

On kernel estimators of second-order measures for spatial point processes *

Jorge Mateu†, Miguel Montenegro‡

†Universitat Jaume I. Department of Mathematics.

Campus Riu Sec, E-12071 Castellón, Spain. ‡Universidad Tecnológica Metropolitana.
Department of Mathematics. Santiago de Chile, Chile.

Abstract

Second-order measures such as the well known and widely used *product density* or *pair correlation function* are often used to describe spatial variability and correlations in spatial point processes. Traditionally, the estimation procedure of these functions is based on kernel smoothing and, in the spatial statistics literature, the Epanechnikov kernel has often been used. However, this preference is not proved through analytical nor empirical grounds. We thus define a new non-negative kernel family, analyze its properties and present a simulation study to prove its improvement compared to other kernels used in spatial smoothing.

Keywords: Kernel smoothing, Matérn and hard-core processes, Pair-correlation, Product density, Spatial point processes.

1 Introduction

Spatial point processes can be considered random geometric structures where the typical measure is based on the spatial locations of the individuals considered. When describing variability and correlations in individuals stands, we have to consider pairs of individuals, and the corresponding characteristics are called *second-order measures*, such as the *product density* (PD). This function is of interest as can be used to discriminate amongst several spatial point structures. Moreover, it plays a central role in point processes theory as it is closely related to the well known cumulative second-order characteristics, the *K*-function (KF), and the *pair correlation function* (PCF).

Corresponding author: J. Mateu. Department of Mathematics, Campus Riu Sec, Universitat Jaume I, E-12071, Castellón, Spain. Tel.: +34.964.728391. Fax: +34.964.728429. Email: mateu@mat.uji.es. Work partially funded by project BFM2001-3286.

A second-order measure, such as PD or PCF, has to be empirically estimated from a data set and this can be done in a similar way to the density estimation for continuous variables.

One way to obtain the PCF can be done through estimation of the K -function and later on its derivative using smoothing splines. This is the way chosen, for example, by Baddeley (2003) [1]. This method chooses among all functions $r(x)$ with two continuous derivatives the one that minimizes the penalized residual sum of squares

$$\sum_{i=1}^n \{y_i - r(x_i)\}^2 + \lambda \int_a^b \{r''(t)\}^2 dt \quad (1)$$

where λ is a fixed constant, and $a \leq x_1 \leq \dots \leq x_n \leq b$. The first term measures closeness to the data while the second term penalizes curvature in the function. Reinsch (1967) [9] showed that (1) has an explicit, unique minimizer which is a natural cubic spline with knots at the unique values of x_i . Large values of λ produce smoother curves while smaller values produce more wiggly curves. At the one extreme, as $\lambda \rightarrow \infty$, the penalty terms dominates, forcing $f''(x) = 0$ everywhere, and thus the solution is the least square line. At the other extreme, as $\lambda \rightarrow 0$, the penalty term becomes unimportant and the solution tends to an interpolating twice-differentiable function.

Another way to estimate the pair correlation function is using a direct method by means of kernel smoothing with a particular bandwidth value. This is the most extended estimation method in the spatial point process context (Cressie, 1993 [2]; Stoyan & Stoyan, 1994 [12]) and the one we focus on here. A kernel smoother uses an explicitly defined set of local weights, defined by the kernel, to produce the estimate at each target values. Usually a kernel smoother uses weights that decrease in a smooth fashion as one moves away from the target points. The weight given to the j th point in producing the estimate at i th point is defined by

$$s_{ij} = \frac{c_i}{\varepsilon} K \left(\left| \frac{s_i - s_j}{\varepsilon} \right| \right)$$

where $K(t)$ is an even function decreasing in $|t|$. The parameter ε is the bandwidth, and the constant c_i is usually chosen so that weights sum to one, although there are slight variations on this.

In spatial statistics literature, the *Epanechnikov kernel* (Epanechnikov, 1969 [4]) has often been used, but no analytical nor empirical grounds have been published for this preference. The extended use of the Epanechnikov kernel in the spatial point pattern literature is due to the fact that it is the *optimum second order kernel*, asymptotically minimizing the mean integrated square error (MISE). However,

the use of this kernel under several practical circumstances, for example when the number of spatial points is not too high, may be inappropriate.

In this paper, we propose the use of alternatives to the Epanechnikov kernel to reduce the importance of the choice of the correct bandwidth. In particular, we present a *new family of non-negative kernels* that has the property that include as a particular case the Epanechnikov kernel, and furthermore we propose an increase in the polynomial degree of the non-negative kernel family, as the data available become lower. This family of non-negative kernels is also related to the optimum kernel family.

In particular, we present and define the theoretical properties of the family of non-negative kernels, and show a simulation study comparing the behaviour of the new family with respect to the usual Epanechnikov kernel and other known optimum kernels (Gasser *et al.*, 1985 [6]). Also, information on adequate bandwidth parameters will be given. The simulation study is carried out on the three types of basic spatial structures randomness, aggregation and inhibition, given by three spatial structures: complete spatial randomness (CSR), Thomas field cluster model (Stoyan *et al.*, 1995 [13]) and Matérn hard-core model (Stoyan *et al.*, 1995 [13]). For each combination of model, number of points, type of kernel and bandwidth, the product density is estimated and the corresponding *MSE* calculated.

The plan of the paper is the following. The family of optimum and non-negative kernels is presented in Section 2. Section 3 shows a general theory for estimation of densities. Then, Section 4 compares the optimum versus the non-negative kernels. This paper focuses on the analysis of spatial point processes, and particularly on kernel smoothing-based estimation for second-order measures. In this line, Section 5 presents some basic theoretical background of point processes. A simulation study is shown in Section 6 presenting several recommendations. The paper ends with some conclusions.

2 Optimum and Non-negative kernels

One possibility for choosing the kernel is by using the so-called optimum kernels, of which the Epanechnikov kernel corresponds to $k = 2$ (Gasser *et al.*, 1985 [6]).

It is known that under common regularity conditions, including $k - th$ order differentiability of the curve to be estimated for a given $k > 0$, and insertion of the asymptotically term of mean or integrated mean square error of kernels estimator for a variety of "rate functions" (regression, density or intensity function), is proportional to the functional $T(K)^{2/(2k+1)}$, where

$$T(K) = \left(\int K^2(x) dx \right)^k \left| \int K(x) x^k dx \right| \quad (2)$$

This functional incorporates the dependence of the (integrate) mean squared error on the kernel function K . The kernel function K in (2) is assumed to be of order k , i.e., satisfies the moment conditions $K \in \mathcal{K}_k$, where

$$\mathcal{K}_k = \left\{ f \in L^2 : \int f(x) x^j dx = 0, j = 0, \dots, k-1, \int f(x) x^k dx \neq 0 \right\} \quad (3)$$

As usual, L^2 is the space of square integrable functions on \mathbb{R} .

The problem of minimizing the above functional $T(K)$ over the set of functions \mathcal{K}_k , has not solutions since the second factor in (2) can be made arbitrarily close to zero. Therefore, it was suggested in Gasser *et al.*, 1985 [6] to impose on K the following additional side condition, called *minimality of sign changes*, which secures the closedness of the set of considered kernel:

$$K \in \mathcal{N}_{k-2} := \{ f \in L^2 : f \text{ has exactly } k-2 \text{ sign changes on } \mathbb{R} \} \quad (4)$$

Here $k-2$ is equal to the lower bound of the number of signs required in order to satisfy (3), which was established in Mammitzsch (1985) [7] and Müller (1985) [8].

The variational problem to consider thus becomes

$$T(K) = \min, \quad \text{subject to } K \in \mathcal{K}_k \cap \mathcal{N}_{k-2} \quad (5)$$

Solutions of (5) as referred to as *optimal kernel*.

Given the idea that the kernels must be non-negative and that the weight must diminish as it moves away from the central point, in this section we propose a family of non-negative kernels and analyze their properties. Without losing generality, we can suppose that the kernel reaches its maximum at some $x = x_0$, and diminishes as it moves away from the center to the point where it has zero value at $x = -1$ and $x = 1$. In order to carry this out, we use the family

$$K^{(\alpha, \beta)}(x) = M(1-x)^\alpha(1+x)^\beta$$

in such a way that $\alpha + \beta = k$, $K^{(\alpha, \beta)}(x) = O(x^k)$ and M is such that

$$\int_{-1}^1 K^{(\alpha, \beta)}(x) dx = 1$$

By making the change to variable $y = \frac{x+1}{2}$ in the integral

$$\int_{-1}^1 M(1-x)^\alpha(1+x)^\beta dx$$

we obtain

$$\int_{-1}^1 M \cdot 2^{\alpha+\beta+1}(1-y)^\alpha y^\beta dy = M \cdot 2^{\alpha+\beta+1} B(\alpha+1, \beta+1)$$

where $B(\cdot, \cdot)$ stands for the beta function defined as

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt, \quad x > 0, \quad y > 0 \quad (6)$$

Thus, the value of M obtained is

$$M = \frac{1}{2^{\alpha+\beta+1} B(\alpha+1, \beta+1)}$$

The maximum of $K^{(\alpha, \beta)}(x)$ is found at $x = \frac{\beta-\alpha}{\alpha+\beta}$. In the case that $\alpha = \beta$, the family is symmetric and its maximum is found at $x = 0$. In this case $\alpha = \beta = \frac{k}{2}$, and the kernel family gives

$$K_k(x) = \begin{cases} \frac{\Gamma(k+2)}{2^{k+1}\Gamma(\frac{k}{2}+1)^2} (1-x^2)^{\frac{k}{2}} & \text{si } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

For $k = 0$, the uniform kernel is obtained, $k = 2$ results in the Epachnenikov kernel, and in the case $k = 4$ the biweight kernel is obtained.

The non-negative kernel family (7) has the following properties:

$$\text{P1 } \int K_k(x) dx = 1$$

$$\text{P2 } \int K_k(x) x^j dx = 0 \text{ si } x \text{ es impar, } j = 0, \dots, k$$

$$\text{P3 } \int K_k(x) x^k dx = \frac{(1+(-1)^k)\Gamma(k+1)(\frac{k}{2}+\frac{3}{2})}{2^{k+1}\Gamma(\frac{k}{2}+1)\Gamma(k+\frac{3}{2})}$$

$$\text{P4 } \int K_k(x)^2 dx = \frac{\Gamma(k+2)^2\Gamma(k+1)^2}{2\Gamma(\frac{k}{2}+1)^4\Gamma(2k+2)}$$

3 Estimation of densities

One of the main uses of kernels is to estimate one-dimensional probability density functions from a random sample. Let us suppose that X_1, \dots, X_n is a set of random variables with a common density of f . The parametric approximation of the estimation f considers that f comes from a specific parametric family of functions such as normal, Poisson, gamma etc., and therefore it aims to determine the unknown parameter, using for instance, a maximum probability estimator. However, a non-parametric density estimator does not assume a specific form of the functional form of f .

In this context, the Parzen-Rosenblatt estimator via kernel (Rosenblatt, 1971 [10]) is used, which is given by

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (8)$$

K is usually chosen in such a way that it is a symmetric, unimodal probability density function of around zero. This guarantees that $f(x; h)$ is also a density. However, kernels are sometimes used which are not densities.

3.1 *MSE and MISE*

Consider $\hat{f}(x; h)$ an estimator of the density function $f(x)$ at some point $x \in \mathbb{R}$. To compute $MSE(\hat{f}(x; h))$ we will require expressions for the mean and the variance of $\hat{f}(x; h)$. Let X be a random variable having density f . Firstly,

$$E[\hat{f}(x; h)] = E[K_h(x - X)] = \int K_h(x - y)f(y)dy \quad (9)$$

where $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$. Using the convolution notation

$$(f * g)(x) = \int f(x - y)g(y)dy$$

allows us to write the bias of $\hat{f}(x; h)$ as

$$E[\hat{f}(x; h)] - f(x) = (K_h * f)(x) - f(x)$$

Similar calculation leads to

$$V[\hat{f}(x; h)] = \frac{1}{n} \left((K_h^2 * f)(x) - (K_h * f)^2(x) \right) \quad (10)$$

and these expressions may be combined to give

$$MSE [\hat{f}(x; h)] = \frac{1}{n} \left((K_h^2 * f)(x) - (K_h * f)^2(x) \right) + ((K_h * f)(x) - f(x))^2 \quad (11)$$

For the *mean integrated square error (MISE)* we have

$$\begin{aligned} MISE [\hat{f}(\cdot; h)] &= \int E [\hat{f}(x; h) - f(x)]^2 dx \\ &= \int MSE [\hat{f}(x; h)] dx \end{aligned}$$

so we can use (11) to obtain

$$MISE [\hat{f}(\cdot; h)] = \frac{1}{n} \int \left((K_h^2 * f)(x) - (K_h * f)^2(x) \right) dx + \int ((K_h * f)(x) - f(x))^2 dx$$

However, some straightforward manipulation leads to the more handy expression

$$\begin{aligned} MISE [\hat{f}(\cdot; h)] &= \frac{1}{nh} \int K^2(x) dx + \left(1 - \frac{1}{n} \right) \int (K_h * f)^2(x) dx \\ &\quad - 2 \int (K_h * f)(x) f(x) dx + \int f(x)^2 dx \end{aligned} \quad (12)$$

3.2 Asymptotic MSE and MISE approximations

The problem with the MSE and MISE expressions is that they depend on a complicated form of the bandwidth h , which makes it difficult to interpret the bandwidth influence on the performance of the density estimator via kernel. In order to execute a more simplified interpretation we use an asymptotic development, which means that the sample must be large.

In this section, we suppose that:

- A1** The density function f is such that its second derivative f'' is continuous, with integrable square and such that M exists in such a way that f is monotonous in the $(-\infty, -M)$ and (M, ∞) intervals

A2 The bandwidth $h = h_n$ is a non-random succession of positive numbers. For ease of notation, we will leave out the dependence of h on n .

We assume that h satisfies

$$\lim_{n \rightarrow \infty} h = 0 \text{ y } \lim_{n \rightarrow \infty} nh = \infty$$

which is equivalent to saying that h approaches zero, but at a rate lower than n^{-1}

A3 The kernel K is a bounded probability density function having finite fourth moment and symmetry about the origin.

We consider the estimation of $f(x)$ at $x \in \mathbb{R}$. Making the change of variable $z = \frac{x-y}{h}$ at (9)

$$E \left[\hat{f}(x; h) \right] = \int K(z) f(x - hz) dz$$

Expanding $f(x - hz)$ in a Taylor series about x we obtain

$$f(x - hz) = f(x) - hz f'(x) + \frac{1}{2} h^2 z^2 f''(x) + o(h^2)$$

uniformly in z . This leads to

$$E \left[\hat{f}(x; h) \right] = f(x) + \frac{1}{2} h^2 f''(x) \int z^2 K(z) dz + o(h^2)$$

where we have used

$$\int K(z) dz = 1; \quad \int z K(z) dz = 0; \quad \text{and} \quad \int z^2 K(z) dz < \infty$$

each of which follows from assumption (A3), which finally gives the bias expression

$$E \left[\hat{f}(x; h) \right] - f(x) = \frac{1}{2} h^2 \left(\int z^2 K(z) dz \right) f''(x) + o(h^2) \quad (13)$$

Given that the bias is of the order h^2 , $\hat{f}(x; h)$ is asymptotically unbiased. It can also be noted that the bias depends on the second derivative of f in such a way that its points are greater where the density function curvature is very large. At the points near a peak or a valley $\hat{f}(x; h)$ tends to be smoother than f , in such a way that at the peaks $\hat{f}(x; h)$ is below the f values and in the valleys, $\hat{f}(x; h)$ is above f .

For the variance we have

$$\begin{aligned}
V \left[\hat{f}(x; h) \right] &= \frac{1}{nh} \int K(z)^2 f(x - hz) dz - \frac{1}{n} \left(E \left[\hat{f}(x; h) \right] \right)^2 \\
&= \frac{1}{nh} \int K(z)^2 (f(x) + o(1)) dz - \frac{1}{n} (f(x) + o(1))^2 \\
&= \frac{1}{nh} \int K(z)^2 dz f(x) + o \left(\frac{1}{nh} \right)
\end{aligned} \tag{14}$$

Since the variance is of order $(nh)^{-1}$, $V \left[\hat{f}(x; h) \right]$ converges to zero.

By using the previously calculated bias and variance, the expression for the *MSE* is

$$\begin{aligned}
MSE \left[\hat{f}(x; h) \right] &= \frac{1}{nh} \left(\int K(z)^2 dz \right) f(x) + \frac{1}{4} h^4 \left(\int z^2 K(z) dz \right)^2 f''(x)^2 \\
&\quad + o \left(\frac{1}{nh} + h^4 \right)
\end{aligned} \tag{15}$$

If this expression is then integrated, under the integrability condition of f , we obtain

$$MISE \left[\hat{f}(\cdot; h) \right] = AMISE \left[\hat{f}(\cdot; h) \right] + \left(\frac{1}{nh} + h^4 \right) \tag{16}$$

where

$$AMISE \left[\hat{f}(\cdot; h) \right] = \frac{1}{nh} \int K(z)^2 dz + \frac{1}{4} h^4 \left(\int z^2 K(z) dz \right)^2 \int f''(x)^2 dx \tag{17}$$

This final quantity is known as *asymptotic mean integrated square error*.

The *AMISE* calculation comes from a simple method for the approximation of *MISE* in a case of a large sample.

A further advantage of the *AMISE* is that the optimum bandwidth concerning this criterion, which is noted down as h_{AMISE} , has the following expression

$$h_{AMISE} = \left[\frac{\int K(z)^2 dz}{\left(\int z^2 K(z) dz \right)^2 \left(\int f''(x)^2 dx \right) n} \right]^{1/5} \tag{18}$$

The previous expression is obtained by deriving (17) with respect to h , and setting the derivative equal to zero. However, it is not possible to directly determine the optimum bandwidth value using (18) as the quantity $\int f''(x)^2 dx$ is unknown. From (18), it can be seen that the optimum bandwidth is inversely proportional to $(\int f''(x)^2 dx)^{1/5}$, which measures the total curvature of f . Therefore, for a density with a low curvature, the bandwidth to be used must be large, and the opposite will occur if the density has a high curvature.

Asymptotic theory analyses kernel behaviour when the size of the sample $n \rightarrow \infty$, the bandwidth $h \rightarrow 0$, slowly enough for $nh \rightarrow \infty$ to be met.

4 Comparison of optimum kernels and non-negative kernels

Asymptotic theory analyses kernel behavior when the size of the sample $n \rightarrow \infty$, the bandwidth $h \rightarrow 0$, slowly enough for $nh \rightarrow \infty$ to be met. However, in many applications the size of the sample is not large. In this paper we analyze small, medium and large size samples.

The *MSE* and *MISE* depend on $\int K(z)^2 dz$ and $\int z^2 K(z) dz$. We then compare these quantities in both cases, for the optimum and non-negative kernels. We need the following definition.

Definition 1 *Gamma function, $\Gamma(x)$, is defined as*

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (19)$$

which has the following properties:

(g1) $\Gamma(n) = (n - 1)!$ where n is a positive number.

(g2) $\Gamma(z + 1) = z\Gamma(z)$

(g3) $\Gamma(z + \frac{1}{2}) = \frac{\sqrt{\pi}\Gamma(2z)}{2^{2z-1}\Gamma(z)}$

Proposition 2 *Let $K_{opt}(0, k)$ and $K_{non}(k)$ be the optimum kernel of order $(0, k)$ and the non-negative kernel of degree k . If k is even, then $B(K_{non}(k)) = |B(K_{opt}(0, k))|$ and $V(K_{non}(k)) \leq V(K_{opt}(k))$*

Proof. Given that k is even, we can write $k = 2p$, and then (Gasser, Müller & Mammitzsch, 1985 [6])

$$B(K_{opt}(0, 2p)) = \frac{(2p+1)![(2p)!]^2}{(p!)^2(4p+1)!}$$

On the other hand, using P3 above, we have

$$B(K_{non}(2p)) = \frac{\Gamma(2p+1)\Gamma(p+\frac{3}{2})}{2^{2p}\Gamma(p+1)\Gamma(2p+\frac{3}{2})}$$

Using the properties of the Gamma function, particularly (g1),

$$B(K_{opt}(2p)) = \frac{\Gamma(2p+2)\Gamma(2p+1)^2}{\Gamma(p+1)^2\Gamma(4p+2)} \quad (20a)$$

and using (g3),

$$\begin{aligned} \Gamma(2p+2) &= \frac{2^{2p+1}}{\sqrt{\pi}}\Gamma(p+1)\Gamma\left(p+\frac{3}{2}\right) \\ \Gamma(4p+2) &= \frac{2^{4p+1}}{\sqrt{\pi}}\Gamma(2p+1)\Gamma\left(2p+\frac{3}{2}\right) \end{aligned}$$

which replacing these results into (20a) the first part of the proposition is proved.

Let us now analyze the kernel variances. The variance of the optimum kernel is (Gasser, *et al.*, 1985 [6])

$$V_o = V(K_{opt}(k)) = \frac{(k+1)k^2\Gamma(k+1)^2}{(2k+1)2^{2k}\Gamma(\frac{k}{2}+1)^4} \quad (21)$$

and using P4,

$$V_n = V(K_{non}(k)) = \frac{\Gamma(k+2)^2\Gamma(k+1)^2}{2\Gamma(\frac{k}{2}+1)^4\Gamma(2k+2)}$$

The quotient V_o/V_n will be then

$$\frac{V_o}{V_n} = \frac{k^2\Gamma(2k+1)}{2^{2k-1}(k+1)\Gamma(k+1)^2}$$

and using again (g2) and (g3) we obtain

$$\frac{V_0}{V_n} = \frac{2k}{\sqrt{\pi}(k+1)} \frac{\Gamma(k + \frac{1}{2})}{\Gamma(k)}$$

The first factor is greater than 1 if $k \geq \frac{\sqrt{\pi}}{2-\sqrt{\pi}} \approx 7.8$, the second factor is always greater than 1, given that the Γ function is an increasing monotone function. Then, from $k = 8$ onwards, $V_0 > V_n$.

Note that for $k = 2, 4, 6$ the quotient $\frac{V_0}{V_n}$ gives 1, 1.75, 2.34 respectively. In other words, the non-negative family of kernels proposed has an equal bias at absolute values, but lower variance than optimum kernels, to which the condition $k - 2$ sign changes were aggregated.

■

5 Spatial point processes

5.1 Setup

A *point process* is a stochastic model governing the locations of events $\{s_i\}$ in some set X (a bounded region in \mathbb{R}^d or a torus, but more generally X could be any locally compact Hausdorff space whose topology has a countable base (Diggle, 2003 [3]; Cressie, 1993 [2])). If the locations contain associated measurements or marks, the point process is referred to as a *marked point process*. For example, events may be trees in a forest, towns in a geographic region, or epicenters of earthquakes. The corresponding marks may be, respectively, species types or diameters of trees, population sizes of towns, or magnitudes of earthquakes. It is often assumed that the locations of events $\{s_i\}$ and their corresponding marks $\{Z(s_i)\}$ are the realization of some stochastic process of the form $\{Z(s) : s \in D\}$, where both $Z(\cdot)$ and D are random.

The general setup for an unmarked point process is the following. Let (X, χ, ν) be a measure space such that $\nu(X) < \infty$, and let μ denote the Poisson process on X with intensity measure λ . We require λ to be atomless so that the process be simple. We are interested in point processes that have densities with respect to μ . Let $X^n = X \times X \times \dots \times X$ (n factors) be the set of collections of n (not necessarily distinct) elements of X . Take $X^0 = \emptyset$. If $n \neq k$, then X^n and X^k are disjoint. Let X_e be the disjoint union of all finite Cartesian products X^n , $X_e = \bigcup_{n=0}^{\infty} X^n$.

Every set $B \in X_e$ can be expressed uniquely as the union of disjoint sets; that is, $B = \bigcup_{n=0}^{\infty} B^{(n)}$, where $B^{(n)} = B \cap X^n$.

Let $\chi^{(n)}$ be the smallest σ algebra of sets in X^n containing all product sets $B_1 \times B_2 \times \dots \times B_n$, such that $B_i \in \chi$, $i = 1, \dots, n$. Let χ_e be the class of all sets

$\cup_{n=0}^{\infty} B^{(n)}$ in X_e , such that $B^{(n)} \in \chi^{(n)}$, $n = 0, 1, \dots$. Then, χ_e is the smallest σ algebra of sets in X_e generated by sets $B^{(n)}$. The pair (X_e, χ_e) is called an *exponential space* (Cressie, 1993 [2]).

Let ν^n denote the n -fold product of ν on X^n for $n \geq 1$ and let λ^0 denote the measure on X^0 that gives mass one to the point \emptyset . Then μ defined by

$$\mu(B) = \sum_{n=0}^{\infty} \frac{\nu^n(B \cap X^n)}{n!}, \quad B \in \chi_e \quad (22)$$

is a probability measure on (X_e, χ_e) . An element $\mathbf{s} \in X^n$ is visualized as a pattern of n points in X . The element \emptyset of X^0 is interpreted as the pattern with no points.

5.2 Second-order measures: Product density and pair correlation function

For a unmarked point process N and a Borel set A , the number of points in A , $N(A)$, is a random variable with first moment

$$\mu_N(A) = E(N(A)) = \int_{\Phi} \phi(A) \prod_N(d\phi). \quad (23)$$

The measure μ_N is called the *mean measure* or *first-moment measure*. The *kth-factorial moment measure* of N is defined by

$$\begin{aligned} \alpha_N^{(k)}(A_1 \times \dots \times A_k) &= \int_{\Phi} \sum_{s_1, \dots, s_k \in \phi}^{\text{distinct}} I(s_1 \in A_1) \dots I(s_k \in A_k) \prod_N(d\phi) \\ &= E \left[\sum_{s_1, \dots, s_k \in \phi}^{\text{distinct}} I(s_1 \in A_1) \dots I(s_k \in A_k) \right] \end{aligned}$$

Particularly, for $k = 2$, $\alpha_N^{(2)}(A_1 \times A_2)$ is the mean number of point pairs where one of the points is in A_1 and the other in A_2 .

Frequently, $\alpha_N^{(2)}$ has a density function $l_N^{(2)}$ called the *second-order product density*,

$$\alpha_N^{(2)}(A_1 \times A_2) = \int_{A_1} \int_{A_2} l_N^{(2)}(s, t) ds dt. \quad (24)$$

For infinitesimal ds and dt , $l_N^{(2)}(s, t) ds dt$ can be interpreted as the probability that there is a point of the point process in each of two specified infinitesimal sets with areas ds and dt respectively.

If the point process is homogeneous and isotropic, the product density depends only on the distance r between the points s and t , writing $\ell_N^{(2)}(r)$.

In statistical physics, it is frequently used the *pair-correlation function* g defined by

$$g(r) = \frac{\ell_N^{(2)}(r)}{\lambda^2}. \quad (25)$$

The pair-correlation function characterizes the spatial structure of a point pattern in the following sense. For a completely random point process $g(r) = 1$. Values of the pair correlation function larger than 1 indicate that the interpoint distances around r are relatively more frequent compared to those in a complete random point process, which is typical of a cluster process, and conversely, values of $g(r)$ smaller than 1 indicate that the corresponding distances are rare and this is typical of an inhibition process.

There also exists a close relationship between the pair-correlation function and the well known Ripley's K -function (Stoyan, Kendall & Mecke, 1995 [13]). $\lambda K_N(r)$ represents the expected number of extra events of N within distance r of an event of N .

Under stationarity and isotropy we have that

$$\lambda^2 dK_N(r) = 2\pi r \ell_N^{(2)}(r) dr. \quad (26)$$

Consequently, we obtain

$$g(r) = \frac{1}{2\pi r} \frac{d}{dr} K_N(r) \quad (27)$$

or alternatively, for $r > 0$

$$K_N(r) = \int_0^r g(u) 2\pi u du \quad (28)$$

For a homogeneous Poisson process $K_N(r) = \pi r^2$.

5.3 Statistics for homogeneous point patterns

Let us assume we have n points, $s = \{s_1, s_2, \dots, s_n\}$ observed in a window X . Each point $s_i = (x_i, y_i)$, $i = 1, 2, \dots, n$. Usually λ is estimated by $\hat{\lambda} = \frac{n}{\nu(X)}$ (often, $\nu(X) = |X|$, the Lebesgue measure). To estimate the pair-correlation function, the product density $l_N^{(2)}(r)$ is first estimated, and then $\hat{g}(r) = \frac{\hat{l}_N^{(2)}(r)}{\hat{\lambda}^2}$.

Some form of smoothing is clearly needed to estimate functions like the product density or pair-correlation. Particularly, kernel estimators have often been used (Fiksel, 1988 [5]). Most of the literature on statistical physics and spatial statistics focusses on the Epanechnikov kernel given by (Epanechnikov, 1969 [4])

$$K_\varepsilon(h) = \begin{cases} (1 - h^2/\varepsilon^2) \frac{3}{4\varepsilon}, & -\varepsilon \leq h \leq \varepsilon \\ 0, & \text{otherwise} \end{cases}. \quad (29)$$

The parameter ε determines the smoothness of the estimation. Following these lines, two possible known kernel-based estimators, under homogeneity and isotropy, are (Stoyan & Stoyan (1994) [12]; Stoyan, Kendall & Mecke, 1995 [13])

$$\hat{l}_N^{(2)}(r) = \frac{1}{2\pi r} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{K_\varepsilon(r - \|s_i - s_j\|)}{\nu(X_{s_j} \cap X_{s_i})}. \quad (30)$$

and

$$\hat{l}_N^{(2)}(r) = \frac{1}{2\pi r c_X(r)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_\varepsilon(r - \|s_i - s_j\|) \quad (31)$$

where $c_X(r)$ is the isotropized set covariance function of the window X , which in general, for $h = (r, \varphi)$ is defined as averaging over φ the set covariance function (or geometric covariogram) $\gamma_X(h) = \nu(X \cap X_h)$ (the area of intersection of X and X shifted by h),

$$c_X(h) = \frac{1}{\pi} \int_0^\pi \gamma_X(r, \varphi) d\varphi.$$

So, using for example (31), a possible kernel-based estimator for the pair-correlation function is given by

$$\hat{g}(r) = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n k_\varepsilon(r - \|s_i - s_j\|) / (\hat{\lambda}^2 2\pi r c_X(r)), \quad r > 0 \quad (32)$$

5.4 Examples of spatial point structures

Stochastic geometry is the study of random patterns, whether of points, line segments, or objects. In particular, point processes are random point patterns in the plane or in space. The simplest example of these is the *Poisson point process*, which models "completely random" distribution of points in the plane or space (Figure

(1)). It is a basic building block in stochastic geometry and elsewhere. Its many symmetries typically reduce calculations to computations of area or volume.

In stochastic geometry the classic generalization of the Poisson process is the *Boolean model*; the set-union of (possibly random) geometric figures or grains located one at each point or germ of the underlying Poisson process. Much of the amenability of the Poisson process is inherited by the Boolean model. It is often convenient to think of Boolean models as being derived from marked Poisson processes, in which the points are supplemented by marks encoding the geometric data (such as disk radius) of the grains; in effect the Poisson process now lives on $R^2 \times T$ where T is the mark space. *Poisson cluster processes* form a special case, in which the grains are simply finite clusters of points. In general, a point process is a *Poisson cluster process* if: (a) *Parents* form a homogeneous Poisson process of intensity ρ ; (b) The number of *offsprings* per parent is a random variable M , realised independently for each parent; (c) The position of each offspring relative to its parent is a bivariate random variable Y , realised independently for each offspring; (d) The observed point process consists of the superposition of offsprings from all parents. The famous *Neyman-Scott processes* arise when the clusters are produced by independent sampling. How can we escape the limitations of the independence property of Poisson processes? One possibility is to randomize the intensity of the Poisson process, which produces the *Cox process* (also termed the doubly stochastic Poisson process).

Neyman-Scott processes are *Poisson cluster processes* because they are based on an initial homogeneous Poisson process with intensity λ whose events are called parent points. Around each parent point, a cluster of daughter points is scattered. The number of points per cluster is randomly generated according to a given discrete probability distribution function. The location of the daughter points with respect to their parent center is independently generated for each parent following a given density function. This law is the same for all parents. The final process is formed only by the offsprings.

The *Thomas process* belongs to the Neyman-Scott family, generating a uniform Poisson point process of “parent” points with intensity λ . Then each parent point is replaced by a random cluster of points, the number of points per cluster being Poisson distributed with mean μ , and their positions being isotropic Gaussian displacements from the cluster parent location. Several realizations are shown in Figure (2).

Opposite to cluster structures are the regular or inhibitory ones. A good reference to this class of spatial point pattern structures is the *Matérn's inhibition process* which reads as follows. The process is constructed by first generating a uniform Poisson point process of “proposal” points with some intensity λ inside the window. A proposal point is deleted if it lies within r units distance of another proposal point.

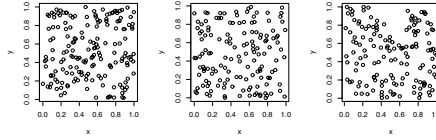


Figure 1: Realizations of a Poisson process

Otherwise it is retained. The retained points constitute Matérn’s inhibition model. Realizations of this model are shown in Figure (3).

In this paper we focus our attention to the Poisson process, the Thomas cluster process and the Matérn’s inhibition process, over which we carry out the simulation study presented in next section. Thomas and Matérn processes have been widely used in literature, and cover the wide scope of spatial point pattern structures.

5.4.1 Second-order properties for the Poisson, Thomas cluster and Matérn’s inhibition process

In the Poisson case, the points are placed independently of one another, and the number of points follows a Poisson distribution with intensity (mean number of points per unit area) λ_P . Thus, the second-order intensity function (Cressie, 1993 [2]) is given by

$$\lambda_2(u, v) = \lambda_2(r) = \lambda_P^2,$$

the covariance density function is

$$\gamma_2(u, v) = \gamma_2(r) = 0,$$

and the second-order characteristics pair-correlation and K -function are given by

$$\begin{aligned} g(r) &= 1 \\ K(r) &= \pi r^2 \end{aligned}$$

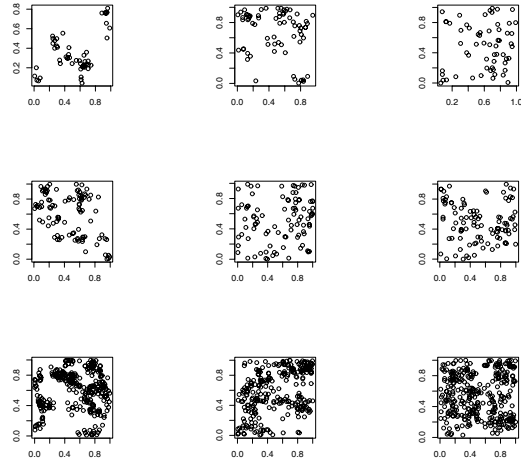


Figure 2: Realizations of Thomas processes: *First row*: Mean number of parents $\lambda = 9$ and mean number of offsprings per parent $\mu = 7$ with a variance of the Gaussian displacements surrounding each parent $R = 0.05$, $R = 0.075$ and $R = 0.10$; *Second row*: Mean number of parents $\lambda = 20$ and mean number of offsprings per parent $\mu = 5$ with a variance of the Gaussian displacements surrounding each parent $R = 0.05$, $R = 0.075$ and $R = 0.10$; *Third row*: Mean number of parents $\lambda = 25$ and mean number of offsprings per parent $\mu = 12$ with a variance of the Gaussian displacements surrounding each parent $R = 0.05$, $R = 0.075$ and $R = 0.10$.

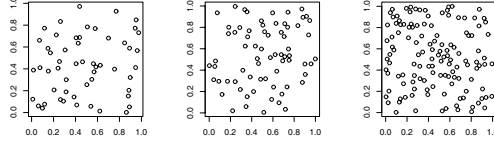


Figure 3: Realizations of Matérn inhibition processes with mean number of points, from left to right, 60, 100 and 300.

In the case of a Thomas cluster process, where the number of offspring per parent follows a Poisson distribution with mean μ , and $h(a)$ is the isotropic bivariate Gaussian distribution function

$$h(a) = (2\pi\sigma^2)^{-1} \exp(-(a_1^2 + a_2^2)/(2\sigma^2)),$$

the intensity of a Thomas cluster process is given by $\lambda_{TCP} = \mu\lambda_P$.

The second-order intensity function is given by

$$\lambda_2(u, v) = \lambda_2(r) = \lambda_{TCP}^2 + \lambda_P \mu^2 (4\pi\sigma^2)^{-1} \exp\{-r^2/(4\sigma^2)\}$$

and the covariance density function is given by

$$\gamma_2(u, v) = \gamma_2(r) = \lambda_P \mu^2 (4\pi\sigma^2)^{-1} \exp\{-r^2/(4\sigma^2)\}$$

Finally, the pair-correlation and K -function take the form

$$\begin{aligned} g(r) &= l_N^{(2)}(r)/\lambda_{TCP}^2 = \lambda_2(r)/\lambda_{TCP}^2 = 1 + \lambda_P^{-1} (4\pi\sigma^2)^{-1} \exp\{-r^2/(4\sigma^2)\} \\ K(r) &= \pi r^2 + \lambda_P^{-1} [1 - \exp\{-r^2/(4\sigma^2)\}] \end{aligned}$$

Finally, when we focus on Matérn inhibition model, where λ_I is the intensity of the inhibition process per unit area. The parameters ν, σ^2 reflect the strength and scale of inhibition.

The pair-correlation function takes the form

$$g(r) = \begin{cases} 0 & r \leq h \\ \frac{(A_1 - B_1) / [(\pi h^2)(C_1)(C_1 - \pi h^2)]}{\left(\frac{1 - \exp(-\lambda \pi h^2)}{\pi h^2}\right)^2} & h < r \leq 2h \\ \frac{(A_2 - B_2) / [(\pi h^2)(C_2)(C_2 - \pi h^2)]}{\left(\frac{1 - \exp(-\lambda \pi h^2)}{\pi h^2}\right)^2} & r > 2h \end{cases}$$

where

$$\begin{aligned} A_1 &= 2C_1 (1 - \exp(-\lambda \pi h^2)) \\ B_1 &= 2\pi h^2 (1 - \exp(-\lambda C_1)) \\ C_1 &= 2\pi h^2 - 2h^2 \arccos\left(\frac{r}{2h}\right) + \frac{r}{2} \sqrt{4h^2 - r^2} \end{aligned}$$

and

$$\begin{aligned} A_2 &= 2C_2 (1 - \exp(-\lambda \pi h^2)) \\ B_2 &= 2\pi h^2 (1 - \exp(-\lambda C_2)) \\ C_2 &= 2\pi h^2 \end{aligned}$$

6 Comparative simulation study

The dependence of the optimum kernel to the number of points and also to the spatial structure lying the an specific spatial pattern is well known. To determine and analyze this dependence, we run a simulation study in which three point pattern models (imitating the widely known three different spatial structures) were simulated in the unit square, with several number of points per pattern: (a) Random (Poisson) patterns were simulated with a mean number of 60, 100 and 300 points; (b) Matérn hard-core patterns were generated with a mean number of 60, 100 and 300 points, and hard-core parameter 0.03; (c) Finally, Thomas cluster patterns were generated with a mean number of 60, 100 and 300 points and the following parameter combinations: (c1) Mean number of parents $\lambda = 9$ and mean number of offsprings per parent $\mu = 7$ with a variance of the Gaussian displacements surrounding each parent $R = 0.05$, $R = 0.075$ and $R = 0.10$; (c2) Mean number of parents $\lambda = 20$ and mean number of offsprings per parent $\mu = 5$ with a variance of the Gaussian displacements surrounding each parent $R = 0.05$, $R = 0.075$ and $R = 0.10$; (c3) Mean number of parents $\lambda = 25$ and mean number of offsprings per parent $\mu = 12$

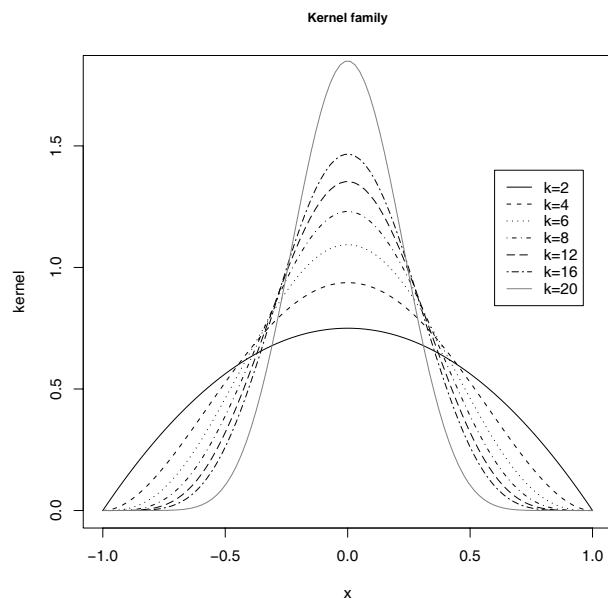


Figure 4: The non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

with a variance of the Gaussian displacements surrounding each parent $R = 0.05$, $R = 0.075$ and $R = 0.10$.

For each combination of model, number of points, and parameter combinations, the PCF function was estimated using several non-negative kernel functions (as presented in section 2, see Figure (4)), and the corresponding MSE versus a sequence of bandwidths was calculated. Note that we know the theoretical expressions for the PCF for these three families of processes (see section 5.4.1). Furthermore, for each combination of model, number of points, type of kernel and bandwidth, a number of 100 simulations (repetitions) were carried out, showing here in the corresponding averages.

Analyzing Figure (4) we see that when the order of the kernel becomes bigger those values closer to the center of the selected bandwidth become more important. In these distributions, there exists places within which the order of the kernels becomes unimportant.

The results for the MSE for the Poisson case with a mean number of points of 60, 100 and 300 and using the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$ are shown in Figures (5), (6) and (7). Note that each line in the graph represents

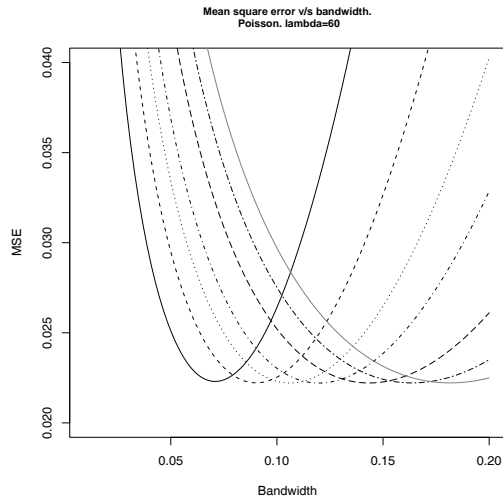


Figure 5: Mean square error versus bandwidth for the Poisson case with a mean of 60 points in the unit square. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

the mean of 100 repetitions.

The results for the MSE for the Matérn hard-core case with a mean number of points of 60, 100 and 300 and using the non-negative kernel family for orders $k = 2, 4, 6, 8, 12$ are shown in Figures (8), (9) and (10). Note that each line in the graph represents the mean of 100 repetitions.

The results for the MSE for the Thomas case with a mean number of points of 60, 100 and 300 and using the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$ and $R = 0.05, 0.075, 0.10$ are shown in Figures (11) to (19). Note that each line in the graph represents the mean of 100 repetitions.

In the case of the completely random (Poisson) process, it is shown that for a small number of points, a any kernel provides a smaller MSE than the Epanechnikov kernel. By using a larger bandwidth, it is also possible to see that as the order of the kernel increases, the zone in which the optimum bandwidth can be found is greater, in such a way that it does not depend, in such a crucial way, on a bandwidth value, and has a lower range to be chosen from. As the number of points increases, in the zone where the optimum bandwidth is found, the MSE obtained by the various kernels tend to be similar, and have less influence on the bandwidth as the order increases. In general, in the Poisson case, there is no gaining with a bigger kernel

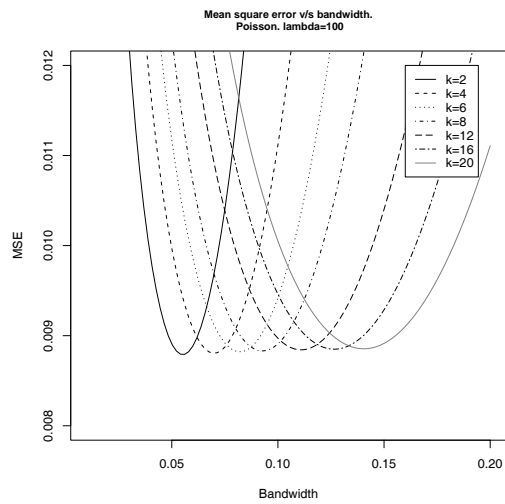


Figure 6: Mean square error versus bandwidth for the Poisson case with a mean of 100 points in the unit square. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

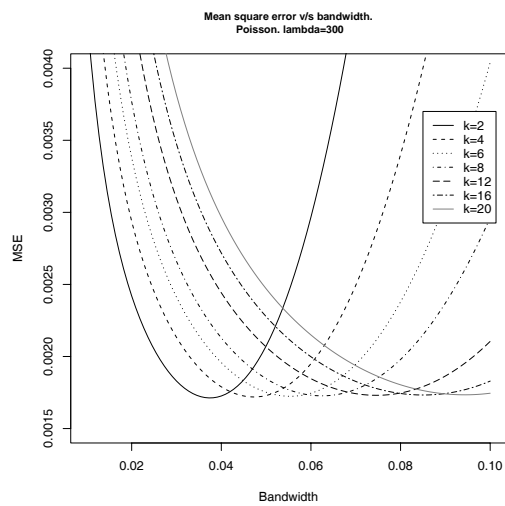


Figure 7: Mean square error versus bandwidth for the Poisson case with a mean of 300 points in the unit square. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

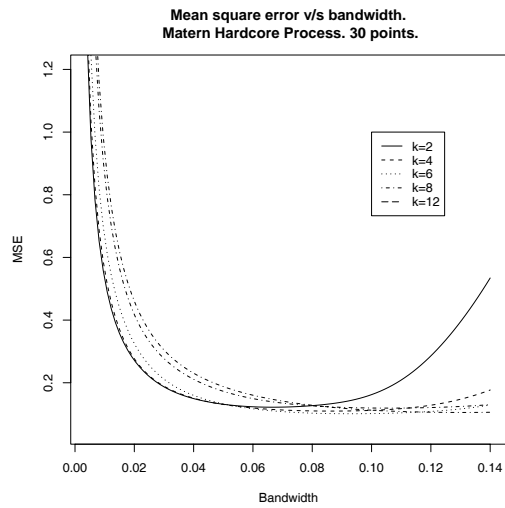


Figure 8: Mean square error versus bandwidth for the Matérn hard-core case with a mean of 60 points in the unit square. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

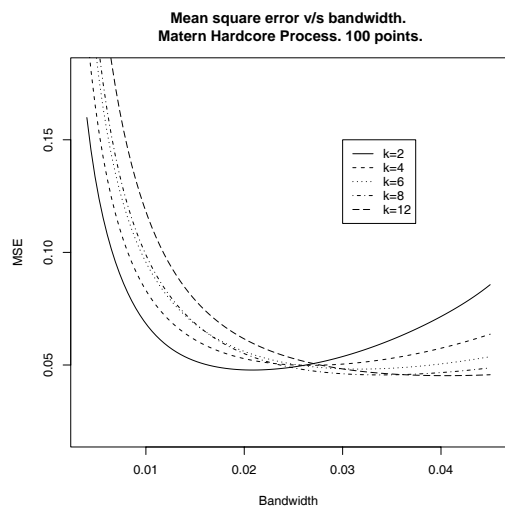


Figure 9: Mean square error versus bandwidth for the Matérn hard-core case with a mean of 100 points in the unit square. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

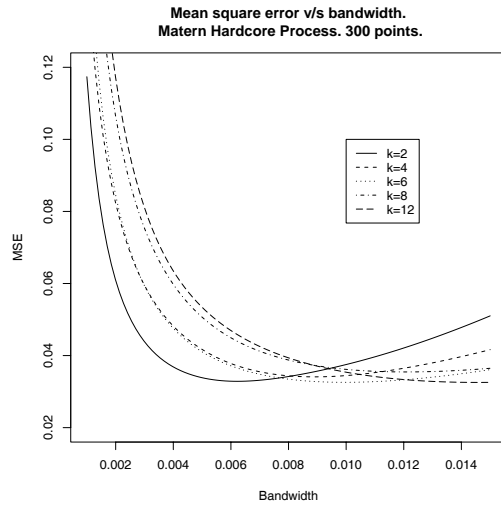


Figure 10: Mean square error versus bandwidth for the Matérn hard-core case with a mean of 300 points in the unit square. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

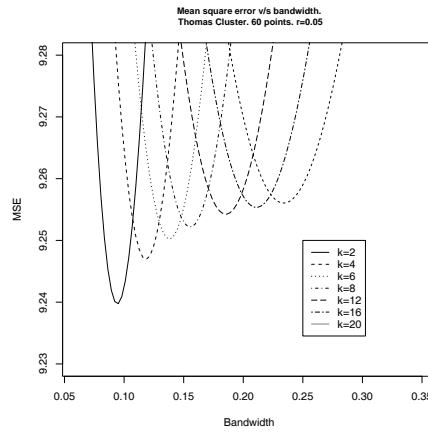


Figure 11: Mean square error versus bandwidth for the Thomas case with a mean of 60 points in the unit square with $R = 0.05$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

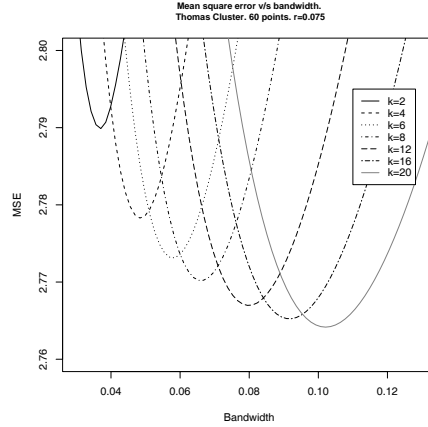


Figure 12: Mean square error versus bandwidth for the Thomas case with a mean of 60 points in the unit square with $R = 0.075$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

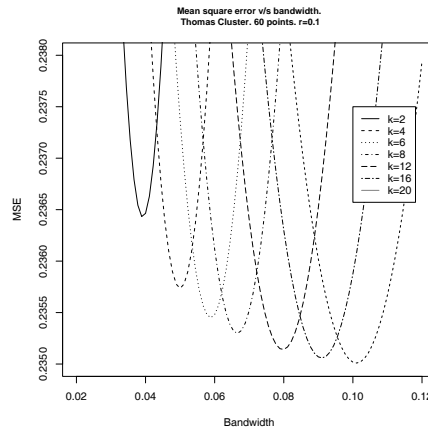


Figure 13: Mean square error versus bandwidth for the Thomas case with a mean of 60 points in the unit square with $R = 0.10$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

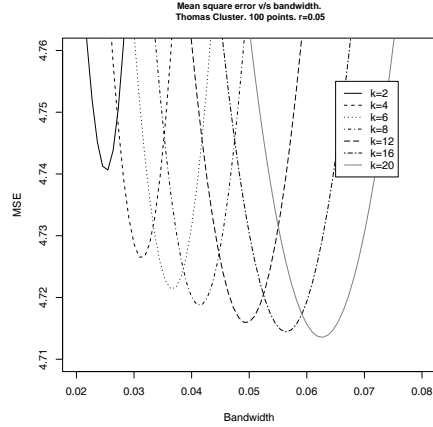


Figure 14: Mean square error versus bandwidth for the Thomas case with a mean of 100 points in the unit square with $R = 0.05$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

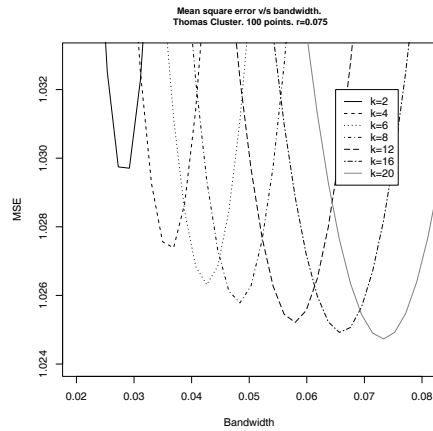


Figure 15: Mean square error versus bandwidth for the Thomas case with a mean of 100 points in the unit square with $R = 0.075$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

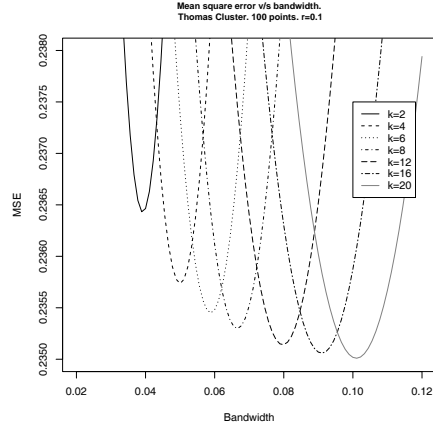


Figure 16: Mean square error versus bandwidth for the Thomas case with a mean of 100 points in the unit square with $R = 0.10$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

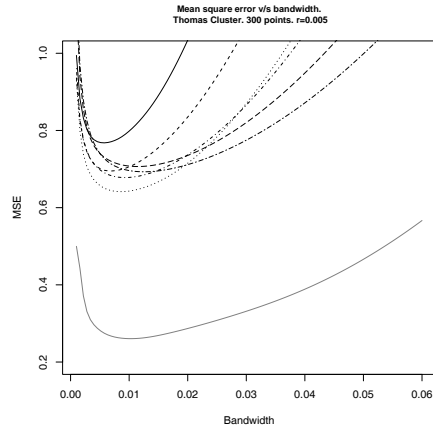


Figure 17: Mean square error versus bandwidth for the Thomas case with a mean of 300 points in the unit square with $R = 0.05$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

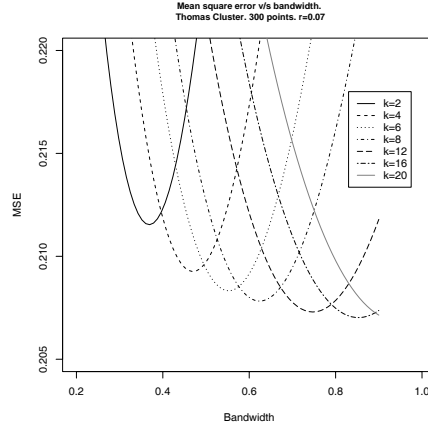


Figure 18: Mean square error versus bandwidth for the Thomas case with a mean of 300 points in the unit square with $R = 0.075$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

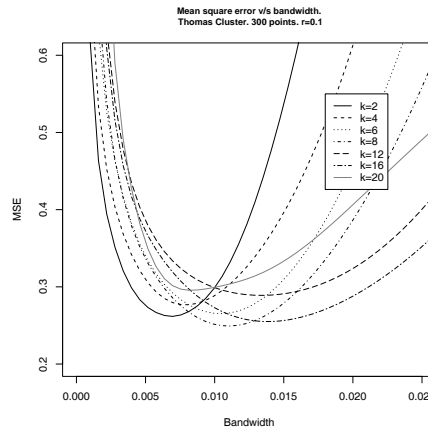


Figure 19: Mean square error versus bandwidth for the Thomas case with a mean of 300 points in the unit square with $R = 0.10$. Results for the non-negative kernel family for orders $k = 2, 4, 6, 8, 12, 16, 20$.

order. This was expected as bigger weights given by big orders to points close to the center can be compensated with bigger weights given by lower orders to far away points.

In the hard-core case, and for any sample size, we recommend using a large order kernel noting that a lower MSE is obtained for large order kernel values, as well as a wider zone from which to choose the optimum bandwidth.

The order of the kernels seems important in the cluster situations, where bigger orders provide better fits. This is more clear when the number of points increases together with the strength of the attraction.

7 Conclusions

In the area of spatial statistics, the Epanechnikov kernel has traditionally been used to estimate density functions such as the pair correlation or product density. This widely used is clearly due to the fact that the Epanechnikov kernel is the optimum 2nd order kernel.

The application of the non-negative kernel family (of which the Epanechnikov kernel is a special case), shows that in the case of a random process, better performance can be obtained with a higher order kernel, supposing that a small number of points are available for its estimation. As the number of points increases, the kernel chosen becomes irrelevant.

In the hard-core process, it is also a good choice to increase the kernel order to alleviate the problem of the jump near the inhibition distance, and to be able to choose the optimum bandwidth from a higher range.

In the case of the cluster process, the triweight kernel shows a very good performance for a sample of 50 to 100 points. Again, when the sample size is increased, the choice of kernel becomes less relevant, although the fact remains that an increase in the order of the kernel gives a wider range of bandwidth choice.

References

- Baddeley, A. (2003). *Analysis of spatial point processes*. Spatstat package. CRAN.
- Cressie, N.A. (1993). *Statistics for Spatial Data*. Wiley, Revised edition.
- Diggle, P.J. (2003). *Statistical Analysis of Spatial Point Processes*. Second edition. London: Edward Arnold.

- Epanechnikov, V. (1969). Nonparametric estimates of a multivariate probability density. *Theory Probab. Appl.*, **14**, 153-158.
- Fiksel, T. (1988). Edge-corrected density estimators for point processes. *Statistics*, **1**, 67-75.
- Gasser, T., Müller, H. & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society*, **B47**, 238-252.
- Mammitzsch, V. (1985). The fluctuation of kernel estimators under certain moment conditions. *Proceedings of 45th Session of ISI*, Amsterdam, Book I, 17-18.
- Müller, H. (1985). On the number of sign changes of a real function. *Periodica Mathematica Hungarica*, **16**, 209-213.
- Reinsch, C. (1967). Smoothing by spline function. *Numerical Mathematics*, **10**, 177-183.
- Rosenblatt, M. (1971). Curve Estimates. *Annals of Statistics*, **42**, 1815-1842.
- Stoyan, D. & Stoyan, H. (1985). On One of Matern's Hard-Core Point Process Models. *Math. Nachr.*, **122**, 205-214.
- Stoyan, D. & Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*. John Wiley & Sons, New York.
- Stoyan, D., Kendall, W. & Mecke, J. (1995). *Stochastic Geometry and its Applications*. John Wiley & Sons, New York. Second edition.