Annals of the Institute of Statistical Mathematics (2024) 76:313–331
https://doi.org/10.1007/s10463-023-00890-6

Non-parametric adaptive bandwidth selection for kernel estimators of spatial intensity functions

M. N. M. van Lieshout[1,2]

Received: 8 December 2022 / Revised: 17 October 2023 / Accepted: 23 October 2023 /
Published online: 22 December 2023

Abstract
We introduce a new fully non-parametric two-step adaptive bandwidth selection method for kernel estimators of spatial point process intensity functions based on the Campbell–Mecke formula and Abramson's square root law. We present a simulation study to assess its performance relative to other adaptive and global bandwidth selectors, investigate the influence of the pilot estimator and apply the technique to two data sets: A pattern of trees and an earthquake catalogue.

Keywords Adaptive kernel estimation · Bandwidth selection · Campbell–Mecke formula · Intensity function · Poisson leave-one-out cross-validation log likelihood · Point process

1 Introduction

The first step in any analysis of a spatial point pattern is usually estimating its intensity function (Diggle, 2014; Illian et al., 2008; Van Lieshout, 2019). To do so, various techniques exist. Perhaps the oldest is quadrat counting (Du Rietz, 1929) in which one simply reports the number of points falling in each quadrat scaled by the quadrat volume. Instead of fixed quadrats, one might use the cells in a tessellation formed by the pattern itself as the units for counting (Barr and Schoenberg, 2010; Moradi et al., 2019; Ord, 1978; Schaap and Van de Weygaert, 2000). However, the most popular technique by far seems to be kernel estimation (Diggle, 1985).

This research was supported by The Netherlands Organisation for Scientific Research NWO (project DEEP.NL.2018.033).

✉ M. N. M. van Lieshout
m.n.m.van.lieshout@cwi.nl

[1] CWI, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands

[2] Department of Applied Mathematics, University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands

The classic kernel estimator for the spatial intensity function uses a constant bandwidth. However, intuitively, such a 'one size fits all' approach would tend to over-smooth in dense areas, while not smoothing enough in sparse regions. As a consequence, finer detail in areas with many points may be lost, whereas the few points in sparser areas might give rise to spurious hot spots. Motivated by similar considerations for density estimation of one-dimensional random variables, Abramson (1982) proposed to use a kernel smoother in which the bandwidth at each observation is weighted by a power of the density at that observation. Doing so reduces the bias significantly, at least asymptotically (Hall et al., 1995). In the spatial context of intensity function estimation, the same idea applies (Davies et al., 2018; Van Lieshout, 2021).

Clearly, the crucial parameter, both in classic and adaptive kernel estimation, is the bandwidth. It is often chosen by visual inspection or by using a rule of thumb (see e.g. Baddeley et al. (2015, Section 6.5), Illian et al. (2008, Section 3.3) or Scott (1992, Section 6)). Obviously, though, such procedures are rather ad hoc and subjective.

A more rigorous class of techniques is based on asymptotics. For instance for classic kernel estimators, Brooks and Marron (1991) considered a Poisson process on the real line and assumed a simple multiplicative model for the intensity function to derive an asymptotically optimal least-squares cross-validation estimator when the number of points tends to infinity. Lo (2017) picked up the baton and studied the asymptotic (integrated) mean squared error in any dimension without imposing a specific intensity model, again in the regime that the number of points goes to infinity. Van Lieshout (2020) generalized Lo's work to point processes that may exhibit interaction between the points under the assumption that replicated patterns are available so that infill asymptotics apply. Davies et al. (2018) considered asymptotic expansions for a spatial analogue of the Abramson estimator for Poisson processes as the number of points increases; Van Lieshout (2021) studied infill asymptotics that allow for interaction between the points. It is important to note that the resulting optimal bandwidths depend on the unknown intensity function and cannot be computed in practice without resorting to iterative techniques.

Most practical procedures to select a suitable global bandwidth rely on a specific model. For example, leave-one-out cross-validation (Loader, 1999, Section 5.3) assumes the data come from a Poisson process. Another common approach is to minimize the mean squared error in state estimation for a planar stationary isotropic Cox process (Diggle, 1985). The disadvantage of such techniques is that the underlying assumption may not hold for the pattern at hand, which motivated Cronie and Van Lieshout (2018) to propose a fully non-parametric technique based on the Campbell–Mecke formula (Chiu et al., 2013). Compared to leave-one-out cross-validation, this approach is also computationally cheaper and does not require numerical approximation of integrals. A simulation study that investigates the relative performance in practice can be found in Cronie and Van Lieshout (2018).

In this article, we extend the leave-one-out cross-validation and Campbell–Mecke-based approaches to adaptive bandwidth selection. The plan is as follows. Section 2 recalls crucial concepts and fixes notation. In Sect. 3, we discuss adaptive kernel estimators and present algorithms for selecting the bandwidth. The results of a simulation

study into the efficacy of the new approaches are given in Sect. 4; applications to two data sets concerning, respectively, induced earthquakes and forestry, are presented in Sects. 5.1 and 5.2. The paper closes with a discussion on computational complexity and ideas for future research.

2 Preliminaries and notation

First, let us introduce some notation. Let Ψ be a simple point process (Chiu et al., 2013) in d-dimensional Euclidean space $\mathbb{R}^d$ that is observed in a bounded, non-empty and open subset W of $\mathbb{R}^d$. We assume that the first order moment measure Λ of Ψ defined by

$$\Lambda(A) = \mathbb{E}\left(\sum_{x\in\Psi} 1\{x \in A\}\right),$$

the expected number of points of Ψ that fall in Borel subsets A of $\mathbb{R}^d$, exists as a locally finite Borel measure and is absolutely continuous with respect to d-dimensional Lebesgue measure ℓ^d with a Radon–Nikodym derivative $\lambda : \mathbb{R}^d \to [0, \infty)$. We will refer to the function λ as the *intensity function* of Ψ.

The *kernel estimator* of the intensity function of a point process was introduced by Diggle (1985) as

$$\widehat{\lambda}(x_0;h,\Psi,W) = \frac{1}{h^d}\sum_{y\in\Psi\cap W}\kappa\left(\frac{x_0-y}{h}\right), \quad x_0 \in W, \tag{1}$$

possibly divided by a global edge correction factor

$$w(x_0,h,W) = \frac{1}{h^d}\int_W \kappa\left(\frac{x_0-z}{h}\right)\mathrm{d}z.$$

An alternative, local, edge correction can be found in Van Lieshout (2012). The function $\kappa : \mathbb{R}^d \to [0, \infty)$ is supposed to be a kernel, that is, a d-dimensional probability density function (Silverman, 1986, p. 13) that is even in all its arguments. When κ is positive in a neighbourhood of the origin, since W is assumed to be open, the global edge correction factor is nonzero for all $x_0 \in W$.

The crucial parameter in (1) is the *bandwidth* $h > 0$, which determines the amount of smoothing. For large h, the mass of κ is spread far and wide, which reduces the variance but may lead to a large bias. For small h, the mass of κ is concentrated around the observed points of $\Psi \cap W$. Thus, the bias is reduced at the price of a larger variance.

Popular choices of kernel include those belonging to the Beta class (Hall et al., 2004)

$$\kappa^\gamma(x) = \frac{\Gamma(d/2+\gamma+1)}{\pi^{d/2}\Gamma(\gamma+1)}(1-x^Tx)^\gamma\, 1\{x \in B(0,1)\}, \quad x \in \mathbb{R}^d, \tag{2}$$

for $\gamma \geq 0$. Here $B(0, 1)$ is the closed unit ball in $\mathbb{R}^d$ centred at the origin. Beta kernels are supported on the compact unit ball and their smoothness is governed by the parameter γ. Indeed, the box kernel defined by $\gamma = 0$ is constant and therefore continuous on the interior of the unit ball; the Epanechnikov kernel corresponding to the choice $\gamma = 1$ is Lipschitz continuous. For $\gamma > k$, the function κ^γ is k times continuously differentiable on $\mathbb{R}^d$. An alternative with unbounded support is the Gaussian kernel

$$\kappa(x) = (2\pi)^{-d/2} \exp\left(-x^T x/2\right), \quad x \in \mathbb{R}^d. \tag{3}$$

Bandwidth selection techniques tend to be based either on asymptotic expansions (Van Lieshout, 2020) or specific model assumptions (Baddeley et al., 2015; Berman and Diggle, 1989; Loader, 1999). For example, the widely used leave-one-out cross-validation technique (Baddeley et al., 2015; Loader, 1999) assumes that Ψ is an inhomogeneous Poisson process with log likelihood function

$$\sum_{x \in \Psi \cap W} \log \lambda(x) - \int_W \lambda(u)\, \mathrm{du}.$$

Upon plugging in an estimator for the unknown intensity function, the leave-one-out cross-validation log likelihood reads

$$L_\kappa(h; \Psi, W) = \sum_{x \in \Psi \cap W} \log \widehat{\lambda}(x; h, \Psi \setminus \{x\}, W) - \int_W \widehat{\lambda}(u; h, \Psi, W)\, \mathrm{du}, \tag{4}$$

which is then maximized over h to select the bandwidth. Note that conditions have to be imposed to ensure that the function $\widehat{\lambda}$ is strictly positive.

In a recent paper, Cronie and Van Lieshout (2018) proposed a non-parametric alternative based on the Campbell–Mecke formula (Chiu et al., 2013, p. 130). For the function $f : \mathbb{R}^d \to \mathbb{R}^+$, $f(x) = 1\{x \in W\}/\lambda(x)$ known as the Stoyan–Grabarnik statistic (Stoyan and Grabarnik, 1991), which is measurable if $\lambda(x) > 0$ for all $x \in W$, the Campbell–Mecke formula states that

$$\mathbb{E}\left(\sum_{x \in \Psi \cap W} \frac{1}{\lambda(x)} \right) = \int_W \frac{1}{\lambda(x)} \lambda(x)\, \mathrm{dx} = \ell(W). \tag{5}$$

To select a bandwidth, one may simply replace λ by an estimator $\widehat{\lambda}(\cdot; h, \Psi, W)$ in the left hand side of the equation and minimize the discrepancy between $\ell(W)$ and the sum of $\widehat{\lambda}(x; h, \Psi, W)^{-1}$ over points x in $\Psi \cap W$. Formally, set

$$T_\kappa(h; \Psi, W) = \begin{cases} \displaystyle\sum_{x \in \Psi \cap W} \frac{1}{\widehat{\lambda}(x; h, \Psi, W)}, & \Psi \cap W \neq \emptyset, \\ \ell(W), & \text{otherwise}, \end{cases}$$

and choose bandwidth $h > 0$ by minimizing

$$F_\kappa(h;\Psi,W) = \left|T_\kappa(h;\Psi,W) - \ell(W)\right|. \tag{6}$$

Since W is assumed to be open and κ is strictly positive in a neighbourhood of the origin for the kernels considered in this paper, T_κ and therefore (6) is well-defined with or without edge correction. Moreover, without edge correction, a zero point of the equation (6) exists (Cronie and Van Lieshout, 2018, Theorem 1). Briefly, when $h \to 0$, the kernel estimator places more and more mass on the observations, so that $T_\kappa(h;\Psi,W)$ tends to zero. When on the other hand h grows to infinity, the variance of the kernel estimator decreases as it spreads its mass out more and more uniformly over W, at the cost of an increase in bias. The function $T_\kappa(h;\Psi,W)$ now tends to infinity. Thus, requiring that $T_\kappa(h;\Psi,W)$ be equal to $\ell(W)$ can be interpreted as striking a balance between bias and variance. For a graphical illustration, see the middle panel in the first row of Fig. 2.

Our goal in the next section is to extend the ideas outlined above to adaptive bandwidths.

3 Adaptive bandwidth selection

3.1 The Abramson adaptive kernel estimator

As noted in the introduction, ideally the bandwidth should be adapted to the local density of points. In the context of random variables, Abramson (1982) proposed to scale the bandwidth in proportion to a power of the probability density function. Similarly, in the point pattern setting, an *adaptive kernel estimator* (Davies et al., 2018; Van Lieshout, 2021) is defined as

$$\widehat{\lambda}_A(x_0;h,\Psi,W) = \sum_{y\in\Psi\cap W} \frac{1}{c(y)^d h^d}\kappa\left(\frac{x_0-y}{hc(y)}\right) w(y,h,W)^{-1} \tag{7}$$

where

$$c(y) = \left(\frac{\lambda(y)}{\prod_{z\in\Psi\cap W}\lambda(z)^{1/N(\Psi\cap W)}}\right)^{\alpha}, \tag{8}$$

$N(\Psi\cap W)$ denotes the number of points of Ψ that fall in W and $w(y, h, W)$ is an edge correction weight. The power α is set to $-1/2$ when considering asymptotic expansions (Abramson, 1982; Van Lieshout, 2021). In practice, a power $\alpha = -1/d$ when $d \geq 2$, is considered best (Bowman and Foster, 1993; Silverman, 1986). In the simulation studies to be presented in Sect. 4, we will work in the plane and the two powers are identical.

Let us make a few observations. First, points y located in regions with a low intensity are given a larger bandwidth $hc(y)$ than those in high-intensity regions, as desired. Secondly, we must assume that $\lambda(y) > 0$ for each $y \in \Psi\cap W$. The normalization by the geometric mean is used to obtain a dimensionless quantity for the bandwidth (Silverman, 1986). When focussing on a single point x_0 (Abramson,

1982; Van Lieshout, 2021), one could normalize simply by $\lambda(x_0)$. Finally, edge correction techniques carry over in a straightforward manner. For example, a local edge correction weight factor in this context takes the form

$$w(y,h,W) = \frac{1}{c(y)^d h^d} \int_W \kappa\left(\frac{z-y}{hc(y)}\right) \mathrm{d}z$$

and is mass preserving:

$$\int_W \widehat{\lambda}_A(z;h,\Psi,W)\,\mathrm{d}z = N(\Psi \cap W).$$

Since the local bandwidth $hc(y)$ depends on the unknown intensity function, (7) cannot be calculated. A common solution is to estimate $c(y)$ by plugging-in a *pilot intensity estimator* (Chacón and Duong, 2018; Silverman, 1986; Wand and Jones, 1994). For example, one could estimate $\lambda(y)$ by a global bandwidth kernel estimator of the form (1) and set

$$\widehat{c}(y;h,\Psi,W) = \left(\frac{\widehat{\lambda}(y;h,\Psi,W)}{\prod_{z\in\Psi\cap W}\widehat{\lambda}(z;h,\Psi,W)^{1/N(\Psi\cap W)}}\right)^{\alpha}.$$

3.2 Adaptive bandwidth selection algorithms

The two global bandwidth selection algorithms presented in Sect. 2 can be modified for adaptive bandwidth selection by means of a two-step approach. In the first step, (4) or (6) is used to select a global bandwidth h_g. In this step, no edge correction is applied as the clearest optimum is obtained that way (Cronie and Van Lieshout, 2018). This h_g is then used to calculate an edge-corrected pilot estimate $\widehat{\lambda}_p$, which in turn is plugged in (8) to obtain $\widehat{c}(y)$.

In the second step, we apply (4) or (6) to $\widehat{\lambda}_A$ with local bandwidths $h\widehat{c}(y)$ and optimize over h. Again, no edge correction is applied in the bandwidth selection phase. Having selected the optimal h_a, (7) is calculated with local edge correction.

The following results guarantee the validity of the second step. For the first step, we refer to Theorem 1 in Cronie and Van Lieshout (2018).

Theorem 1 *Let ψ be a locally finite point pattern of distinct points in $\mathbb{R}^d$ observed in some non-empty open and bounded window W such that $\psi \cap W \neq \emptyset$. Let κ be the Gaussian kernel or a Beta kernel with $\gamma > 0$, and $w \equiv 1$. Write $\widehat{\lambda}_A$ for the Abramson estimator (7) with*

$$c(y;\psi,W) = \left(\frac{1}{\lambda_p(y)} \prod_{z\in\psi\cap W} \lambda_p(z)^{1/N(\psi\cap W)}\right)^{1/2}$$

for some pilot estimates $\lambda_p(y)$ that are strictly positive for all $y \in \psi \cap W$. Then the criterion function

$$T_\kappa(h;\psi, W) = \sum_{x\in\psi\cap W} \frac{1}{\widehat{\lambda}_A(x;h,\psi,W)}$$

is a continuous function of h on $(0,\infty)$. *For the box kernel, it is piecewise continuous. In all cases,*

$$\lim_{h\to 0} T_\kappa(h;\psi, W) = 0; \quad \lim_{h\to\infty} T_\kappa(h;\psi, W) = \infty.$$

Proof We will first look at the limit as $h \to 0$. For all $h > 0$ and $x \in \psi \cap W$,

$$\widehat{\lambda}_A(x;h,\psi,W) \geq \kappa(0)\, c(x;\psi,W)^{-d} h^{-d} > 0.$$

Here, we use that since the pilot estimator λ_p is strictly positive on the non-empty pattern $\psi \cap W$, so is $c(\cdot)$. Also, for all kernels considered, $\kappa(0) > 0$. Consequently,

$$T_\kappa(h;\psi, W) = \sum_{x\in\psi\cap W} \frac{1}{\widehat{\lambda}_A(x;h,\psi,W)} \leq h^d \sum_{x\in\psi\cap W} \frac{c(x;\psi,W)^d}{\kappa(0)}.$$

The right-most expression and therefore $T_\kappa(h;\psi, W)$ tends to 0.

Next let $h \to \infty$. For the box, Beta and Gaussian kernels, $\kappa(\cdot) \leq \kappa(0)$. We already observed that $c(y;\psi, W)$ is strictly positive for $y \in \psi \cap W$ since by assumption λ_p is. Moreover, it does not depend on h and therefore

$$\begin{aligned} T_\kappa(h;\psi, W) &= \sum_{x\in\psi\cap W} \frac{h^d}{\sum_{y\in\psi\cap W} c(y;\psi,W)^{-d}\kappa\left(\frac{x-y}{hc(y;\psi,W)}\right)} \\ &\geq h^d \sum_{x\in\psi\cap W} \frac{1}{\sum_{y\in\psi\cap W} c(y;\psi,W)^{-d}\kappa(0)}. \end{aligned}$$

The right-most expression and therefore $T_\kappa(h;\psi, W)$ tends to ∞ when $\psi \cap W$ is non-empty.

It remains to look at continuity properties. Both the Beta kernels κ^γ with $\gamma > 0$ and the Gaussian kernel are continuous on $\mathbb{R}^d$. The box kernel is discontinuous on the unit sphere $\partial B(0,1)$ only. The function $h \to h^{-1}$ is continuous on $(0,\infty)$. Therefore, for fixed $z \in \mathbb{R}^d$, the function $h \to \kappa(z/h)$ is also continuous when κ is a Gaussian kernel or a Beta kernel with $\gamma > 0$. For the box kernel, this function is piecewise continuous, having a discontinuity at $h = ||z||$. Observe that, since $\psi \cap W$ is non-empty by assumption, $\kappa(0) > 0$ and the pilot estimates $\lambda_p(x)$ are strictly positive for every $x \in \psi \cap W$, also the estimates $\widehat{\lambda}_A(x;h,\psi,W)$ are strictly positive for $x \in \psi \cap W$ and $h > 0$. We conclude that, as a function of h on $(0,\infty)$, $T_\kappa(h;\psi, W)$ is continuous for Gaussian kernels and Beta kernels with $\gamma > 1$, piecewise continuous for the box kernel. □

The theorem above implies that the criterion function $T_\kappa(h;\psi, W)$ attains all positive values, in particular $\ell(W)$. Usually, but not always, it does so for a unique h. In case of multiple solutions, one may pick the smallest.

Theorem 2 *Let ψ be a locally finite point pattern of distinct points in $\mathbb{R}^d$ observed in some non-empty open and bounded window W such that $\psi \cap W$ has at least two points. Let κ be the Gaussian kernel. Write $\widehat{\lambda}_A$ for the Abramson estimator (7) with*

$$c(y;\psi, W) = \left(\frac{1}{\lambda_p(y)} \prod_{z \in \psi \cap W} \lambda_p(z)^{1/N(\psi \cap W)} \right)^{1/2}$$

for some pilot estimates $\lambda_p(y)$ that are strictly positive for all $y \in \psi \cap W$. Then, the criterion function

$$L_\kappa(h;\psi, W) = \sum_{x \in \psi \cap W} \log \widehat{\lambda}_A(x;h, \psi \setminus \{x\}, W) - \int_W \widehat{\lambda}_A(u;h, \psi, W) du$$

is a continuous function of h on $(0, \infty)$. Moreover,

$$\lim_{h \to 0} L_\kappa(h;\psi, W) = -\infty; \quad \lim_{h \to \infty} L_\kappa(h;\psi, W) = -\infty.$$

Proof The Gaussian kernel is continuous on $\mathbb{R}^d$. The function $h \to h^{-1}$ is continuous on $(0, \infty)$. Therefore, for fixed $z \in \mathbb{R}^d$, the function $h \to \kappa(z/h)$ is continuous. Since the logarithm is continuous on $(0, \infty)$, the first term in $L_\kappa(h;\psi, W)$ is continuous in h provided the argument is strictly positive. Now, by assumption, the pilot estimates $\lambda_p(x)$ are strictly positive. Therefore, the same is true for $c(y;\psi, W)$. Under the assumption that $\psi \cap W$ has at least two points, $\psi \setminus \{x\}$ will still contain a point in W and, noting that the Gaussian kernel is strictly positive, one concludes that the argument of the logarithm is strictly positive. Turning to the integral term, continuity follows from the dominated convergence theorem, as $\kappa(\cdot) \leq (2\pi)^{-d/2}$ is bounded. We conclude that, as a function of h on $(0, \infty)$, $L_\kappa(h;\psi, W)$ is continuous.

Next turn to the limits. First let $h \to \infty$. Then $-d \log h \to -\infty$ while

$$\log \left[\sum_{x \neq y \in \psi \cap W} \frac{1}{c(y;\psi, W)^d} \kappa\left(\frac{x - y}{hc(y;\psi, W)} \right) \right] \to \log \left[\sum_{x \neq y \in \psi \cap W} \frac{\kappa(0)}{c(y;\psi, W)^d} \right].$$

Since $h^{-d} c(y)^{-d} \kappa((\cdot - y)/(hc(y;\psi, W)))$ is a probability density function, its integral takes values in between zero and one, so that the contribution from the integral term in $L_\kappa(h;\psi, W)$ is between 0 and the negative number of points in $\psi \cap W$. Collecting all terms, we find that $\lim_{h \to 0} L_\kappa(h;\psi, W) = -\infty$.

Finally, let $h \to 0$. As already noted, the contribution from the integral term in $L_\kappa(h;\psi, W)$ is bounded. For the logarithmic term, recall that

$$\lim_{h \to 0} \frac{1}{h^d c(y;\psi, W)^d} \kappa\left(\frac{z}{hc(y;\psi, W)} \right) = 0$$

when $||z|| \neq 0$. By the assumption that $\psi \cap W$ contains at least two points, $L_\kappa(h;\psi, W)$ tends to $-\infty$ as $h \to 0$. □

Table 1 Intensity functions on the open unit square. Here the set S is the annulus $S = \{(x, y) \in (0, 1)^2 : \left|((x - 0.5)^2 + (y - 0.6)^2)^{1/2} - 0.1\right| < 0.02\}$

Constant	Trend	High-contrast feature
$\lambda_1(x, y) \equiv 50$	$\lambda_3(x, y) = 10 + 80x$	$\lambda_5(x, y) = \lambda_3 + 1000 \times 1_S(x, y)$
$\lambda_2(x, y) \equiv 250$	$\lambda_4(x, y) = 30 + 440x$	$\lambda_6(x, y) = \lambda_4 + 3000 \times 1_S(x, y)$

The validity of the bandwidth selection steps follows from the observation that one may restrict h to a closed and bounded interval on which the continuous function $L_\kappa(h;\psi, W)$ attains its maximum. For compact support kernels such as (2), $\hat{\lambda}_A(x;h, \psi \setminus \{x\})$ may be zero and the logarithmic term in $L_\kappa(h;\psi, W)$ ill-defined.

4 Simulation study

To evaluate the performance of the adaptive bandwidth selection approach and investigate the influence of the pilot estimates, we conduct a simulation study. We consider three types of intensity functions on the unit square in $\mathbb{R}^2$: A constant intensity, a gradual polynomial trend in the horizontal direction and a gradual trend combined with a high intensity feature. For each type of function, we set the parameters in such a way that realizations contain approximately 50 or 250 points. Doing so, we obtain the intensity functions summarized in Table 1.

A convenient way to obtain realizations of point processes with spatially varying intensity functions λ_i is to apply independent thinning to realizations of a stationary point process whose intensity function is known explicitly. Here, we choose a Poisson process, a Matérn cluster process and a Matérn hard core process (Matérn, 1986). We will need the notation $\bar{\lambda}_i = \sup_{(x,y)\in(0,1)^2} \lambda_i(x, y)$.

Poisson process

Let X be a homogeneous Poisson process with intensity function $\bar{\lambda}_i$. Then, its independent thinning with retention probabilities $\lambda_i(x, y)/\bar{\lambda}_i$ is a heterogeneous Poisson process with intensity function λ_i, $i = 1, \ldots, 6$.

Matérn cluster process

Let X_p be a homogeneous Poisson process with intensity κ on $(-0.01, 0.01)^2$. Assume that each 'parent' point $z \in X_p$ generates a Poisson number of 'daughter' points, say with mean ν in the closed ball $B(z, 0.01)$ of radius 0.01 around z and write X for the union of daughter points falling in $(0, 1)^2$. Then, X is homogeneous, and has constant intensity $\kappa\nu$ on $(0, 1)^2$. We will consider two degrees of clustering:

- Parent intensity $\kappa = \bar{\lambda}_i/5$, mean number of daughters $\nu = 5$ in a ball of radius 0.01 around the parent;

Table 2 Mean integrated squared error relative to expected number of points of classic kernel estimates over 100 simulations using a Gaussian kernel with local edge correction and bandwidth chosen by (4) for various point process models having intensity functions λ_i, $i = 1, \ldots, 6$

λ	Poisson	Cluster $\nu = 5$	Cluster $\nu = 10$	Hard core $\nu = 0.9$	Hard core $\nu = 0.5$
λ_1	6.0	3,170.4	6,876.2	5.4	3.2
λ_2	15.7	3,817.8	6,738.7	12.8	7.7
λ_3	9.0	775.1	2,802.4	6.6	5.1
λ_4	16.8	1,295.0	3,981.1	16.3	12.3
λ_5	398.9	433.9	508.4	398.7	443.4
λ_6	617.3	656.4	861.5	617.2	618.9

- Parent intensity $\kappa = \bar{\lambda}_i/10$, mean number of daughters $\nu = 10$ in a ball of radius 0.01 around the parent.

In either case, independent thinning with retention probabilities $\lambda_i(x, y)/\bar{\lambda}_i$ results in a point process X having intensity function λ_i, $i = 1, \ldots, 6$.

Type II Matérn hard core process

Let X_g be a homogeneous Poisson process with intensity κ on $(-r, r)^2$ and assign each 'ground' point $z \in X_g$ a mark according to the uniform distribution on (0, 1) independently of other points. Keep a point $z \in X_g \cap (0, 1)^2$ if no other point of X_g with a larger mark lies within distance $r > 0$. The resulting point process X is homogeneous and has constant intensity $(1 - e^{-\kappa\pi r^2})/(\pi r^2)$ on $(0, 1)^2$. We will consider two degrees of repulsion:

- Ground intensity $\kappa = -10\bar{\lambda}_i \log \nu$ with $\nu = 0.9$ and hard core distance $r = (10\pi\bar{\lambda}_i)^{-1/2}$;
- Ground intensity $\kappa = -2\bar{\lambda}_i \log \nu$ with $\nu = 0.5$ and hard core distance $r = (2\pi\bar{\lambda}_i)^{-1/2}$.

In both cases, independent thinning with retention probabilities $\lambda_i(x, y)/\bar{\lambda}_i$ results in a point process X having intensity function λ_i, $i = 1, \ldots, 6$.

The results of the simulation study are presented in Tables 2, 3, 4, 5, 6 and 7. For each intensity function and each point process model, we generated 100 simulations in the unit square and calculated the optimal global and adaptive bandwidths using a Gaussian kernel and criterion functions (4) and (6). In the adaptive case, we used two different pilot estimators. The tabulated values are the mean integrated squared errors after local edge correction over the 100 patterns scaled by the expected number of points. All calculations were done in the R-package `spatstat` (Baddeley et al., 2015) to which we contributed the function `bw.CvL.adaptive`.

A line-by-line comparison of Tables 2 and 3 confirms the conclusions in Cronie and Van Lieshout (2018). For a Poisson process, the leave-one-out cross-validation approach which assumes the process is Poisson is best. For clustered

Table 3 Mean integrated squared error relative to expected number of points of classic kernel estimates over 100 simulations using a Gaussian kernel with local edge correction and bandwidth chosen by (6) for various point process models having intensity functions $\lambda_i, i = 1, \ldots, 6$

λ	Poisson	Cluster $\nu = 5$	Cluster $\nu = 10$	Hard core $\nu = 0.9$	Hard core $\nu = 0.5$
λ_1	10.2	23.7	36.6	10.4	7.8
λ_2	32.4	63.4	74.9	29.5	23.5
λ_3	10.6	20.3	27.5	8.7	8.0
λ_4	23.7	55.1	73.5	23.6	19.6
λ_5	418.4	433.2	448.3	416.0	484.2
λ_6	663.5	697.7	720.4	664.2	666.3

Table 4 Mean integrated squared error relative to expected number of points of adaptive kernel estimates over 100 simulations using a Gaussian kernel with local edge correction and bandwidth chosen by (4) with classic kernel pilot estimate with bandwidth chosen by (4) for various point process models having intensity functions $\lambda_i, i = 1, \ldots, 6$

λ	Poisson	Cluster $\nu = 5$	Cluster $\nu = 10$	Hard core $\nu = 0.9$	Hard core $\nu = 0.5$
λ_1	11.5	3,661.9	7,730.7	10.4	5.4
λ_2	35.1	3,586.5	7,281.5	28.4	14.6
λ_3	17.0	1,293.2	3,938.8	11.9	8.3
λ_4	37.9	2,298.4	4,769.7	35.3	22.0
λ_5	343.4	745.8	1,410.9	337.7	419.7
λ_6	438.7	1,234.8	3,015.8	426.8	420.2

Table 5 Mean integrated squared error relative to expected number of points of adaptive kernel estimates over 100 simulations using a Gaussian kernel with local edge correction and bandwidth chosen by (4) with classic kernel pilot estimate with bandwidth chosen by (6) for various point process models having intensity functions $\lambda_i, i = 1, \ldots, 6$

λ	Poisson	Cluster $\nu = 5$	Cluster $\nu = 10$	Hard core $\nu = 0.9$	Hard core $\nu = 0.5$
λ_1	13.1	3,370.5	7,413.7	11.9	6.1
λ_2	42.2	3,193.3	6,908.3	34.4	18.1
λ_3	17.7	1,087.1	3,480.3	12.7	9.1
λ_4	41.3	2,086.7	4,594.0	38.6	24.9
λ_5	350.2	531.6	751.8	344.3	436.1
λ_6	443.9	909.6	1,902.7	432.6	430.5

patterns, the Cronie–Van Lieshout approach is far better, unless the intensity exhibits a strong feature (λ_5 and λ_6). In the latter case, both approaches have similar mean squared error. For repulsive patterns, the approach based on (4) is somewhat better than that based on (6).

Next, let us compare Table 2 with Tables 4 and 5. As expected, when the intensity function is constant, an adaptive approach to bandwidth selection tends to be

Table 6 Mean integrated squared error relative to expected number of points of adaptive kernel estimates over 100 simulations using a Gaussian kernel with local edge correction and bandwidth chosen by (6) with classic kernel pilot estimate with bandwidth chosen by (4) for various point process models having intensity functions $\lambda_i, i = 1, \ldots, 6$

λ	Poisson	Cluster $\nu = 5$	Cluster $\nu = 10$	Hard core $\nu = 0.9$	Hard core $\nu = 0.5$
λ_1	14.0	29.2	41.3	14.2	9.6
λ_2	44.9	77.8	85.1	39.5	28.5
λ_3	18.2	41.6	46.5	14.5	12.0
λ_4	47.9	100.4	125.2	46.0	35.5
λ_5	355.7	477.7	529.0	349.5	464.4
λ_6	463.0	651.2	780.3	452.8	450.4

Table 7 Mean integrated squared error relative to expected number of points of adaptive kernel estimates over 100 simulations using a Gaussian kernel with local edge correction and bandwidth chosen by (6) with classic kernel pilot estimate with bandwidth chosen by (6) for various point process models having intensity functions $\lambda_i, i = 1, \ldots, 6$

λ	Poisson	Cluster $\nu = 5$	Cluster $\nu = 10$	Hard core $\nu = 0.9$	Hard core $\nu = 0.5$
λ_1	15.7	35.5	51.5	16.0	10.9
λ_2	52.1	106.4	120.1	46.6	35.0
λ_3	19.3	44.3	52.2	15.3	12.9
λ_4	50.9	132.1	172.2	49.6	38.0
λ_5	361.6	437.2	464.7	355.1	471.8
λ_6	468.0	617.8	723.3	457.3	457.7

worse than a global one. The same remark applies when the intensity function is slowly varying. However, for λ_5 and λ_6, it is better to use an adaptive approach unless the pattern exhibits clustering. Regarding the influence of the pilot estimates, note that using (4) gives slightly better results than (6) for Poisson and repulsive point patterns, whereas the Campbell–Mecke-based approach mitigates the increase in mean squared error for clustered patterns somewhat.

When comparing Table 3 with Tables 6 and 7, as before the adaptive approach to bandwidth selection tends to be worse than a global one for constant and slowly varying intensity functions. However, for λ_5 and λ_6, it is better to use an adaptive approach, unless the pattern exhibits clustering. Regarding the influence of the pilot estimates, as before (4) gives slightly better results than (6) for Poisson and repulsive point patterns, whereas the Campbell–Mecke-based approach mitigates the increase in mean squared error for clustered patterns somewhat.

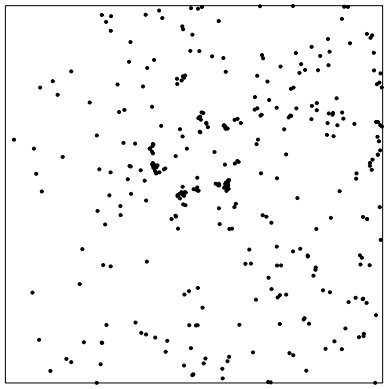

Fig. 1 Realization from a Matérn cluster process with intensity function λ_6 based on a ground process that has on average $\nu = 5$ daughters in a ball of radius $r = 0.01$ around each parent

A comparison of Table 4 with Table 6 or of Table 5 with Table 7 confirms that the criterion (4) is somewhat better suited to Poisson and regular patterns, whereas (6) is far better in the case of clustering.

To illustrate the different approaches and explain why the leave-one-out cross-validation approach is not appropriate for clustered patterns, Fig. 1 shows a simulation from a Matérn cluster process with intensity function λ_6 based on a ground process that has on average $\nu = 5$ daughters in a ball of radius $r = 0.01$ around each parent. Both the gradient in intensity from left to right, the increased density of points near the centre of the plot and the clustering are readily apparent.

The true (top left panel) and estimated (middle and bottom rows) intensity functions are shown in Fig. 2. The graphs of the criterion functions are also shown in the top row of the figure. The left-most panels show classic kernel estimates using bandwidths chosen according to (6) for the middle row and according to (4) for the bottom row. The Campbell–Mecke-based approach captures both the trend and to some extent the increased intensity near the centre of the plot, whereas the leave-one-out cross-validation approach concentrates more on the central feature. This phenomenon is emphasized even more in the adaptive kernel estimates whose bandwidths are selected using (6) for the middle row and by leave-one-out cross-validation for the bottom row. For this pattern, the best estimate in terms of mean squared error is the one shown in the middle right panel.

5 Applications

To illustrate the conclusions of the simulation study of Sect. 4 on real data, we will look at two patterns. The first one shows induced earthquake occurrences whose intensity varies gradually over the field; the second one is a mapped tree pattern with intensity hot spots due to disturbances during harvesting.

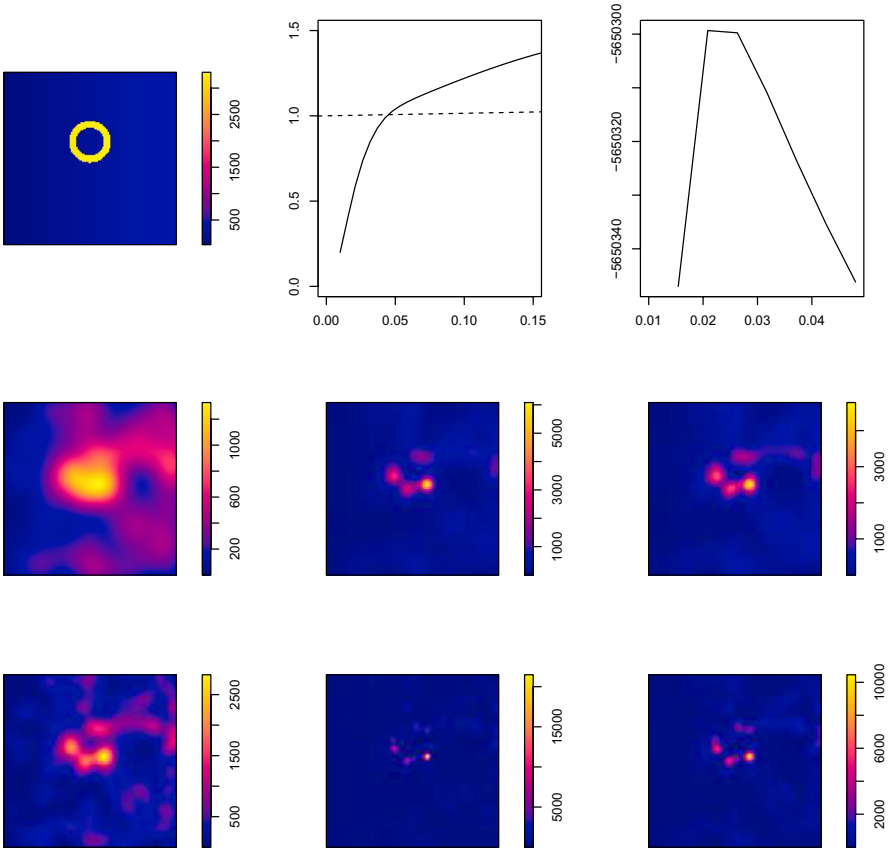

Fig. 2 Intensity function estimates for the point pattern depicted in Fig. 1. Top row: True intensity function (left), graph of $T_\kappa(h)$ (middle) and graph of $L_\kappa(h)$ (right) as functions of h. Middle row: Global (left) and adaptive kernel estimates using a Gaussian kernel with local edge correction and bandwidth chosen by (6) (middle and right). Middle row: Global (left) and adaptive kernel estimates using a Gaussian kernel with local edge correction and bandwidth chosen by (4) (middle and right). The pilot estimates are classic kernel estimates using a Gaussian kernel with local edge correction and bandwidth chosen by (4) (middle), respectively (6) (right)

5.1 Pattern of induced earthquakes

In 1959, a large gas field of about 969 square km was discovered in Groningen, a province in the north of The Netherlands. Initially, the benefits from the sale of gas were a boon to the Dutch economy. However, from the 1990 s earthquakes were being registered in the previously tectonically inactive Groningen region. A catalogue is maintained by the Royal Dutch Meteorological Office.[1] The pattern of 332 induced earthquakes of magnitude 1.5 and larger during the period 1995–2021 is depicted in the left-most panel in Fig. 3. Note that most earthquakes

[1] www.knmi.nl/kennis-en-datacentrum/dataset/aardbevingscatalogus.

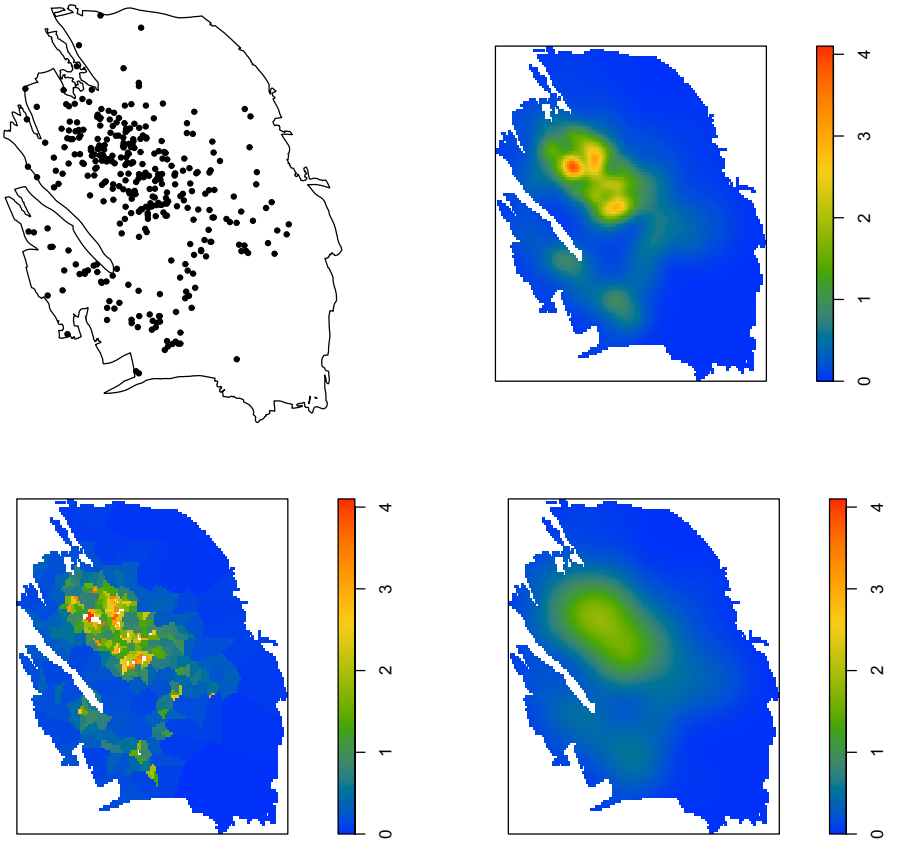

Fig. 3 Map of earthquakes of magnitude $M \geq 1.5$ that occurred during the time period 1995–2021 in the Groningen gas field (top left panel). Top right panel: Adaptive kernel estimates of the intensity function using a Gaussian kernel with local edge correction and bandwidth as well as pilot bandwidth selected by leave-one-out cross-validation. Bottom row: A classic kernel estimate with bandwidth chosen according to the Scott rule (right panel) and Voronoi-based estimate of the intensity function (left panel)

occurred in the central and western regions but there do not appear to be any clusters.

We therefore estimate the intensity function by adaptive kernel estimation and select the bandwidths using the leave-one-out cross-validation principle (4). We apply a Gaussian kernel and local edge correction. The result is shown in the top right panel in Fig. 3. For comparison, the bottom row shows a Voronoi tessellation based estimate (Barr and Schoenberg, 2010) and a classic kernel estimate obtained upon applying the Scott rule (Scott, 1992) from classic multivariate density estimation.

In the absence of a ground truth, we can only judge the performance of the various estimators by eye. The adaptive kernel estimator seems to outperform its competitors. It captures the central high-risk region and its internal structure. Also, the U-bend to the west is visible. The classic kernel estimate does capture the central

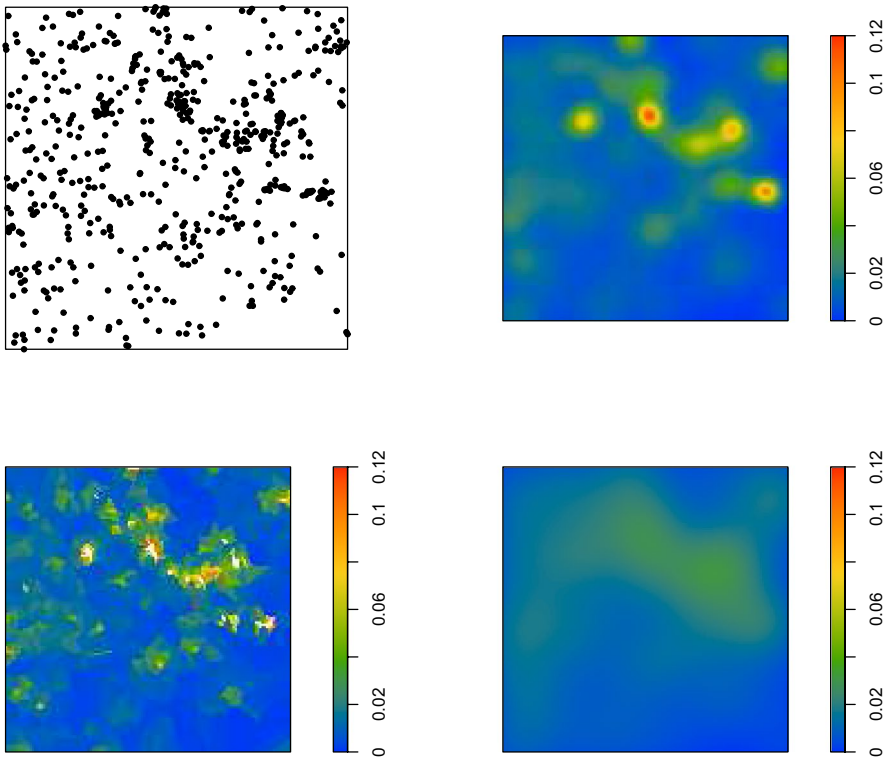

Fig. 4 Locations of pine trees (top left panel) in a square plot in Thomas County, Georgia, censused in 1979. Top right panel: Adaptive kernel estimates of the intensity function using a Gaussian kernel with local edge correction and bandwidth selected by (6). The pilot bandwidth is chosen according to (4). Bottom row: A classic kernel estimate with bandwidth chosen according to the and Scott rule (right panel) and Voronoi-based estimate of the intensity function (left panel)

region but misses all internal structure. The Voronoi estimate on the other hand seems to capture spurious detail and, being discontinuous, is less pleasing to the eye. To alleviate this undesirable behaviour to some extent, one could apply repeated thinning and smoothing, as suggested by Moradi et al. (2019).

5.2 Pattern of pine trees

The map in the left-most panel of Fig. 4 depicts the locations of 584 Longeaf pine trees in a 200 by 200 metre square region in Thomas County, Georgia (USA) (Platt et al., 1988; Rathbun and Cressie, 1994), the coordinates of which are available in the R-package `spatstat` (Baddeley et al., 2015). The pattern displays some marked clusters of elevated intensity, probably due to disturbance of the ground cover during the salvage of dead timber.

Thus, we estimate the intensity function by adaptive kernel estimation and select the bandwidth using (6) with a pilot bandwidth chosen according to (4). We apply a Gaussian kernel and local edge correction. The result is shown in the top right panel in Fig. 4. For comparison, the bottom row shows a Voronoi tessellation-based estimate (Barr and Schoenberg, 2010) and a classic kernel estimate obtained upon applying the Scott rule (Scott, 1992).

One sees that the adaptive Cronie–Van Lieshout algorithm succeeds in finding some hot spots, which are completely missed by the classic kernel estimate. The Voronoi estimate finds more hot spots than the adaptive kernel estimate, but these could simply be clusters of trees.

6 Discussion

In this article, we compared two algorithms for adaptive bandwidth selection for kernel estimators of the spatial intensity function and proved their validity. Simulations showed that for patterns with strong contrasts in point densities, the adaptive kernel estimator outperforms the classic kernel estimator in terms of integrated squared error. We also demonstrated the feasibility of the proposed algorithms in practice. The algorithm based on the Poisson leave-one-out cross-validation likelihood works well for Poisson processes and for models for repulsion. The Campbell–Mecke-based algorithm is more suited to models for clustering. Note that this algorithm applies for a wider range of kernels than the cross-validation approach that requires the argument of the log to be nonzero.

Let us briefly consider the numerical complexity of the two approaches. For a pattern with n points, the calculation of $\widehat{\lambda}_A$ without edge correction requires n function evaluations per point. Therefore, calculation of the criterion function F_κ is quadratic in n. Discretizing the range of bandwidth values into n_h steps, the total computational load is therefore of the order $n_h n^2$. Evaluation of the function L_κ is rather more costly as, additionally, for every potential bandwidth also the integral of the estimated intensity function over W must be approximated. If a grid of size M is used to do so, the numerical complexity is $n_h Mn$ which makes the total computational load $n_h n(n + M)$. The computational cost of calculating the edge correction weights depends on the weights. When direct computation is too costly, fast approximation techniques exist (Davies and Baddeley, 2018).

Finally, in this article, we only considered isotropic kernels. It would be interesting to study adaptive kernel estimators with different bandwidths for the various components.

Acknowledgements Many thanks are due to the anonymous referees for thoughtful suggestions that helped to improve the manuscript.

References

Abramson, I. A. (1982). On bandwidth variation in kernel estimates—A square root law. *The Annals of Statistics, 10*, 1217–1223.

Baddeley, A., Rubak, E., Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. Boca Raton: CRC Press.

Barr, C. D., Schoenberg, F. P. (2010). On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. *Biometrika, 97*, 977–984.

Berman, M., Diggle, P. J. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 51*, 81–92.

Bowman, A. W., Foster, P. J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *Journal of the American Statistical Association, 88*, 529–537.

Brooks, M. M., Marron, J. S. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications, 38*, 157–165.

Chacón, J. E., Duong, T. (2018). *Kernel smoothing and its applications*. Boca Raton: CRC Press.

Chiu, S. N., Stoyan, D., Kendall, W. S., Mecke, J. (2013). *Stochastic geometry and its applications*, 3rd ed., Chichester: Wiley.

Cronie, O., van Lieshout, M. N. M. (2018). A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika, 105*, 455–462.

Davies, T. M., Baddeley, A. (2018). Fast computation of spatially adaptive kernel estimates. *Statistics and Computing, 28*, 937–956.

Davies, T. M., Flynn, C. R., Hazelton, M. L. (2018). On the utility of asymptotic bandwidth selectors for spatially adaptive kernel density estimation. *Statistics and Probability Letters, 138*, 75–81.

Diggle, P. J. (1985). A kernel method for smoothing point process data. *Applied Statistics, 34*, 138–147.

Diggle, P. J. (2014). *Statistical analysis of spatial and spatio-temporal point patterns*, 3rd ed., Boca Raton: CRC Press.

Du Rietz, G. E. (1929). The fundamental units of vegetation. *Proceedings of the International Congress of Plant Science, 1*, 623–627.

Hall, P., Hu, T. C., Marron, J. S. (1995). Improved variable window kernel estimates of probability densities. *The Annals of Statistics, 23*, 1–10.

Hall, P., Minnotte, M. C., Zhang, C. (2004). Bump hunting with non-Gaussian kernels. *The Annals of Statistics, 32*, 2124–2141.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester: Wiley.

van Lieshout, M. N. M. (2012). On estimation of the intensity function of a point process. *Methodology and Computing in Applied Probability, 14*, 567–578.

van Lieshout, M. N. M. (2019). *Theory of spatial statistics: A concise introduction*. Boca Raton: CRC Press.

van Lieshout, M. N. M. (2020). Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. *Methodology and Computing in Applied Probability, 22*, 995–1008.

van Lieshout, M. N. M. (2021). Infill asymptotics for adaptive kernel estimators of spatial intensity functions. *Australian and New Zealand Journal of Statistics, 63*, 159–181.

Lo, P. H. (2017). An iterative plug-in algorithm for optimal bandwidth selection in kernel intensity estimation for spatial data, PhD Thesis, Technical University of Kaiserslautern.

Loader, C. (1999). *Local regression and likelihood*. New York: Springer.

Matérn, B. (1986). *Spatial variation*. Berlin: Springer.

Moradi, M. M., Cronie, O., Rubak, E., Lachièze-Rey, R., Mateu, J., Baddeley, A. (2019). Resample-smoothing of Voronoi intensity estimators. *Statistics and Computing, 29*, 995–1010.

Ord, J. K. (1978). How many trees in a forest? *Mathematical Sciences, 3*, 23–33.

Platt, W. J., Evans, G. W., Rathbun, S. L. (1988). The population dynamics of a long-lived Conifer (Pinus palustris). *The American Naturalist, 131*, 491–525.

Rathbun, S. L., Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in southern Georgia. *Journal of the American Statistical Association, 89*, 1164–1173.

Schaap, W. E., van de Weygaert, R. (2000). Letter to the editor. Continuous fields and discrete samples: Reconstruction through Delaunay tessellations. *Astronomy and Astrophysics, 363*, L29–L32.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice and visualization*. New York: Wiley.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Stoyan, D., Grabarnik, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachrichten, 151*, 95–100.

Wand, M. P., Jones, M. C. (1994). *Kernel smoothing*. Boca Raton: Chapman & Hall.