University of Wisconsin — Madison
Department of Statistics

Methods for Estimating the Interaction Parameter of Strauss Disc Processes

Elaine Borghi
Brian S. Yandell
Department of Statistics

Technical Report # 1019
March 2000

Department of Statistics
University of Wisconsin–Madison
1210 West Dayton Street
Madison WI 53706-1685

Phone: 608/262-2598
Fax: 608/262-0032
Internet: yandell@stat.wisc.edu

Methods for Estimating the Interaction Parameter of Strauss Disc Processes

Elaine Borghi [1]
Universidade Estadual de Campinas

Brian S. Yandell
University of Wisconsin

Abstract

The Strauss disc process can potentialy model many applications in biological systems, physics, environmental science and other fields. We perform a simulation study to compare methods of estimating the interaction parameter of the density of this process. We considered minimum-contrast, pseudo-likelihood and Takacs-Fiksel estimators and for this case a pseudo-likelihood estimator presented better performance compared to the others.
Keywords : Strauss disc process, Metropolis-Hasting algorithm, pseudo-likelihood, minimum-contrast, Takacs-Fiksel.

1 Introduction

Many of the applied problems that are related to spatial statistics fit inside a special class of point processes, the pairwise interaction point processes. This particular case of Gibbs point process is widely used in applications not just because they efficiently model many patterns that appear in physics, mechanics, environmental sciences and other fields (e.g. Percus, 1964; Penttinen *et al.*, 1992), but also because they have few parameters to be estimated.

There is a considerable literature related to the inference on pairwise interaction point processes in general. Various methods were proposed that apply to more general Gibbs point processes. Maximum likelihood for spatial point processes was studied by Ogata and Tanemura (1981), and later extended to marked point processes (Ogata and Tanemura,

[1] *Address for correspondence* : Elaine Borghi, Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola, Cidade Universitária Zeferino Vaz, Cx.Postal 6011, Campinas SP CEP:13026-430, Brazil
e-mail : elaine@agr.unicamp.br

1985). Fiksel (1984) generalized the idea earlier introduced by Takacs and proposed the Takacs-Fiksel method of estimating parametrized pair potentials of gibssian point processs. The pseudo-likelihood method was considered by Jensen and Møller (1991) for general Markov spatial point processes. Later, Goulard *et al.* (1996) investigated this method for marked Gibbs point processes using a pseudo-likelihood approach, and worked on cases like the bivariate Gibbs point process and the Strauss disc process. The pseudo-likelihood method is a particular case of the Takacs-Fiksel method for suitable test functions. Some non-parametric estimation methods were proposed in the past few years but we will not cover them in the present paper.

One approach that we consider is a very simple and well known method, the method of minimum-contrast. Although there are many functions that can be used for this method, here we consider the reduced second moment measure K-function. This method appears frequently in the literature, especially for goodness-of-fit tests or exploratory analysis (Stoyan and Stoyan, 1994; Stoyan *et al.*, 1987; Cressie, 1991).

Although the literature is broad in this subject, studies of properties of different estimators were not throughout investigated for the marked case, especially in terms of comparison of performance. Diggle *et al.* (1994) investigates three methods of estimation for pairwise interaction point processes, maximum likelihood, pseudo-likelihood and Takacs-Fiksel, presenting a comparative simulation study. In the same direction, we perform a simulation study for comparison of estimation methods for the particular case of the Strauss disc process in this paper. Properties and the behavior of each of the estimators for the interaction parameter are investigated.

2 The Strauss disc process

Random disc processes (Baddeley and Møller, 1989) are special cases of the germ-grain model (Stoyan *et al.*, 1987) which can be used to describe a pattern of randomly distributed discs (or spheres in higher dimensional case), where discs may overlap.

Consider a marked point process $\Psi = \{[x_i, r_i]\}$ with finite number of points on a bounded space S and marks on $\Delta \subset [0, \infty)$, where the density is the Radon-Nikodym derivative with respect to the distribution of a homogeneous Poisson process (which, with no loss of generality, can be taken to have unit rate). Let $x = (x_1, ..., x_n)$,

$r = (r_1, ..., r_n)$, and consider $[x_i, r_i]$ to be a disc with radius r_i centered at x_i. A Strauss disc process has density

$$f(x,r) = z\beta^n \gamma^{s(x,r)}, \tag{1}$$

where β is a parameter related to the intensity of the point process, γ is the interaction parameter, n is the number of points in Ψ and $s(x,r)$ is the number of pairs of points x_i and x_j that are closer than $r_i + r_j$ apart in Ψ.

3 Simulation of the Strauss disc process

Statistical inference for spatial point processes can be very complicate, which explains why Markov chain Monte Carlo methods are being used extensively in this field. Some of these methods of simulating a Markov chain are conditional on the number of points n, what makes them simple in terms of computational aspects (Ripley, 1979b). Methods that are unconditional on the number of points seem to be more complicate but the choice must be made depending on the goal of the study. For the present research, we have to generate Strauss disc processes and conditioning on the number of points could interfere in an undesirable way on the final distribution of the process, based on the theory involved. For these reasons, we decided on an approach that is unconditional on the number of points.

3.1 A Metropolis-Hastings algorithm

Geyer and Møller (1994) suggested a simulation procedure for the unconditional case using a Metropolis-Hastings algorithm (Hastings, 1970). This algorithm is simpler than the spatial birth-and-death technique and can be used for Markov chain Monte Carlo methods. We now give a brief review of it.

For an usual finite point process defined on a bounded Borel set $S \subset \Re^d$ equipped with the Borel σ-field and Lebesgue measure ν restricted to S, suppose the point process has density π on $(\Omega, \mathcal{F})$ which is absolutely continuous with respect to τ, the density of the usual homogeneous Poisson process with intensity 1 restricted to S. The function $f = d\pi/d\tau$ is the density of the point process with respect to the Poisson process.

Informally, $\Omega = \bigcup_{n=0}^{\infty} \Omega_n$, where $\Omega_n = \Big\{ \{x_1, \cdots, x_n\} \subset S \Big\}$ is the set of all point configurations of n (not necessarily distinct) points included in S. Geyer and Møller (1994) proposed for this Metropolis-Hastings algorithm the transition kernel to be a mixture of two transition kernels Q_0, controlling displacements, and Q_1, controlling deletion and addition of points,

$$Q_p(F \mid x) \; = \; (1-p)\, Q_0(F \mid x) + p\, Q_1(F \mid x), \quad 0 \leq p \leq 1,$$

for any $F \in \mathcal{F}$. This transition kernel satisfies the condition for the Markov chain to be time reversible, given that Q_0 and Q_1 satisfy the same condition.

$Q_0(. \mid x)$ is concentrated on $H_n = \Omega_n \cap H$ with $H = \{x \in \Omega_n \mid f(x) > 0\}$ and $Q_1(. \mid x)$ is constructed so that it is concentrated on $H_{n-1} \cup H_n \cup H_{n+1}$ (or $H_0 \cup H_1$ if $n = 0$). In this way, with probability $q(x)$, we generate a new point ξ from some density $b(x, \xi)$ with respect to $\nu(d\xi)$, and with probability $1 - q(x)$ we either delete a random point $\eta \in x$ which is selected with some probability $d(x, \eta)$, or if $n = 0$ we do nothing.

According to the authors, the time reversibility holds if, denoting the acceptance probability by A_1, and

$$q(x \cup \xi) < 1, \;\; d(x, \xi) > 0, \;\; q(x) > 0, \;\; b(x, \xi) > 0 \;\text{ whenever }\; x \cup \xi \in H,$$

$$A_1(x \mid x \cup \xi) = \begin{cases} \min\{1, 1/r(x,\xi)\} & \text{if } \; x \cup \xi \in H, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$A_1(x \cup \xi \mid x) = \begin{cases} \min\{1, r(x,\xi)\} & \text{if } \; x \cup \xi \in H, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$r(x, \xi) = \frac{f(x \cup \xi)}{f(x)} \; \frac{1 - q(x \cup \xi)}{q(x)} \; \frac{d(x, \xi)}{b(x, \xi)} \;\; \text{if } \; x \cup \xi \in H.$$

Geyer and Møller considered one of the simplest situations for some cases of point processes like the Strauss process, when

$$q(\cdot) \equiv \frac{1}{2}, \;\; b(\cdot, \cdot) \equiv \frac{1}{\nu(S)} \;\text{ and }\; d(x, \cdot) \equiv \frac{1}{n+1} \;\text{ if }\; x \in \Omega_n.$$

In our case, for the Strauss disc process, we have to consider the mark distribution μ_1 (for the radii), which we will consider here as the "primary" mark distribution (as in Stoyan and Stoyan (1994), Section 16.3). So we have the probability of birth of a disc given by

$$b(\cdot,\xi) \equiv \frac{\mu_1(\xi)}{\nu(S)},$$

and

$$\frac{f(x \cup \xi)}{f(x)} \;=\; \beta\,\gamma^{s(\xi)}\,\mu(\xi),$$

where $s(\xi)$ is the number of discs in the process that overlap with a disc centered at ξ and β and γ are the parameters of the Strauss disc process (as in Equation 1). Note that, for the density of the points, the mark distribution μ may be different from μ_1, but we do not have control over μ. Therefore, we initially consider them to be equal, so that the mark distributions cancel. Later we examine how robust this is through simulations. A relation between μ and μ_1 is given by a mean-relation with respect to the Palm distribution (Stoyan and Stoyan, 1994).

Using similar notation introduced in Section 2, we denote $\Omega_n = \Big\{ \{x_1, \cdots, x_n\} \subset S \Big\}$ and $\Delta_n = \Big\{ \{r_1, \cdots, r_n\} \subset [0,\infty)^n \Big\}$.

The algorithm is constructed such that, given that the current state of the Markov chain is $\Psi = \{(x_1, r_1), \cdots, (x_n, r_n)\}$, we generate the next proposal Ψ' as follows.

If $n = 0$ then the proposal is either

(a) remain in the point configuration, or

(b) become a single disc with uniformly distributed center in S and randomly chosen radius from μ_1.

and if $n \geq 1$ then the proposal is either obtained by

(c) replacing a randomly picked disc in Ψ by a disc with uniformly distributed center in S and randomly chosen radius from μ_1, or

(d) deleting a randomly picked disc in Ψ, or

(e) adding a new disc with uniformly distributed center in S and randomly chosen radius from μ_1 to Ψ.

Here (a) and (b) occur with probabilities $1 - p/2$ and $p/2$, respectively, while (c), (d),(e) have probabilities $1 - p$, $p/2$, $p/2$, respectively. Note that $0 \leq p \leq 1$ and is related to the optimization of the simulation process. Then in any of the cases, the proposal Ψ' is accepted with probability

$$A(\Psi' \mid \Psi) = min\left(1, \frac{\{\nu(S)\ \beta\ \gamma^{s(\xi)}\}^{m-n}\ n!}{m!}\right),\ \text{if}\ \Psi' \in \Omega_m \times \Delta_m.$$

This is called a basic step of the chain. Geyer and Møller (1994) investigated this algorithm for the Strauss point process (unmarked), among others. They reported that in this case $p = 1$ (discarding replacements) is optimal. Also, according to the authors, the Markov chain appears to reach equilibrium, starting from the empty state, in fewer than 2,500 basic steps, although they consider a "burn-in" period of 40,000 basic steps to be safe. They considered a 200 basic step spacing between samples.

3.2 Study of the simulation of Strauss disc processes

There are some aspects of this simulation method that have not been investigated as far as we know for the specific case of a Strauss disc process. It would be worth exploring the properties of the simulation method in order to get some feeling about the behavior of, for example, the intensity of the processes generated or the mark distribution, as we vary the interaction parameter γ. Because we are going to focus on the effect of γ over the number of points of the generated Strauss disc process, we fix $\beta = 50$. In some of our preliminary investigations, we noticed that, if β is too big with the mean radii we are considering, the final mark distributions of the generated patterns are skewed with respect to the "primary" distribution.

On the other hand, if β is too small and we have a too sparse pattern, it is hard to detect any effect of changes in the parameter γ.

Although we know that choosing the an optimal "burn-in" period and spacing between the samples is very useful and important, we are also aware of the time that it would take for us to investigate those. As an attempt, we used the information given by Geyer and Møller (1994) in the Strauss point process case and do some diagnostic using simple Time Series techniques. In some of the problems we investigated, 40,000 basic steps became too expensive in terms of computational time, so we decreased the "burn in" period to 20,000, and we always used $p = 1$. Choosing the number of basic steps betwen samples equal to 200,

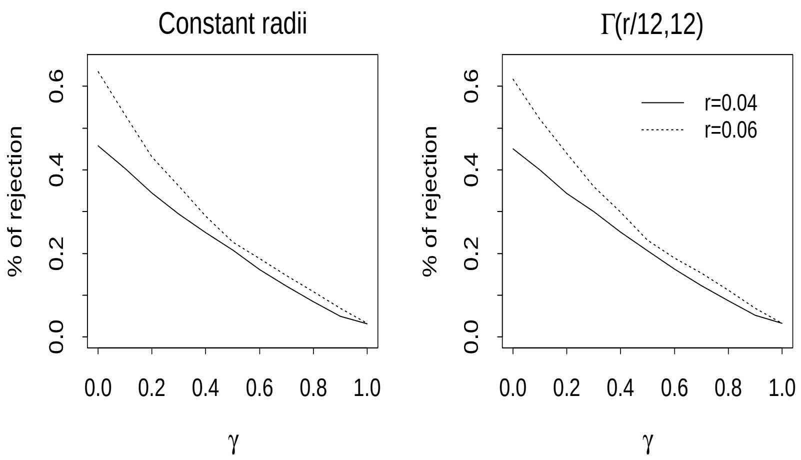

Figure 1: Percentage of rejection versus γ for the simulated patterns

we found that the autocorrelation function gives us a weak correlation (between -0.2 and 0.2 for most of the cases, just getting close to its maximum absolute value, 0.4, when γ gets closer to 1).

The estimate of the probability of rejection is shown in Figure 1. The behavior seems no to be different for the considered distributions. The percentage of rejection decreases as γ increases and it is overall higher for bigger mean radius.

The random number generator used was from Press *et al.* (1994), the subroutine RAN2. In order to reduce the boundary effects in our simulation studies, we simulated the point patterns in a bigger window (square $(-0.5, 1.5) \times (-0.5, 1.5)$), although our region of interest was always considered to be the unit square in $\Re^2$.

Figure 2 shows some examples of Strauss disc patterns generated using this algorithm for different γ values.

In order to look at the behavior of the intensity of the patterns from the simulations, we have in Figure 3 box-plots for 100 patterns for γ varying from zero, non-intersecting disc process, to one, when we have a Boolean model (homogenous Poisson process with independent disc radii). We considered the case where the radii have $\Gamma(r/12, 12)$ "primary" distribution for mean radius r equal to 0.04 and 0.06. We can see that, as r increases, the slope of the curve increases, which means we have a bigger change in the number of points with respect to change in the parameter γ when r is bigger. We have some indication that the intensity of the Strauss disc processes does not have a linear

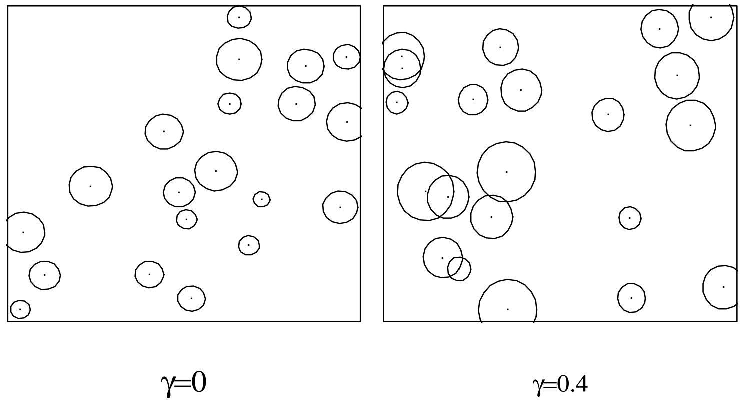

Figure 2: Examples of Strauss disc patterns for a $\Gamma(0.005, 12)$ mark distribution (mean radius r equal to 0.06) on an unit square ($d = 0.06$,$\beta = 50$).

form with respect to γ, but we did not investigate further about this issue.

As we initiate the process of generating marks from the "primary" mark distribution μ_1, we expect that the final distribution of the marks μ in the pattern is going to be skewed with respect to μ_1, for small values of γ. As showed in Figure 1, big radii are going to be more likely to be rejected than small radii. As we approach the Boolean model this does not happen.

Figures 4 and 5 show the Monte Carlo empirical mark distribution for a sample of 50 Strauss disc patterns and the theoretical "primary" mark distribution for different values of γ and r, for $\beta = 50$ for uniform and gamma mark distributions, respectively. We notice that the mark distribution gets more skewed when γ is equal to zero or as r increases. For the uniform "primary" distribution this effect is even bigger, given the shape of the distribution. For the gamma distribution this effect becomes smaller. This problem of getting a final mark distribution μ too far from the "primary" mark distribution μ_1 may affect some of our simulation studies when we have to guess the mark distribution.

Overall, this method of simulation seems to work well for Strauss disc processes if we are careful not to get too close to the extreme cases.

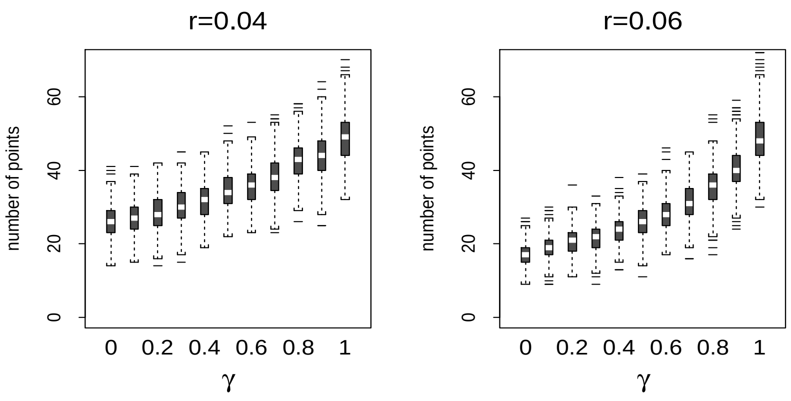

Figure 3: Box-plot of the number of points in 100 simulated patterns across γ with radii distribution $\Gamma(r/12, 12)$ and $\beta = 50$.

4 Estimation of γ

The Strauss disc process depends basically on parameters β, which is related to the intensity of the process, γ, the so-called interaction parameter, and a distribution of the marks (radii) μ. We concentrated our effort on the estimation of γ, fixing β and μ. We considered known methods including minimum-contrast, Takacs-Fiksel and pseudo-likelihood for estimating γ. We describe each method and present respective simulation study results.

For all the simulations performed in this chapter we used the Metropolis-Hastings algorithm described in Section 3.1. We fixed the parameter β to be equal to 50 and considered as mark distributions either uniform in $[2r/3, 4r/3]$, uniform in $[r/2, 3r/2]$, or gamma $\Gamma[r/12, 12]$, with r in $\{0.04, 0.06\}$. Sample sizes of patterns were always 50 for estimation purposes. For comparison of the estimators we used the mean square error, given by $\sum(\hat{\gamma} - \gamma)^2/49$ and the mean bias.

4.1 Minimum-contrast method

This method is being widely used, especially in goodness-of-fit tests for point field models. Suppose we have access to some characterizing function U_ϑ from the point pattern depending on the parameter ϑ from the model considered. One example of how this method can be applied

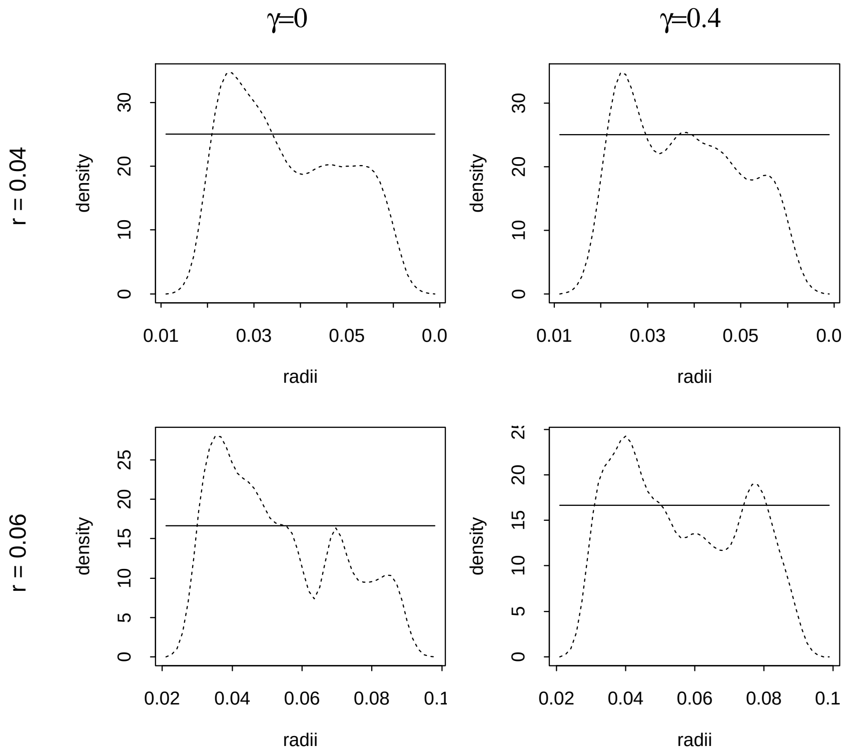

Figure 4: Empirical mark density function (dashed line) for a sample of 50 Strauss disc patterns with "primary" mark distribution Uniform[r/2,3r/2] (solid line being the theoretical density).

consists in finding ϑ that minimizes a Von-Mises-type statistic

$$\int [\hat{U}(s) - U_\vartheta(s)]^2 \; f(s) \; ds,$$

where $f(s)$ is a suitable weight function (Stephens, 1986).

Although this is a very simple method in principle, computing work can be intensive if we cannot calculate U_ϑ analytically. We focused on the Strauss disc model and considered U_ϑ to be the K-function that depends on the interaction parameter γ, which is our primary interest. We fixed $f(s) = 1$. Thus we have that our minimum-contrast estimator

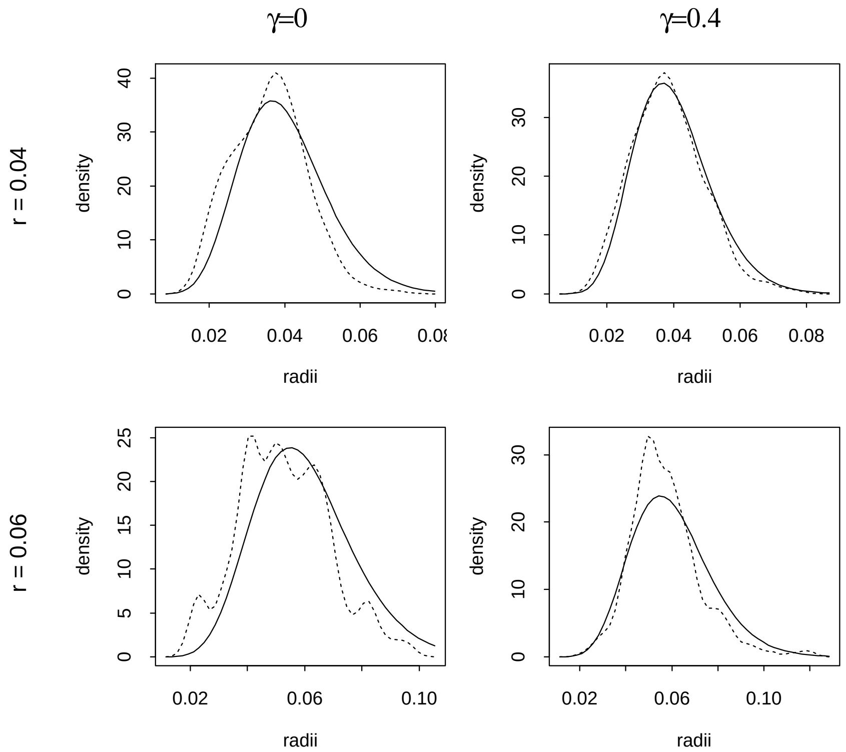

Figure 5: Empirical mark density function (dashed line) for a sample of 50 Strauss disc patterns with "primary"mark distribution Γ(r/12,12) (solid line being the theoretical density).

of γ is given by γ which minimizes the integral

$$\int_0^{t_0} [\hat{K}(s) - K_\gamma(s)]^2 \, ds,$$

for some appropriate t_0.

Other functions can be used instead of the K-function, such as the L-function, the pair correlation function or the nearest-neighbour distance distribution function (Stoyan *et al*, 1987). We considered the K-function for purposes of comparison among methods and used estimator proposed by Ripley (1979a) for estimating this function.

As we do not have an explicit analytical form for the K-function for Strauss disc processes, we approached this problem by simulating

100 patterns and calculating the MCMC average of the estimated K-function.

Figures 6 and 7 show, respectively, the mean square error and the mean bias of 50 estimates of γ for all considered methods, for different mark distributions. In Figure 6 we noticed that, for the minumum-contrast estimator, the mean square error is slightly smaller overall for the mean radius $r = 0.06$. In Figure 7, we see that the mean bias is significantly higher in absolute value at extreme values of γ for the same estimator, what makes this method not attractive for these particular cases of Strauss disc processes.

4.2 Pseudo-likelihood method

Goulard *et al.* (1996) gave a general definition and derived the pseudo-likelihood function for marked Gibbs point processes. They developed formulae for two particular cases, the bivariate Gibbs point processes and the Strauss disc processes. Further, they applied the maximum pseudo-likelihood for one data set for each of the cases. For the bivariate Gibbs point process, they presented simulation results. We present here their formulae for Strauss disc processes and describe an algorithm for the pseudo-likelihood estimator.

We are going to use the same notation as in Section 2, but definitions here require a reparametrization, in order to better understand the basis for the methodology. Consider a marked point process $\Psi = \{([x_i, r_i]\}$, with finite number of points on a bounded space Ω and marks on $\Delta \subset [0, \infty)$ with density function μ. Let $x = (x_1, ..., x_n)$, $r = (r_1, ..., r_n)$ and consider $[x_i, r_i]$ to be a disc with radius r_i centered at x_i. A general pairwise interaction process has density

$$f_{\gamma,\beta}(x, r) = z\ \beta^n \prod_{i<j} \phi(\| x_i - x_j \|, r_i, r_j; \gamma), \tag{2}$$

where $\phi : \Omega \times \Omega \times \Delta \times \Delta \to [0, \infty)$.

With appropriate reparametrization, we can write 2 in the form

$$f_{\theta,\alpha}(x, r) = z\ exp\{-n\alpha - \sum_{i<j} \varphi(\| x_i - x_j \|, r_i, r_j; \theta)\}, \tag{3}$$

with $\theta = -\log\gamma$ and $\alpha = -\log\beta$, where $\varphi : \Omega \times \Omega \times \Delta \times \Delta \to (-\infty, \infty)$ is called the mark pair potential function, and $\alpha : \Omega \to (-\infty, \infty)$ is the mark chemical activity function. The function φ characterizes interactions between marked points and α describes the ability of the system to receive a point.

Goulard *et al.* (1996) considered the mark chemical activity to depend on the marks instead of being a constant, as a way of dealing with the problem that we do not have access to the "primary" distribution function of the marks (the mark distribution if we had no interaction). They defined the pseudo-likelihood (PL) function of first order by

$$\begin{aligned} logPL^{(1)}(\nu,\theta;\Psi) &= -\sum_{[x_i,r_i]\in\Psi} E_{\nu,\theta}(x_i,r_i,\Psi-\delta[x_i,r_i]) \\ &\quad - \int_\Omega\int_\Delta exp\left(-E_{\nu,\theta}(\xi,r_\xi,\Psi)\right)\ \mu(dr_\xi)\ d\xi, \quad (4) \end{aligned}$$

with

$$E_{\nu,\theta}(x_i,r_i,\Psi)) = \alpha(r_i;\nu) + \sum_{[y_j,r_j]\in\Psi} \varphi(x_i,y_j,r_i,r_j;\theta); \quad (5)$$

being the local energy at $[x_i,r_i]$ with respect to Ψ, $\alpha(\,\cdot\,;\nu)$ and $\varphi(\,\cdot\,;\theta)$, the parametric models for the chemical activity and pair potential, respectively, and θ and $\nu=\{\nu_1,\cdots,\nu_k\}$, with k being the number of subsets constituting the partition of Δ (see below), are the parameters to be estimated.

The Strauss disc process has pair potential function given by

$$\varphi(\parallel x_i-x_j\parallel,r_i,r_j;\theta) = \theta\cdot\mathbf{1}(\parallel x_i-x_j\parallel\le r_i+r_j),\quad \theta>0. \quad (6)$$

Goulard *et al.* (1996) fixed the mark space to be $\Delta=[D_0,D]$, where D_0 is the minimum and D the maximum of the disc radii from the observed values, and considered the chemical activity as a step function, i.e. for the partition $D_0<\cdots<D_{l-1}<D_l<\cdots<D_k=D$ of Δ, $\alpha(r_i;\nu)=\nu_l$ on $(D_{l-1},D_l]$. Thus the pseudo-likelihood function (4) takes the form

$$\begin{aligned} &logPL(\nu,\theta;\Psi) = \\ &\quad -\sum_{[x_i,r_i]\in\Psi}\sum_{l=1}^{k} \nu_l\ \mathbf{1}_{(D_{l-1},D_l]}(r_i) \\ &\quad -\ \theta\sum_{[x_i,r_i]\in\Psi}\ \sum_{[x_j,r_j]\in\Psi,i\neq j} \mathbf{1}(\parallel x_i-x_j\parallel\le r_i+r_j) \quad (7) \\ &\quad -\frac{1}{|D-D_0|}\int_\Omega\int_{D_0}^{D}\exp\left(-\sum_{l=1}^{k}\nu_l\ \mathbf{1}_{(D_{l-1},D_l]}(r_\xi)\right. \\ &\quad \left.-\ \theta\cdot\sum_{[x_j,r_j]\in\Psi}\mathbf{1}(\parallel\xi-x_j\parallel\le r_\xi+r_j)\right)\ dr_\xi\ d\xi. \end{aligned}$$

Thus the pseudo-likelihood estimators for the parameters are given by the values of $\nu_1, \cdots, \nu_k$ and θ that maximize (7).

We describe now the algorithm that was used to calculate the estimator of the parameter of interest θ. Let's take

$$S(\xi, r_\xi) = \sum_{[x_j, r_j] \in \Psi} \mathbf{1}(\| \xi - x_j \| \leq r_\xi + r_j),$$

$$SS = \sum_{[x_i, r_i] \in \Psi} \sum_{[x_j, r_j] \in \Psi, j \neq i} \mathbf{1}(\| x_i - x_j \| \leq r_i + r_j)$$

and

$$r^l = \sum_{[x_i, r_i] \in \Psi} \mathbf{1}_{(D_{l-1}, D_l)}(r_i).$$

Then

$$\frac{\partial logPL(\nu, \theta; \psi)}{\partial \nu_l} = -r^l + \frac{e^{-\nu_l}}{|D - D_0|} \int_\Delta \int_{D_{l-1}}^{D_l} e^{-\theta S(\xi, r_\xi)} \, dr_\xi \, d\xi, \quad (8)$$

and setting $\frac{\partial \log PL(\nu, \theta; \psi)}{\partial \nu_l} = 0$ we have that

$$\frac{e^{-\hat{\nu}_l}}{|D - D_0|} = \frac{r^l}{\int_\Delta \int_{D_{l-1}}^{D_l} e^{-\theta S(\xi, r_\xi)} \, dr_\xi \, d\xi}. \quad (9)$$

On the other hand, if we take the derivative of the pseudo-likelihood with respect to θ and substitute (9), we get

$$\frac{\partial logPL(\nu, \theta; \psi)}{\partial \theta} = -SS + \sum_{l=1}^{k} r^l \, J_l(\theta), \quad (10)$$

with

$$J_l(\theta) = \frac{\int_\Delta \int_{D_{l-1}}^{D_l} e^{-\theta S(\xi, r_\xi)} S(\xi, r_\xi) \, dr_\xi \, d\xi}{\int_\Delta \int_{D_{l-1}}^{D_l} e^{-\theta S(\xi, r_\xi)} \, dr_\xi \, d\xi}.$$

The pseudo-likelihood estimator of θ is the solution of the equation $\frac{\partial logPL(\nu, \theta; \psi)}{\partial \theta} = 0$.

In terms of computing the estimator, we need an algorithm to find the value of $J_l(\theta)$. An interpretation of $J_l(\theta)$ can be made as follows: Note that the function $S(\xi, r_\xi)$ of the process evaluates the number of discs in the process that overlap a disc centered at ξ with radius r_ξ. Thus $S(\xi, \cdot)$ is a counting process and each r_ξ can be considered as the "time" for the i-th interaction in a particular process observed in a window. Therefore we can let r_ξ vary over the range of integration and sum up the areas of the rectangles formed by the step functions depending on $S(\xi, r_\xi)$ to calculate the inner integrals in $J_l(\theta)$. We can approximate the outer integrals with respect to ξ by selecting random points in the window of interest and averaging over the inner integral using Monte Carlo.

In the simulation, we had a sample of 50 Strauss disc processes, for which we calculate the pseudo-likelihood estimators. We considered a grid of 500 points in the window that is the unit square in $\Re^2$ to approximate the integrals.

Figures 6 and 7 show that this estimator has a stable performance for the cases we considered. We see that the estimator improves its performance slightly when the mean radius is bigger, which it is to be expected. In Figure 7 we see that a bias is introduced for the smaller mean radii for big values of γ, most probably because these cases get very close to the Boolean pattern. But overall we can say this method has nice properties.

4.3 Takacs-Fiksel method

Using the same reparametrization used in the last section, we say that the distribution of a homogeneous and isotropic marked Gibbs field satisfies some continuity properties and the following mean-value relation :

$$\lambda \int E_{0,r}(T(\Psi, r))\ M(dr) = \tag{11}$$
$$\int E\left\{T(\Psi, r)\ exp\left[-\alpha(r) - \sum_{[x_i, r_i]\in\psi} \varphi(\| x_i \|, r_i, r, \theta)\right]\right\}\ M_1(dr),$$

where M is the distribution function of the marks, M_1 is the "primary" distribution funtion, and $E_{0,r}$ is the Palm mean-value operator (gives means under the condition that there is a point with the mark r at 0) (Stoyan and Stoyan, 1994).

Equation 11 is a generalization of the mean-value relation for unmarked Gibbs point processes. The idea of the Takacs-Fiksel method is to choose a series of test functions $T_v(\Psi, r)$, $v = 1, \cdots, V$, where V is at least equal to the dimension of (α, θ), compute estimates $\hat{L}_v(\alpha, \theta)$ and $\hat{R}_v(\alpha, \theta)$ of the left and right sides of (11) for each $T_v(\Psi, r)$, and estimate (α, θ) to minimize the sum of squares

$$S(\alpha, \theta) = \sum_{v=1}^{V} \left\{ \hat{L}_v(\alpha, \theta) - \hat{R}_v(\alpha, \theta) \right\}^2 . \tag{12}$$

Since Takacs (1983) proposed this estimation method, various test functions T_v have been considered, especially for the unmarked case. According to Stoyan and Stoyan (1994), experience shows that with point processes that present regularity (or inhibitory processes),

$$T_v(\Psi) = N(t_v) = \text{ number of points } x_i \text{ in } \Psi \text{ with } \parallel x_i \parallel \leq t_v$$

is preferable. The advantage of using this test function is that we get $\lambda^2 K(t_v)$ on the left side of (11). For the case of marked point processes, of course we would be using a marginal test function, since it does not depend on the marks explicitly. But we should consider the fact that the K-function indirectly depends on the marks, since the distance between points depends on the marks. Besides, the right side takes into account the "primary" mark distribution. If we knew M_1 or could estimate it efficiently, maybe we would be able to compensate for the estimation of θ.

In the previous section, we considered the pseudo-likelihood estimators for (α, θ). If we consider a little different characterization of the mean-value relation given by

$$\int_{\Omega \times \Delta} \sum_{[x_i, r_i] \in \Psi} h(x_i, r_i, \Psi - \delta([x_i, r_i]))\, P(d\psi) =$$
$$\int_{\Delta} \int_{\Omega} \int_{\Omega \times \Delta} h(x_i, r_i, \Psi)\, exp(-E(x_i, r_i, \psi))\, P(d\psi)\, dx_i M(dr_i),$$

where $\Psi - \delta([x_i, r_i])$ can be interpreted as the point process without the point x_i, then the pseudo-likelihood estimators could be considered as a particular case of the Takacs-Fiksel estimators. The test functions in that case are $h_1(x_i, r_i, \Psi) = \dfrac{\partial \alpha(r, \nu)}{\partial \nu}$ and $h_2(x_i, r_i, \Psi) = \dfrac{\partial \varphi(x_i, x_j, r_i, r_j; \theta)}{\partial \theta}$, with functions α and φ as in (5). Based on this fact, Goulard *et al.* (1996) expressed their concern with respect to the advantage the pseudo-likelihood estimator considered by them would have over other Takacs-Fiksel estimators.

In order to make a comparison with the pseudo-likelihood method and the minimum-contrast based on the K-function presented in the first section, we considered the Takacs-Fiksel estimator based on the K-function, that is, considering the test functions of the kind $N(t_v)$. The left side of (11) is defined by the K-function with which we can use Ripley's estimator (Ripley, 1979a). The right hand side can be estimated by

$$\hat{R}_v(\alpha,\theta) = \frac{1}{n_u\, n_y} \sum_{l=1}^{n_u} \sum_{j=1}^{n_y} N_j(t_v) \exp\left\{ -\alpha - \sum_{[x_i,r_i]\in\Psi} \varphi(\| x_i - y_j \|, r_i, u_l) \right\},$$

where the y_j form a lattice of n_y points in W, u_l form a sample of n_u marks from the distribution chosen to estimate M_1, and $N_j(t_v)$ denotes the number of events $[x_i, r_i]$ with $\| x_i - y_i \| \leq r_i + u_l$.

To implement (12), we used a regularly spaced grid of $n_y = 81$ fixed points in the unit square, $V = 5$ and $t_v = 0.05v$. We chose these values trying to get a reasonable but not too expensive estimator, computationally speaking. Of course the accuracy of the estimation can always be improved.

We considered for the estimation the same cases we considered in the pseudo-likelihood estimation in terms of mean radius, distributions of the marks and β. As the distribution from which we sample the u_l's (weight distribution), we considered the uniform distribution (as in the pseudo-likelihood method) and the distribution we used to generate samples of the Strauss disc processes (μ_1).

Figure 6 shows the mean square error. There is little difference between the two weight distributions results (curves TFM and TFU) when we have a gamma distribution as "primary" distribution. This means that the estimator is not sensitive to the choice of weight distribution for the cases we considered. Figure 7 shows that a consistent bias is present, although small. The bias is slightly bigger for the smaller mean radius in this case too.

4.4 Comparison of methods

As far as the cases we considered here, the pseudo-likelihood estimator proposed by Goulard *et al.* (1996) presented the best performance in terms of estimation of the parameter γ. For comparison purposes we have that the standard deviation for the mean square error (MSE) can be approximated by $MSE\,\sqrt{\frac{2}{k}}$, with k being the degrees of freedom (sample size minus one) and we can also calculate the standard deviation of the mean bias for the sample. We found that the standard deviation for the mean square error is approximately 0.038 for the Takacs-Fiksel methods and maximum pseudo-likelihood method and 0.054 for the minimum-contrast methods. The standard deviation for the mean bias is approximately 0.004 for the Takacs-Fiksel methods and maximum pseudo-likelihood method and 0.005 for the minimum-contrast methods.

The minimum-contrast estimator presents a significant bias, compared to the others (Figure 7). Further, the mean square errors are higher than for the other two methods, especially for strong interactions (Figure 6). But an improvement of performance with respect to both bias and mean square error can be seen as we increase the radius mean. If we look at both Figures 6 and 7, we see that the curves approach the ones for the pseudo-likelihood estimator.

The Takacs-Fiksel estimators based on the K-function present closer results to the pseudo-likelihood than to the minimum-contrast method in terms of mean square error (Figure 6), but there is the problem of bias appearing for all considered cases (Figure 7). The choice of the weight distribution seemed not to matter when the radii distribution was gamma. We can notice almost no difference between the curves TFM and TFU in both Figure 6 and Figure 7. Most probably what happens with respect to the bias is that although we introduced in the equation some correction for the marks on the right side of the mean-value relation equation, we are losing some information related to the intensity of the marks. This is taken care of in the pseudo-likelihood estimation when they consider the chemical activity as nonconstant.

Thus, although Goulard *et al.* (1996) implied that the choice of the weight distribution may influence the pseudo-likelihood estimation performance, which means that the estimator can still be improved, the best results were found using the uniform distribution as weight over the cases and methods of estimation we considered. In this case, the mean bias is consistently around zero and smaller than any other methods, considering their standard deviation (Figure 7) and the mean square errors are smaller than for the minimum-contrast (Figure 6). However, it would merit further investigations in terms of other weight distributions.

An important feature of this study is that we may be able to get nice comparison tests of Strauss disc patterns based on this pseudo-likelihood estimator for the interaction parameter.

5 Conclusions

We focused our studies on the interaction parameter of the Strauss disc process γ, which in general is of the greatest interest in applications. A question that arises naturally is how we could estimate this parameter in the case of Strauss disc processes. The issue was considered in earlier work by Goulard *et al.* (1996) using the pseudo-likelihood method, but studies about properties has not been reported. We compared three methods, pseudo-likelihood, minimum-contrast based on the K-function and the Takacs-Fiksel method also based on the K-function through MCMC simulation. The pseudo-likelihood estimator was considering the uniform distribution as the weight distribution, while for the Tacaks-Fiksel we considered both uniform and gamma distribution. Our simulation study indicated that the pseudo-likelihood estimator for the interaction parameter has the best performance, compared to the minimum-contrast and the Takacs-Fiksel method. The pseudo-likelihood estimator was shown to be robust in terms of varying mark distribution shape and size of the discs (mean radius). Further investigations should be done for the pseudo-likelihood method using weight distributions other than the uniform. New methods like maximum likelihood and even nonparametric methods could be also investigated.

Another issue that was not covered here which deserves some investigation is about the intensity of the processes, which depends on the interaction parameter, distribution of the marks, and on the intensity β of the process when γ=1. This is related to the chemical activity, and which was considered by Goulard *et al.* (1996) as a corrector factor for the estimation, most probably the reason for the estimator to present better properties than the others.

References

Baddeley, A.J. and Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, **57**, 89-121.

Cressie, N.A.C. (1991). *Statistics for Spatial Data.* New York : John Wiley.

Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes. *International Statistical Review*, **62**, 99-117.

Fiksel, T. (1984) Estimation of parametrized pair potentials of marked and non-marked gibbsian point processes. *Elektronische Informationsverarbeitung und Kybernetik*, **20**, 270-278.

Geyer, C. J. and Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, **21**, 359-373.

Goulard M., Särkka, A. and Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics*, **23**,365-379.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and its applications. *Biometrika*, **57**, 97-109.

Jensen, J.L. and Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability*, **3**, 445-461.

Ogata, Y. and Tanemura, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics - Part B*, **33**, 3115-338.

Ogata, Y. and Tanemura, M. (1985). Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method. *Biometrics*, **41**, 421-433.

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

Percus, J.K. (1964). The pair distribution function in classical statistical mechanics. In: Frish, H.L. and Lebowitz, J.L., eds., *The Equilibrium Theory of Classical Fluids.* Benjamin, New York.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T.

(1993). *Numerical Recipes in C : The Art of Scientific Computing.* 2nd edition. Cambridge : Cambridge University Press.

Ripley, B.D. (1979a). Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society - Series B*, **41**, 368-374.

Ripley, B.D. (1979b). Algorithm AS137. Simulating spatial patterns : dependent samples from a multivariate density. *Applied Statistics*, **28**, 109-112.

Stephens, M.A. (1986).Tests for the uniform distribution. In *Goodness-of-fit Techniques* (eds R.B. D'Agostino and M.A. Stephens), pp 331-366. New York : Dekker.

Stoyan, D., Kendall, W.S. and Mecke, J. (1987). *Stochastic Geometry and Its Applications.* Chichester : John Wiley & Sons.

Stoyan, D., Stoyan, H. (1994). *Fractals,Random Shapes and Point Fields.* Chichester : John Wiley & Sons.

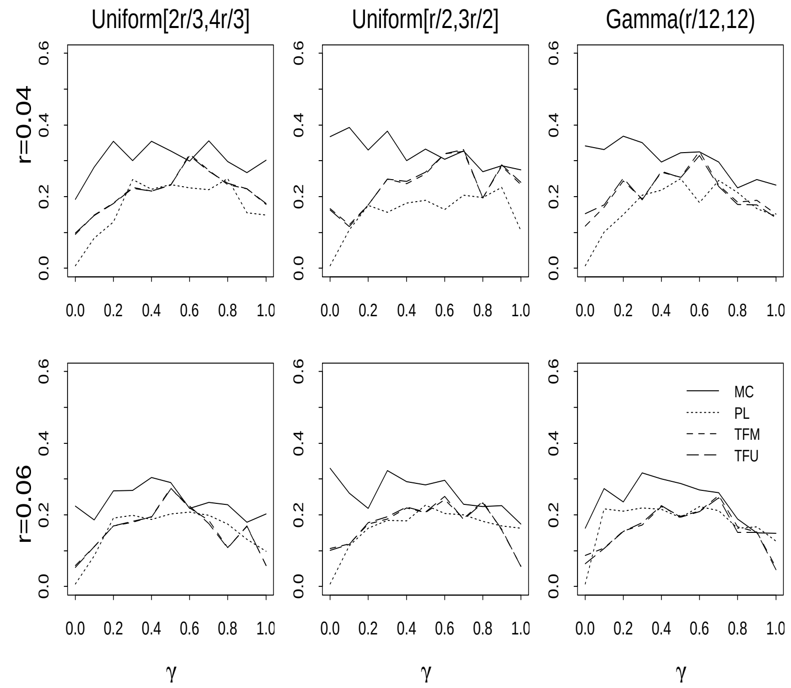

Figure 6: Mean square error versus γ for Strauss disc patterns with mean radius equal to r, for estimates of γ, obtained using the methods minimum-contrast (MC) for the K-function, pseudo-likelihood (PL), Takacs-Fiksel using the mark distribution as weight (TFM) and the uniform distribution (TFU).

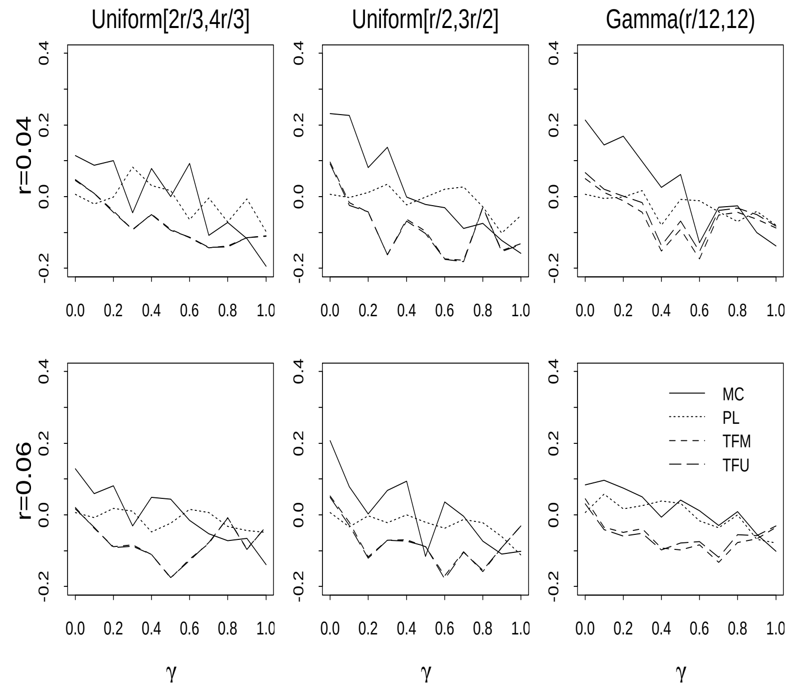

Figure 7: Mean bias versus γ for Strauss disc patterns with mean radius equal to r, for estimates of γ, obtained using the methods minimum-contrast (MC) for the K-function, pseudo-likelihood (PL), Takacs-Fiksel using the mark distribution as weight (TFM) and the uniform distribution (TFU).