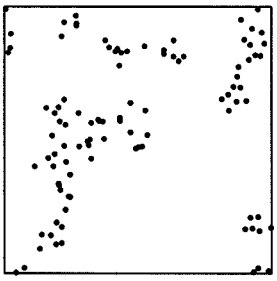
c)



d)



e)



f)

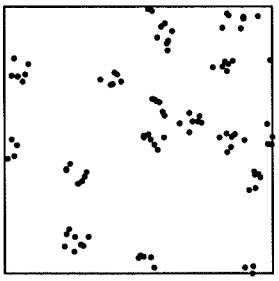
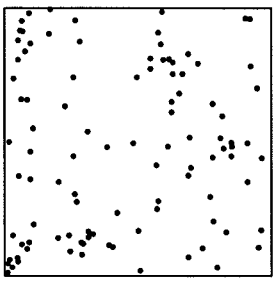
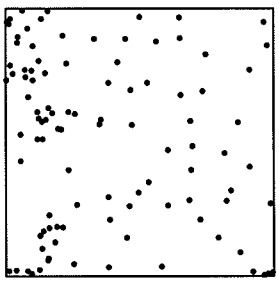


Fig. 4

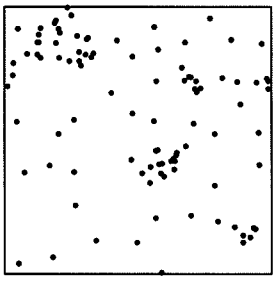
a)



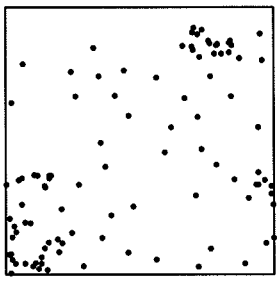
b)



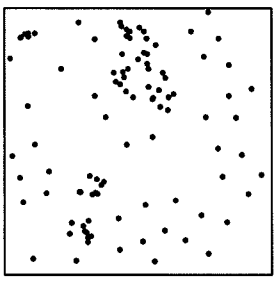
c)



d)



e)



f)

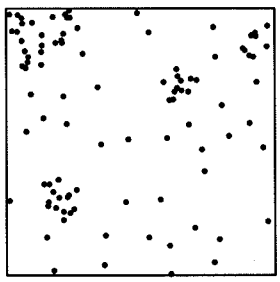
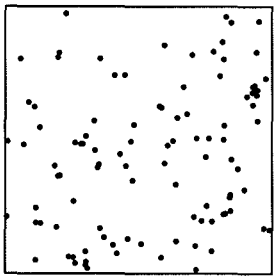
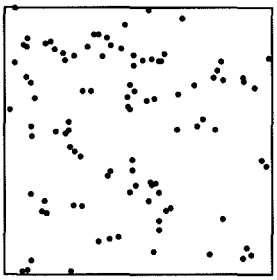


Fig. 5

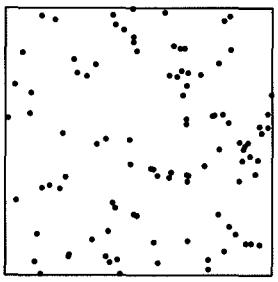
a)



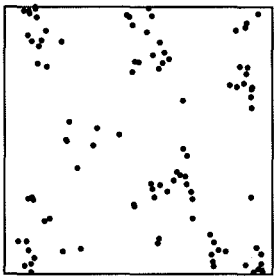
b)



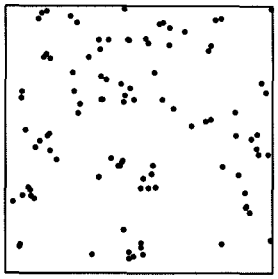
c)



d)



e)



f)

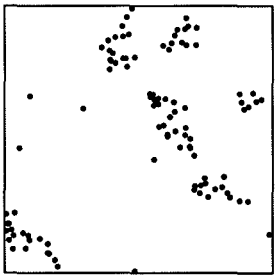
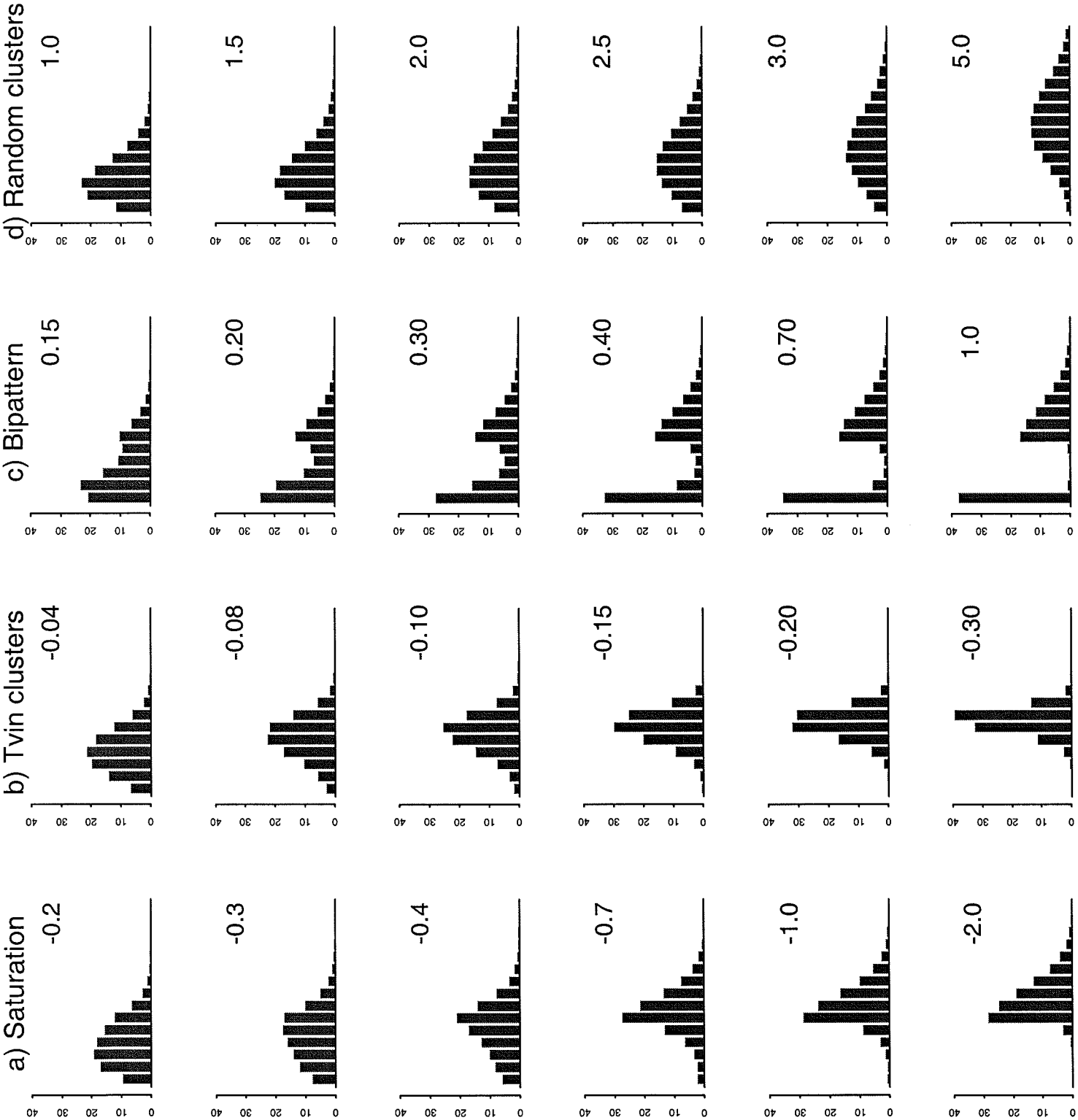


Fig. 6


a) Saturation
-0.2
-0.3
-0.4
-0.7
-1.0
-2.0
b) Tvin clusters
-0.04
-0.08
-0.10
-0.15
-0.20
-0.30
c) Bipattern
0.15
0.20
0.30
0.40
0.70
1.0
d) Random clusters
1.0
1.5
2.0
2.5
3.0
5.0
0
10
20
30
40

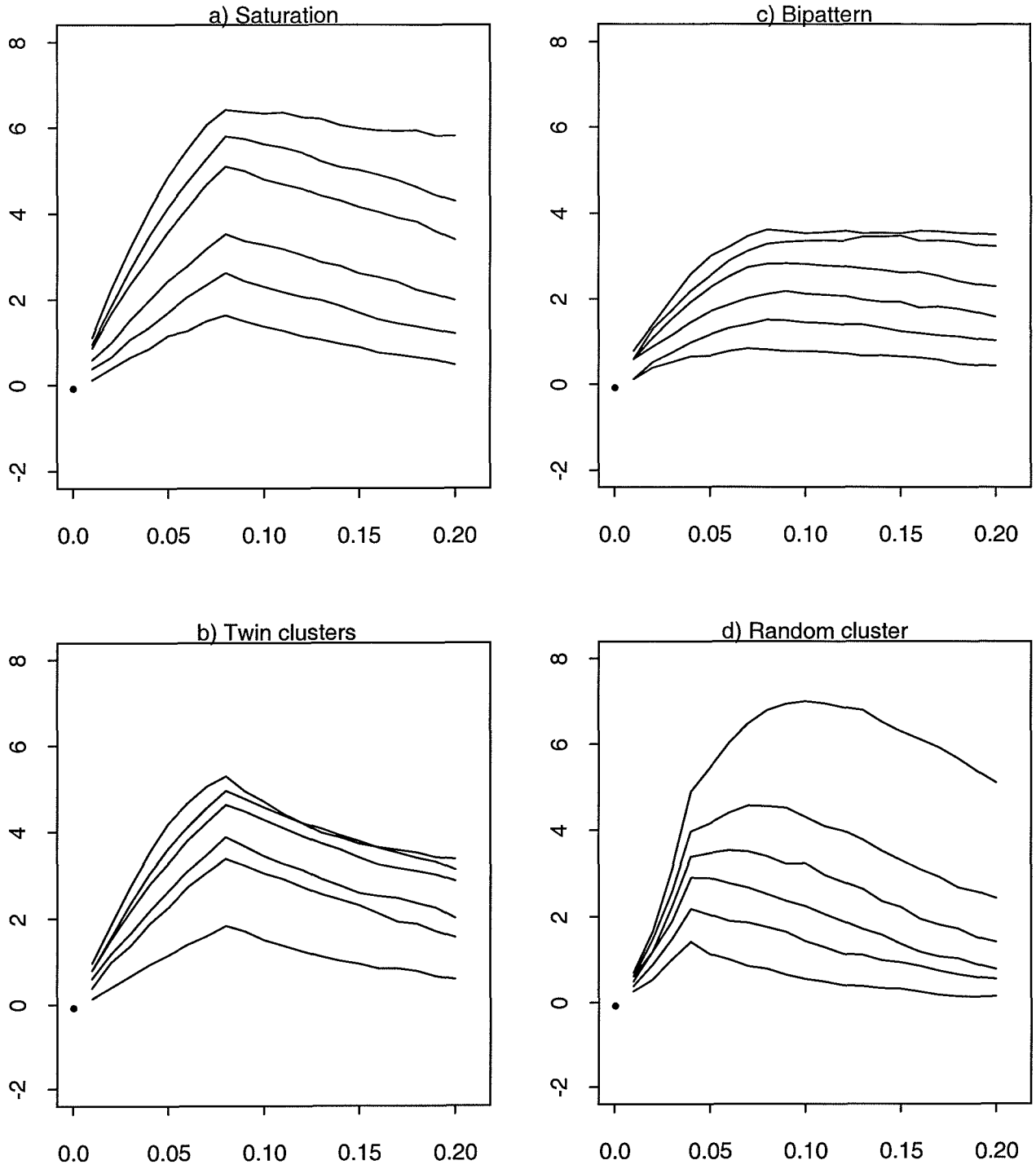

a) Saturation
c) Bipattern
b) Twin clusters
d) Random cluster
8
6
4
2
0
-2
0.0
0.05
0.10
0.15
0.20

Fig. 4

Fig. 8

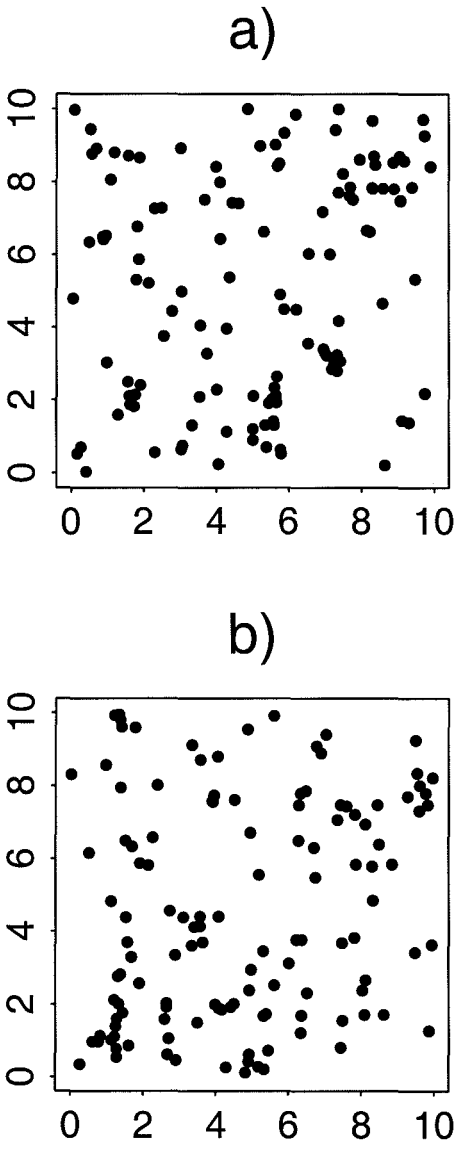


Fig. 9

a)
L*(r)
-0.1
0.0
0.1
0.2
0.3
0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
r
b)
G(r)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
r

Research Report

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.
1998:1	Särkkä, A. & Högmander, H.:	Multiple spatial point patterns with hierarchical interactions.
1998:2	Frisén, M. & Wessman, P.:	Evaluations of likelihood ratio methods for surveillance. Differences and robustness.
1998:3	Järpe, E.:	Surveillance of spatial patterns. Change of the interaction in the Ising model.
1998:4	Pettersson, M.:	Evaluation of some methods for statistical surveillance of an autoregressive process.
1998:5	Pettersson, M.:	On monitoring of environmental and other autoregressive processes.
1998:6	Dahlbom, U.:	Least squares estimates of regression functions with certain monotonicity and concavity/convexity restrictions.
1998:7	Dahlbom, U.:	Variance estimates based on knowledge of monotonicity and concavity properties.