Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

On Kernel-Based Intensity Estimation of Spatial Point Patterns on Linear Networks

M. Mehdi Moradi, Francisco J. Rodríguez-Cortés & Jorge Mateu

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2018, VOL. 27, NO. 2, 302–311
https://doi.org/10.1080/10618600.2017.1360782

Taylor & Francis
Taylor & Francis Group

On Kernel-Based Intensity Estimation of Spatial Point Patterns on Linear Networks

M. Mehdi Moradi [a], Francisco J. Rodríguez-Cortés [b], and Jorge Mateu[b]

[a]Institute of New Imaging Technologies (INIT), University Jaume I, Castellon, Spain; [b]Department of Mathematics, University Jaume I, Castellon, Spain

ABSTRACT
We propose an extension of Diggle's nonparametric edge-corrected kernel-based intensity estimator to the case of events coming from an inhomogenous point pattern on a linear network. We analyze its statistical properties, showing that it is an unbiased estimator of the first-order intensity; we also provide an expression for the variance, and comment on the appropriate bandwidth selection. Our estimator is compared with the current existing equal-split discontinuous kernel density estimator in terms of the mean integrated squared error (MISE). We then use our estimator on two real datasets. We first revisit street crimes in an area of Chicago, obtaining similar results to previously published ones based on a parametric intensity function. Then, we study network-based spatial events consisting of calls to the Police department reporting anti-social behavior in the city of Castellon (Spain).

ARTICLE HISTORY
Received May 2016
Revised April 2017

KEYWORDS
Anti-social behavior; Intensity function; Kernel estimator; Shortest-path distance; Street crimes

1. Introduction

The last decade witnessed an extraordinary increase in interest in the analysis of network-related data within numerous disciplines. This pervasive interest is partly caused by a strongly expanded availability of network data. In the spatial statistics field, there are numerous real problems such as the location of traffic accidents in a geographical area, geo-coded locations of crimes or anti-social behavior events in the streets of cities that need to restrict the support of the process over such linear networks to set and define a more realistic scenario. In this context, the spatial events can be classified into two classes with respect to the linear networks: events that occur directly on a linear network, and events that occur alongside a linear network rather than directly on it (see more specific details in Okabe and Sugihara 2012, chap. 1). We consider here the events directly occurring on a linear network as a spatial point pattern.

The inherent geometry of a linear network makes the analysis of spatial point patterns a sort of different story with respect to the analysis of such patterns occurring on planar regions. For example, the shortest-path distance can be used as an alternative to the Euclidean one, and since this distance strongly depends on the structure of the linear network, it adds a point of difficulty to the computation. Hence, having efficient approaches to analyze spatial point patterns on linear networks is a welcome issue. We note that a particular realization of events may be clustered on a planar region while they may be just randomly distributed on a linear network.

Okabe and Yamada (2001) replaced the Euclidean distance by the shortest-path distance, and modified the empirical Ripley K-function adapting it to the case of a linear network. Borruso (2005) implemented a method to examine clusters of human-related events on a linear network. Ver Hoef, Peterson, and Theobald (2006) developed spatial autocovariance models for stream networks that incorporate flow and use stream distances. Okabe and Satoh (2006) provided a simple procedure to convert a nonuniform network to a uniform one so that the graph of the new network is not a straight-line graph, that is, its edges might be connected by polygonal line segments or curved segments. Xie and Yan (2008) proposed a nonparametric kernel-based estimator to analyze traffic accidents. However, this estimator did not consider the structure of the linear network around the points, and as a result they provided a highly biased estimation. Similarly to what happens in the plane, Xie and Yan (2008) also found that the type of kernel function is not critical in analyzing the intensity and the spatial structure of the pattern on the linear network, whereas the linear pixel length critically impacts the local variation details and the chosen bandwidth has a direct impact over the smoothness. Okabe, Satoh, and Sugihara (2009) introduced equal-split network kernel density estimators, both under continuous and discontinuous schemes. The equal-split estimators are perfectly well-defined functions of location along segment. However, these estimators are based on nodes, and if the path between any two arbitrary points does not contain any node, the corresponding mass is analogous to the mass coming from the uncorrected intensity estimator, and this can affect the estimation of the intensity.

Using a constructive process based on stream distances and moving average functions, Ver Hoef and Peterson (2010) developed spatial autocovariance models for spatially continuous data on stream networks. Okabe and Sugihara (2012) provided a good account of interesting discussion on the spatial analysis along linear networks. The K-function adapted to the linear network case by Okabe and Yamada (2001) was affected by the geometry of the linear network and its interpretation was not easy. Therefore, Ang, Baddeley, and Nair (2012) corrected it by defining a geometrically corrected K-function, which does not

CONTACT M. Mehdi Moradi Moradi@uji.es Institute of New Imaging Technologies (INIT), University Jaume I, Castellon, Spain.

depend on the geometry of the network, and showed that for any Poisson process $K(r) = r$. Similar to the point processes on the plane, this property can be used in exploratory data analysis, hypothesis testing, and goodness-of-fit tests. Baddeley, Jammalamadaka, and Nair (2014) considered multitype point patterns and discussed first- and second-order characterizations of such patterns on a linear network. O'Donnell et al. (2014) analyzed data that are collected over river networks and developed methods of flexible regression for such data, discussed local fitting, penalized methods and spatio-temporal models. To fix the effects of estimation, prediction, and the estimation of covariance parameters, Som et al. (2014) considered optimal sampling designs for stream networks. McSwiggan, Baddeley, and Nair (2017) also developed a kernel density estimation based on the properties of diffusion on the network.

Knowing that Diggle's corrected estimator for the intensity function on the plane has been shown to provide an overall better performance than other existing nonparametric kernel-based estimators (in the sense of a smaller mean square error), and that it is normalized so that the integral of the estimated intensity over the window is exactly equal to the observed number of points (Baddeley, Rubak, and Turner 2015, p. 169), the aim of this article is to propose a new intensity estimator, adapting Diggle's corrected estimator to the case of linear networks. Its statistical properties such as unbiasedness and continuity are also discussed. Moreover, we also discuss on bandwidth selection tools. Under inhomogenous Poisson processes, and considering two different linear networks, we compare the adapted Diggle's corrected estimator with the equal-split discontinuous estimator in terms of the mean integrated squared error of the intensity (MISE). We show that our estimator provides a better performance in the estimation of the intensity function than the equal-split discontinuous one. We finally apply our proposed intensity estimator over street crimes in an area of Chicago (USA), and over network-based spatial events consisting of calls to the Police department reporting anti-social behavior in the streets of Castellon (Spain).

The rest of the article is organized as follows. Section 2 reviews some basic concepts of point processes on linear networks. Details of the equal-split discontinuous and adapted Diggle's corrected estimator for point patterns on linear networks, together with a bandwidth selection method are discussed in Section 3. Section 4 is devoted to a simulation study, comparing the performance of our approach with the equal-split discontinuous method. The real data applications come in Section 5, and the article finally ends with a discussion section.

2. Background and Setup

The study of point patterns on linear networks cannot be approached by just using classical statistical techniques designed for point patterns in a two-dimensional space. In particular, the intensity of events lying on a network may be quite different compared to the two-dimensional spatial case. This section considers some formalism having to do with linear networks and how point processes can be defined in this context.

2.1. Point Processes on Linear Networks

A line segment in the plane with endpoints $\mathbf{u}$ and $\mathbf{v}$ can be written in a parametric form as $[\mathbf{u}, \mathbf{v}] = \{t\mathbf{u} + (1 - t)\mathbf{v} : 0 \le t \le 1\}$ with $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$. Following Ang, Baddeley, and Nair (2012), a linear network L can be defined as the union of a finite collection of line segments embedded in a plane. According to Okabe, Satoh, and Sugihara (2009), the endpoints of segments are called nodes and the degree of a node $\mathbf{n}$ (denoted by δ) is the number of line segments that share the same node. The total length of all line segments in L is denoted by $|L|$. The distance between two points $\mathbf{u}$ and $\mathbf{v}$ in the network L is computed by the shortest-path distance $d_L(\mathbf{u}, \mathbf{v})$, which is the minimum of the length of all possible paths between $\mathbf{u}$ and $\mathbf{v}$. The disk of radius $r > 0$ centered at the point $\mathbf{u} \in L$ is given by $b_L(\mathbf{u}, r) = \{\mathbf{v} \in L : d_L(\mathbf{u}, \mathbf{v}) \le r\}$ and its relative boundary $\partial b_L(\mathbf{u}, r)$ is the set of points lying exactly r units away from $\mathbf{u}$. Moreover, the circumference $m(\mathbf{u}, r)$ is the number of points in the relative boundary with center at $\mathbf{u}$ and radius r, which are on the linear network L, that is, $m(\mathbf{u}, r) = \#\{\partial b_L(\mathbf{u}, r) \cap L\}$. The circumference is finite for $r < \infty$ and $m(\mathbf{u}, \infty) = \infty$ by convention. In addition, $R(L)$ is the largest value so that $m(\mathbf{u}, r) \neq 0$ for all events $\mathbf{u} \in X$ and all $r \le R(L)$ (for more details see Baddeley, Rubak, and Turner 2015, chap. 17).

We consider a point process X on a linear network L with no overlapping points as a random countable subset of $\mathbb{R}^2$. In practice, we observe $n > 0$ events $\{\mathbf{u}_i\}_{i=1}^n$ of X within a linear network $L \subset \mathbb{R}^2$, and let $N(L)$ denote the number of events happened in L. Assume that X has a kth-order product density $\lambda^{(k)}$, for $k = 1, 2, \ldots$. Then for any nonnegative Borel function h defined on $(\mathbb{R}^2)^{\otimes k}$, Campbell's formula for linear networks is given by

$$\int_{L^{\otimes k}} h(u_1, \ldots, u_k)\lambda^{(k)}(u_1, \ldots, u_k)d_1(u_1, \ldots, u_k) = \mathbb{E}\left[\sum_{\mathbf{u}_1,\ldots,\mathbf{u}_k \in X}^{\neq} h(\mathbf{u}_1, \ldots, \mathbf{u}_k)\right], \tag{1}$$

where $\int_{L^{\otimes k}} = \int_L \ldots \int_L$ for k times, $\sum^{\neq}$ means that the sum is over k pairwise distinct points, and $d_1 u$ denotes one-dimensional integration over the line segment (Federer 1996; Ang, Baddeley, and Nair 2012; Chiu et al. 2013, p. 114). Considering (1), in particular for $k = 1$, the kth-order product density function is called the intensity function $\lambda(\cdot)$, which is the point process analog to the mean function for a real-valued stochastic process, and for an arbitrary subset $A \subseteq L$ we may write

$$\mu(A) = \mathbb{E}[N(A)] = \int_A \lambda(u)d_1 u, \tag{2}$$

which is the expected number of points in $A \subset L$.

A point process X on a linear network L with constant intensity $\lambda(\mathbf{u}) = \lambda$ for all $\mathbf{u} \in L$ is called homogenous. Generally, $\lambda(\mathbf{u})$ is interpreted as an expected number of events per unit length of the linear network L in a neighborhood of $\mathbf{u}$.

Similar to spatial point processes on the plane, the Poisson process is an important model to study point patterns on linear networks due to its simplicity, which leads to the known CSR model. Homogenous Poisson point processes on a linear network L with intensity λ can be characterized by three fundamental properties: (i) the random variable $N(L)$ follows a Poisson distribution with mean $\lambda|L|$, (ii) given $N(L) = n$, the n events in L constitute an independent random sample from the uniform distribution on L, and finally (iii) for any finite number of disjoint subsets $l_i \subset L$, the random variables $N(l_i)$ are independent. Under the assumption of homogenous Poisson processes, the intensity function is constant and can be estimated by the quantity $\hat{\lambda} = n/|L|$.

A more general scenario to generate inhomogenous random patterns on a linear network comes from inhomogenous Poisson processes, which can be obtained by replacing the constant intensity λ of the Poisson process by a spatially varying intensity function over the linear network L. In this case, the random variable $N(L)$ follows a Poisson distribution with mean $\int_L \lambda(u) d_1 u$, which is a straightforward extension of the homogenous case. Note that, for any $A \subset L$ with $0 < \mu(A) < \infty$, conditional on $N(A) = m$, $m \in \mathbb{N}$, X_A is a Binomial point process on A with density $\lambda/\mu(A)$.

3. Kernel-Based Intensity Estimation

To estimate the density of points along the network, to detect the significant high-density segments on the linear network, and also to improve the previous published kernel smoothing techniques, Okabe, Satoh, and Sugihara (2009) proposed a method to estimate the density function and called their estimator "*network kernel density estimator*," which is analogous to the uncorrected intensity estimator in the planar case. This method is basically an average of a weight assigned by a one-dimensional kernel corrected by a quantity that depends on the degree of nodes within a specific sub-network for each point on the linear network L. The method has been accommodated in the R package spatstat by the function `density.lpp` (for more details, see Baddeley, Rubak, and Turner 2015, p. 721).

3.1. *Equal-Split Discontinuous Estimator*

Okabe, Satoh, and Sugihara (2009) considered a sub-network $L_{\mathbf{u}}$ of L where the shortest-path distance between $\mathbf{u}$ and any point on $L_{\mathbf{u}}$ is less or equal to ϵ, and they called it "*buffer network*" of $\mathbf{u}$ with width ϵ (bandwidth). This is equivalent to the concept of a disk centered at $\mathbf{u}$ and radius ϵ, that is, $b_L(\mathbf{u}, \epsilon)$ as defined in Ang, Baddeley, and Nair (2012). Okabe, Satoh, and Sugihara (2009) proposed the network kernel density estimator (network KDE), an intensity estimator for a linear network, defined as

$$\hat{\lambda}_{\epsilon}^{(\mathrm{OS})}(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{K}_{\mathbf{v}_i}(\mathbf{u}), \quad \mathbf{u} \in L, \tag{3}$$

where $\mathrm{K}_{\mathbf{v}}(\mathbf{u})$ is the equal-split discontinuous/continuous kernel function, according to which they are defining a discontinuous/continuous function on a network L. It is important to point out that here continuity at nodes is of interest.

We here only focus on the equal-split discontinuous estimator for the reasons given later in Section 4. To analyze and compare the methodology behind the equal-split discontinuous estimation with respect to our proposed estimator on linear networks (see Section 3.2), we first highlight some concepts related to the decomposition of a linear network L in Okabe and Sugihara (2012).

Let $\mathbf{N}_{(1)}$ be the set of nodes with degree 1, and $\mathbf{N}_{(\geq 3)}$ be the set of nodes with degree larger than or equal to 3. Let $L_{\mathbf{n}_{(1)q}}$ be the disk (buffer networks) of the qth node in $\mathbf{N}_{(1)}$ with width 2ϵ, and let $L_{\mathbf{n}_{(\geq 3)q}}$ be the disk (buffer networks) of the qth node in $\mathbf{N}_{(\geq 3)}$ with width 2ϵ. Then

$$L_{(1)} = \{L_{\mathbf{n}_{(1)q}} | \mathbf{n}_{(1)q} \in \mathbf{N}_{(1)}\} \quad \text{and}$$
$$L_{(\geq 3)} = \{L_{\mathbf{n}_{(\geq 3)q}} | \mathbf{n}_{(\geq 3)q} \in \mathbf{N}_{(\geq 3)}\}. \tag{4}$$

Geometrically, it is more natural to consider a linear network where each line segment is larger or equal to 4ϵ (similar results can be obtained without this assumption) as it is proposed by Okabe and Sugihara (2012). The complement set of disks (buffer networks) with respect to L is given by

$$L_{(2)} = \{L_{\mathbf{n}_{(2)q}} | \mathbf{n}_{(2)q} \in (\mathbf{N}_{(1)} \cup \mathbf{N}_{(\geq 3)})^c\}, \tag{5}$$

where $L_{(2)}$ is the set of disks (buffer networks) with exactly two endpoints. Network KDE methods vary form $L_{(1)}$ to $L_{(2)}$ and $L_{(\geq 3)}$. Okabe and Sugihara (2012) claimed that the network KDE on $L_{(1)}$ and $L_{(2)}$ is equivalent to the case of one-dimensional probability density estimation (see Silverman 1986). For $L_{(\geq 3)}$, the equal-split discontinuous kernel density function must be calculated for two different cases as follows:

Case 1: If the point $\mathbf{u} \in L$ does not coincide with a node, the general idea to count is to traverse all the paths over the disk $b_L(\mathbf{u}, \epsilon)$ from $\mathbf{u}$ to the points in $\partial b_L(\mathbf{u}, \epsilon)$ visiting the nodes located on the shortest-path between $\mathbf{u}$ and $\mathbf{v}$, and equally divide the value of $\kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}))$ by the degree of the respective node minus one for each node, and repeating the same procedure for each path on the disk. The final expression for an equal-split discontinuous kernel density function is given by

$$\mathrm{K}_{\mathbf{v}}(\mathbf{u}) = \begin{cases} \dfrac{\kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}))}{(\delta_{i1}-1)(\delta_{i2}-1)\cdots(\delta_{iq}-1)}, & \text{for } d_L(\mathbf{u}, \mathbf{n}_{iq-1}) \geq d_L(\mathbf{u}, \mathbf{v}) \geq d_L(\mathbf{u}, \mathbf{n}_{iq}), \\ 0, & \text{for } d_L(\mathbf{u}, \mathbf{v}) \geq \epsilon, \end{cases} \tag{6}$$

for $q = 1, 2, \ldots, w$, where q is the index for the order of the nodes located on the shortest-path between $\mathbf{u}$ and $\mathbf{v}$, $i = 1, \ldots, m(\mathbf{u}, \epsilon)$, i is the index of paths in $b_L(\mathbf{u}, \epsilon)$, and κ is a one-dimensional kernel function with bandwidth or smoothing parameter ϵ. Note that, if the path between $\mathbf{u}$ and $\mathbf{v}$ does not contain any node, then $\mathrm{K}_{\mathbf{v}}(\mathbf{u}) = \kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}))$, that is, $\mathrm{K}_{\mathbf{v}}(\mathbf{u})$ turns into the uncorrected kernel estimator of the intensity function (see Baddeley, Rubak, and Turner 2015, p. 168). An example that satisfies buffers with no node can be the real line.

Case 2: If the point $\mathbf{u} \in L$ coincides with a node, the expression for an equal-split discontinuous kernel density function is given by

$$\mathrm{K}_{\mathbf{v}}(\mathbf{u}) = \begin{cases} \dfrac{2\kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}))}{\delta_{i1}(\delta_{i2}-1)\cdots(\delta_{iq}-1)}, & \text{for } d_L(\mathbf{u}, \mathbf{n}_{iq-1}) \geq d_L(\mathbf{u}, \mathbf{v}) \geq d_L(\mathbf{u}, \mathbf{n}_{iq}), \\ 0, & \text{for } d_L(\mathbf{u}, \mathbf{v}) \geq \epsilon, \end{cases} \tag{7}$$

for $q = 2, \ldots, w$ and $i = 1, \ldots, m(\mathbf{u}, \epsilon)$. Okabe and Sugihara (2012) showed that the equal-split functions satisfy the definition of a kernel function, and provided a proof for their unbiasedness property when the density comes from a uniform probability density function (i.e., under complete spatial randomness). In case that the observed data are not uniform, these authors suggested to use the probability integral transformation as in Okabe and Satoh (2006). With this transformation, we convert a nonuniform network into a uniform one, and the estimators are unbiased in this transformed network. However, this artifact modifies the structure of network making it not very realistic (see more details in Okabe and Satoh 2006; Okabe, Satoh, and Sugihara 2009).

In brief, the equal-split discontinuous kernel method divides the mass at nodes equally into outgoing segments and it results in a discontinuous function on the network. Okabe and Sugihara (2012) did not find a systematic method available to find a continuous kernel density function on a network (Okabe and Sugihara 2012, p. 183), thus by only changing the way of splitting compared to the equal-split discontinuous estimator, they defined the equal-split continuous one, which produces an unbiased and a continuous function on the linear network (Okabe and Sugihara 2012, p. 192). Note that to produce a continuous function, Okabe and Sugihara (2012) increased the weight of outgoing segments and decreased the weight of incoming segments in the discontinuous algorithm. Hence, although it results in a continuous, unbiased, and mass preserved function on the network, it seems to be somewhat artificial. More formal details about the equal-split continuous kernel density function can be found in Okabe, Satoh, and Sugihara (2009) and Okabe and Sugihara (2012, sec. 9.2.3, p. 183).

In the next section, we propose an alternative kernel-based intensity estimator that produces a continuous, unbiased, and mass preserved function.

3.2. Adapted Diggle's Corrected Estimator

For a point process, the intensity function has a local characterization in terms of an infinitesimal segment $d_1 u$ with total length of line segment $|d_1 u|$ centered at an arbitrary location $\mathbf{u} \in L$, and $\lambda(u) d_1 u$ gives the probability that there is a point of the point process X in $d_1 u$ (see Chiu et al. 2013, p. 113). In addition, the standard estimator of the intensity function for a linear network can be written as

$$\hat{\lambda}_\epsilon(\mathbf{u}) = \frac{N(b_L(\mathbf{u}, \epsilon))}{|b_L(\mathbf{u}, \epsilon)|}, \quad \mathbf{u} \in L. \tag{8}$$

We extend the nonparametric kernel-based edge-corrected intensity estimator (Diggle 1985) to the linear network case following the mathematical arguments in Ang, Baddeley, and Nair (2012). The natural extension can thus be written as

$$\hat{\lambda}_\epsilon^{(\mathrm{D})}(\mathbf{u}) = \sum_{i=1}^{n} \frac{\kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}_i))}{C_{L,\epsilon}(\mathbf{v}_i)}, \quad \mathbf{u} \in L, \tag{9}$$

where κ_ϵ is a one-dimensional kernel function with bandwidth ϵ, and

$$C_{L,\epsilon}(\mathbf{v}) = \int_L \kappa_\epsilon(d_L(\mathbf{v}, \mathbf{u})) d_1 u, \quad \mathbf{v} \in X \cap L, \tag{10}$$

is an edge-correction factor. Following Federer (1996) and Ang, Baddeley, and Nair (2012), the computation of the edge-correction factor in (10) on a linear network is given by

$$\begin{aligned} C_{L,\epsilon}(\mathbf{v}_i) &= \int_L \kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}_i)) d_1 u \\ &= \int_0^\infty \sum_{\mathbf{u} \in L: d_L(\mathbf{u}, \mathbf{v}_i) = r} \kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}_i)) dr \\ &= \int_0^\infty \sum_{\mathbf{u} \in L: d_L(\mathbf{u}, \mathbf{v}_i) = r} \kappa_\epsilon(r) dr \\ &= \int_0^\infty \kappa_\epsilon(r) m(\mathbf{v}_i, r) dr. \end{aligned} \tag{11}$$

Note that the edge-correction in (11) involves the circumference function $m(\mathbf{v}_i, r)$, which provides local information about the structure of the linear network on each point. Therefore, we take into account the essence in Okabe and Sugihara (2012), which is counting over disks (buffer networks), but in our case the weight assigned by the kernel is equally distributed over the segments that composed the respective disk. It is important to point out that the estimator (9) is independent of the choice of the kernel function, that is, usual kernel functions used to estimate the density function can be accommodated here; we thus use the Gaussian kernel in our numerical computations in the next section. We note that the adaptation of classical techniques for the intensity estimation on linear networks is appropriate, and the estimator inherits the properties of its counterpart for the planar case.

3.2.1. Statistical Properties

The analogous estimator to (9) for the planar case has been widely studied in the literature of spatial point processes. Properties such as unbiasedness, the variance, and the appropriate bandwidth selection are analyzed in Møller and Waagepetersen (2003); Illian et al. (2008); Diggle (2013); Baddeley, Rubak, and Turner (2015). We next turn to some statistical properties of the intensity estimator (9).

Property 1. $\int_L \hat{\lambda}_\epsilon^{(\mathrm{D})}(u) d_1 u$ is an unbiased estimator of $\mu(L) = \int_L \lambda(u) d_1 u$.

Proof. Using Campbell's formula (1) and Fubini's theorem, we have that

$$\begin{aligned} \mathbb{E}\left[\int_L \hat{\lambda}_\epsilon^{(\mathrm{D})}(u) d_1 u\right] &= \mathbb{E}\left[\int_L \sum_{\mathbf{v} \in X \cap L} \frac{\kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}))}{C_{L,\epsilon}(\mathbf{v})} d_1 u\right] \\ &= \int_L \mathbb{E}\left[\sum_{\mathbf{v} \in X \cap L} \frac{\kappa_\epsilon(d_L(\mathbf{u}, \mathbf{v}))}{C_{L,\epsilon}(\mathbf{v})}\right] d_1 u \\ &= \int_L \int_L \frac{\kappa_\epsilon(d_L(u, v))}{C_{L,\epsilon}(v)} \lambda(v) d_1 v\, d_1 u \\ &= \int_L \lambda(v) d_1 v = \mu(L). \end{aligned}$$

□

Property 2. If the intensity function λ is constant, then $\hat{\lambda}_\epsilon^{(\mathrm{D})}$ is unbiased.

Proof. Under the assumption of having a constant intensity function, we have

$$\mathbb{E}[\hat{\lambda}_\epsilon^{(D)}(\mathbf{u})] = \mathbb{E}\left[\sum_{\mathbf{v}\in X\cap L} \frac{\kappa_\epsilon(d_L(\mathbf{u},\mathbf{v}))}{C_{L,\epsilon}(\mathbf{v})}\right] = \int_L \frac{\kappa_\epsilon(d_L(u,v))}{C_{L,\epsilon}(v)}\lambda d_1 u = \lambda.$$

□

Property 3. The intensity estimator (9) produces a continuous function on a network L. In particular, all arbitrary points $\mathbf{u}_i \in L$ that have the same distance $d = d_L(\mathbf{u}_i, \mathbf{v})$ to $\mathbf{v} \in L \cap X$ receive the same mass $\kappa_\epsilon(d)/C_{L,\epsilon}(\mathbf{v})$ regardless of the path between $\mathbf{u}_i$ and $\mathbf{v}$. In other words, when the path from $\mathbf{v}$ to $\mathbf{u}_i$ touches a node, each outgoing segment receives a copy of the kernel tail.

Property 4. Assume that X is a Poisson point process on a linear network L. Then, the variance of $\hat{\lambda}_\epsilon^{(D)}(\mathbf{u})$ is given by

$$V(\hat{\lambda}_\epsilon^{(D)}(\mathbf{u})) = \int_L \left[\frac{\kappa_\epsilon(d_L(\mathbf{u},v))}{C_{L,\epsilon}(v)}\right]^2 \lambda(v) d_1 v$$

and its unbiased estimator is of the form $\widehat{V}(\mathbf{u}) = \sum_{i=1}^n [\kappa_\epsilon(d_L(\mathbf{u},\mathbf{v}_i))/C_{L,\epsilon}(\mathbf{v}_i)]^2$ (Baddeley, Rubak, and Turner 2015, p. 173).

3.2.2. Bandwidth Selection

Kernel-based estimates always need to deal with choosing the smoothing bandwidth ϵ, and the estimator (9) is not an exception. Diggle (1985) assumed that the underlying point process is a stationary Cox point process, and estimated ϵ by means of minimizing the mean squared error. However, Cox point processes are not yet developed on linear networks. Techniques for choosing ϵ are studied by Silverman (1986, sec 3.4, p. 43) and Loader (1999, sec. 5.3 and chap. 10) and they can be adapted for linear networks. Nevertheless, we here follow the idea by Cronie and van Lieshout (2016), and use the smoothing bandwidth parameter ϵ, which minimizes

$$\left|\sum_{i=1}^n \frac{1}{\hat{\lambda}_\epsilon^{(D)}(\mathbf{u}_i)} - |L|\right|. \tag{12}$$

Note that, by using Campbell's formula (1),

$$\mathbb{E}\left[\sum_{i=1}^n \frac{1}{\lambda(\mathbf{u}_i)}\right] = \int_L \frac{1}{\lambda(u)}\lambda(u) d_1 u = |L|.$$

Therefore, an optimum bandwidth ϵ can be chosen in a fully nonparametric way using (12), that is, no particular assumption is considered about the point process. Cronie and van Lieshout (2017) compared their approach with likelihood-based cross-validation (Loader 1999), and with state estimation for isotropic Cox processes (Diggle 1985) and found that for clustered patterns, their method appears to be the best choice. For Poisson processes, at least for moderately sized patterns, the likelihood-based cross-validation method seems to give slightly better results than the method defined by Cronie and van Lieshout (2016) and the picture is a bit more varied for regular patterns (see more details in Cronie and van Lieshout 2016). However, in our simulation study, as we know the true intensity function, we use

$$\text{ISE} = \int_L (\hat{\lambda}_\epsilon^{(D)}(u) - \lambda(u))^2 d_1 u, \tag{13}$$

to find the smoothing bandwidth ϵ, that is, an optimum bandwidth is the one that minimizes the integrated squared error (ISE). Note that in case of real data, we consider Equation (12) to find the optimum bandwidth. We point out that $\sum_{i=1}^n 1/\hat{\lambda}_\epsilon^{(D)}(\mathbf{u}_i)$ does not have a monotonic trend but it converges to the total length of the network because we use edge correction (see Cronie and van Lieshout 2016 for details).

In the following section, we show that our estimator in (9) provides a better performance than the equal-split discontinuous one in terms of visualization along the segments and having a smaller MISE $= \mathbb{E}$[ISE].

4. Simulation Study

We carried out a simulation experiment to compare the accuracy and performance of our proposed edge-corrected intensity estimator $\hat{\lambda}_\epsilon^{(D)}(\mathbf{u})$ with the equal-split discontinuous kernel density estimator $\hat{\lambda}_\epsilon^{(OS)}(\mathbf{u})$ under inhomogenous Poisson processes. Okabe and Sugihara (2012) mentioned that both discontinuous and continuous network kernel density estimators are unbiased, but the equal-split continuous density estimator takes longer time to be computed (see Okabe and Sugihara 2012, sec. 9.3 for more details) due to the difference in the way of splitting. In any case, we focus here on the comparison between our proposed method and the equal-split discontinuous estimator defined by Okabe, Satoh, and Sugihara (2009).

We considered two linear networks L_1 and L_2 (shown in Figure 1). L_1 has 19 vertices, 26 lines, and a total length of 39.38 units, within a window $[0, 5] \times [0, 5]$ units^2. It also contains 12 intersections with degrees varying between 2 and 6. Network L_2 has 40 vertices, 48 lines, and a total length of 45.62 units, within a window $[0, 6] \times [0, 7.1]$ units^2. The function `density.lpp` with argument `continuous=FALSE` on the R package `spatstat(1.48-0)` was used to compute the equal-split discontinuous density estimator. It is worth pointing out that throughout this article, both our estimator and the equal-split discontinuous density estimator are computed by using the Gaussian kernel function (Silverman 1986), and the tail of the kernel for distances larger than 4ϵ was removed. In other words, a tail of the kernel with total mass less than M $= 1 - P(X < 4\epsilon)$ has been deleted, where X follows the same distribution as the kernel function κ (Gaussian kernel function), and ϵ is the smoothing bandwidth parameter. In our simulation study, κ is the Gaussian distribution with mean 0 and standard deviation ϵ.

4.1. Accuracy and Performance

We considered two inhomogenous Poisson processes over the two linear networks L_1 and L_2. Two particular realizations of them are shown in Figure 1. In both cases, the degree of inhomogeneity is given by an exponential function, which makes the distribution of events exponentially changing over both networks.

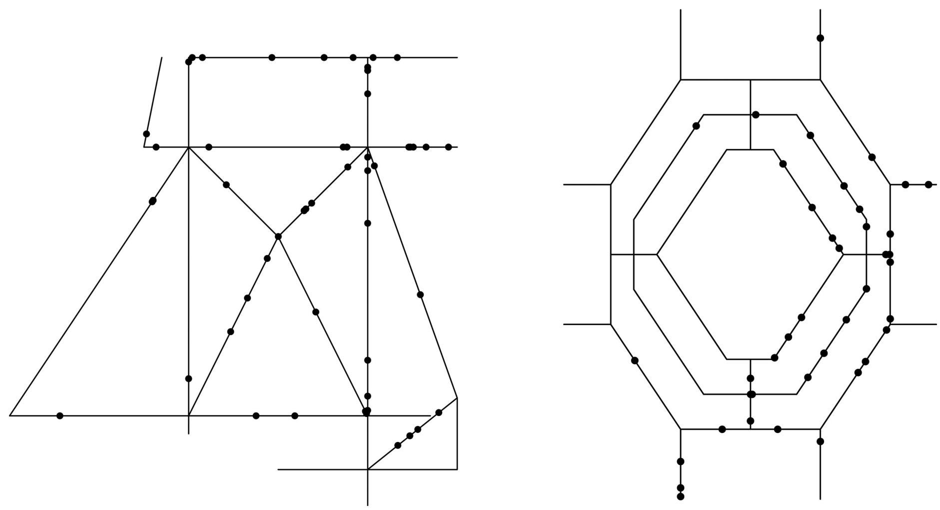


Figure 1. Realization of inhomogenous Poisson processes. Left: with intensity function (14) on network L_1, Right: with intensity function (15) on network L_2.

Example 1. Consider an inhomogenous Poisson process on the linear network L_1 with expected number of points $\mathbb{E}[N(L_1)] \simeq 49$ and intensity function given by

$$\lambda_1(x, y) = 0.2\, e^{0.3(x+y)} \quad \text{with} \quad (x, y) \in L_1. \tag{14}$$

A single realization of this point pattern is shown in the left-hand side of Figure 1. To estimate the intensity function, we first need to find the optimum bandwidth ϵ by using Equation (13). The plot in Figure 2 shows that $\epsilon = 1.2$ units is the optimum value as it minimizes the integrated squared error (ISE).

Using $\epsilon = 1.2$ units, we estimated the intensity of the pattern in the left-hand side of Figure 1 through the adapted Diggle's corrected estimator and the equal-split discontinuous one, together with the true intensity function (see Figure 3).

By comparing both estimated intensities in Figure 3, we note that the equal-split discontinuous density estimator assigns a specific thickness along some line segments, which can be somewhat equivalent to consider a constant estimated intensity value along those line segments of the linear network L_1. However, this is clearly not the case, as the point pattern in Figure 1 is not homogenous and the intensity is changing exponentially over the network. On the other hand, the adapted Diggle's corrected estimator assigns precise values of the intensity to the different parts of each line segment, as shown in the middle plot in Figure 3. It is thus able to identify high/low intensity values along each line segment in the linear network. Moreover, in terms of ISE, the adapted Diggle's corrected estimator shows an ISE = 9.57, outperforming the equal-split discontinuous density estimator with an ISE = 9.66.

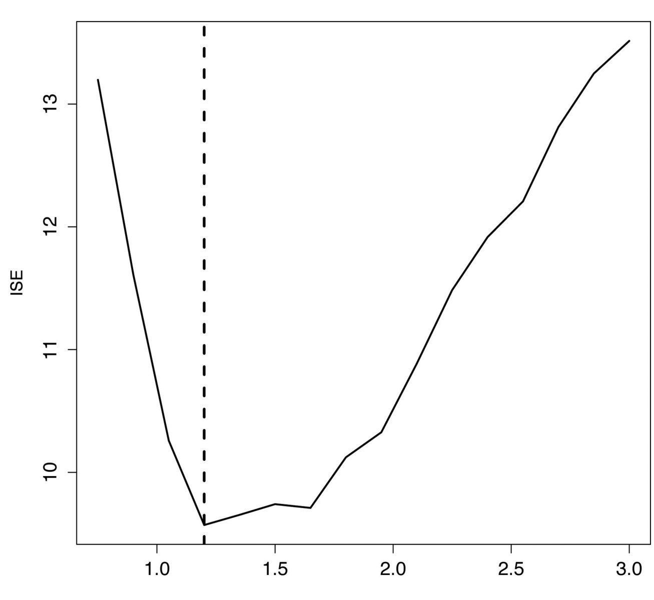


Figure 2. Bandwidth selection, ISE versus a sequence of the bandwidth smoothing parameter ϵ.

To have a better comparison, we then simulated 500 realizations over the linear network L_1 of the inhomogenous Poisson process with intensity function (14). By using $\epsilon = 1.2$ units (see Figure 2), both adapted Diggle's corrected estimator (9) and the equal-split discontinuous (3) were computed for each realization, and we evaluated the MISE $= \mathbb{E}$[ISE]. The adapted Diggle's corrected estimator showed a value of MISE = 10.95, whereas the value for the equal-split discontinuous was MISE = 13.56

Example 2. Consider an inhomogenous Poisson process on the linear network L_2 with expected number of points $\mathbb{E}[N(L_2)] \simeq 42$ and intensity function

$$\lambda_2(x, y) = e^{(x-y)/x} \quad \text{with} \quad (x, y) \in L_2. \tag{15}$$

Similarly to the previous example, we considered the ISE in (13) to find the optimum bandwidth. Figure 4 shows that $\epsilon = 1.95$ units is the optimum bandwidth that minimizes the ISE.

Figure 5 shows the true intensity function (left plot), together with the estimated intensity using the adapted Diggle's corrected estimator (middle plot), and the equal-split discontinuous (right plot). We note that the adapted Diggle's corrected estimator (9) outperforms the equal-split discontinuous estimator (3) along the segments (ISE = 8.54 against ISE = 11.27), while also differentiating much clearly the low/high intensity parts of the segments.

As in Example 1, we simulated 500 realizations of an inhomogenous Poisson process with intensity function (15) on the network L_2 and calculated the MISE $= \mathbb{E}$[ISE] as a means of comparison between the adapted Diggle's corrected estimator and the equal-split discontinuous one. Considering $\epsilon = 1.95$

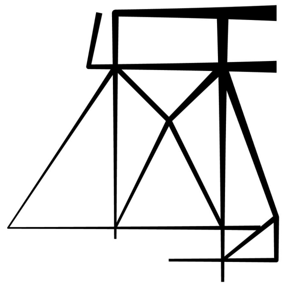


Figure 3. Intensity visualization for the pattern on the left-hand side of Figure 1. Left: true intensity; Middle: adapted Diggle's corrected estimator (9) with ISE = 9.57; Right: equal-split discontinuous with ISE = 9.66.

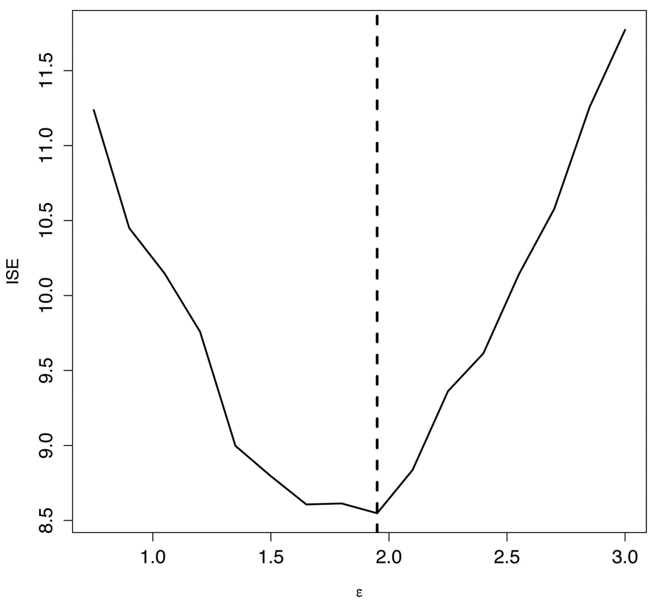


Figure 4. Bandwidth selection, ISE versus a sequence of the bandwidth smoothing parameter ϵ.

units as an optimum bandwidth parameter (see Figure 4), both the adapted Diggle's corrected estimator and the equal-split discontinuous one were computed for each one of the 500 realizations, showing MISE $= 7.81$ and MISE $= 10.00$, for the adapted Diggle's and equal-split discontinuous estimators, respectively.

5. Real Data Applications

This section is devoted to apply the adapted Diggle's corrected estimator to two real datasets in the context of criminology showing events occurring on streets of cities. The analysis of these events on networks provides a more detailed and reliable analysis than if they were analyzed in a planar region. We first consider street crime data in an area of Chicago (USA), a dataset that has previously been studied by Ang, Baddeley, and Nair (2012) using a parametric approach to estimate the intensity function. We then study anti-social behavior data collected by the Police department in the city of Castellon (Spain).

5.1. Chicago Crime Data

The spatial locations of 116 reported street crimes in the period April 25 to May 8, 2002, recorded over a fortnight in an area of Chicago (USA) are shown in Figure 7(a). Although the original data were labeled by the type of crime (assault, burglary, car-theft, damage, robbery, theft, and trespass), following Ang, Baddeley, and Nair (2012) and Baddeley, Rubak, and Turner (2015), we ignore the type of crime.

The street network contains 338 intersections, 503 line segments, and a total length of 31,150 feet (5.9 miles, 9.5 km). For more details about this dataset, see Ang, Baddeley, and Nair (2012) and (Baddeley, Rubak, and Turner 2015, p. 721). Baddeley, Rubak, and Turner (2015) used a nonparametric kernel-based intensity using an equal-split estimator with a small bandwidth parameter of 60 feet. We note that this bandwidth, although allowing to show a nice visualization, might not provide a good intensity estimation. Figure 6 shows that $\epsilon = 650$ feet can minimize (12) while $\epsilon = 60$ feet produces a large bias.

A bandwidth of $\epsilon = 650$ feet is a much more larger value than the one used in Baddeley, Rubak, and Turner (2015, p. 721), but we do have less bias with this amount of bandwidth. Here, we estimated the intensity function through the adapted Diggle's

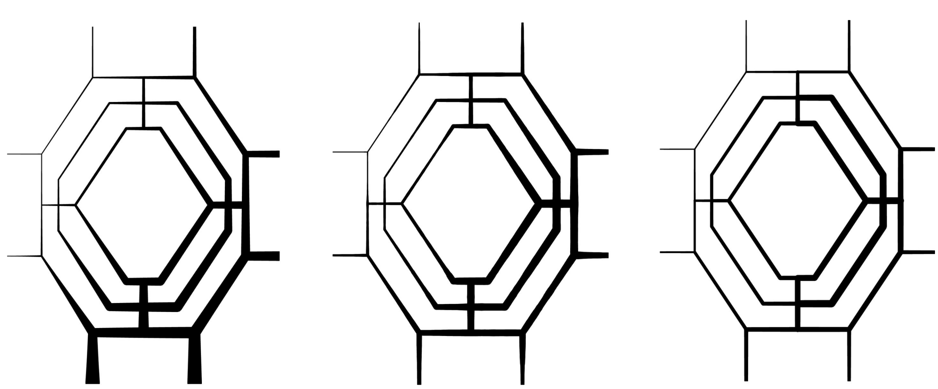


Figure 5. Intensity visualization for the pattern on the right-hand side of Figure 1. Left: true intensity; Middle: adapted Diggle's corrected estimator (9) with ISE = 8.54; Right: equal-split discontinuous with ISE = 11.27.

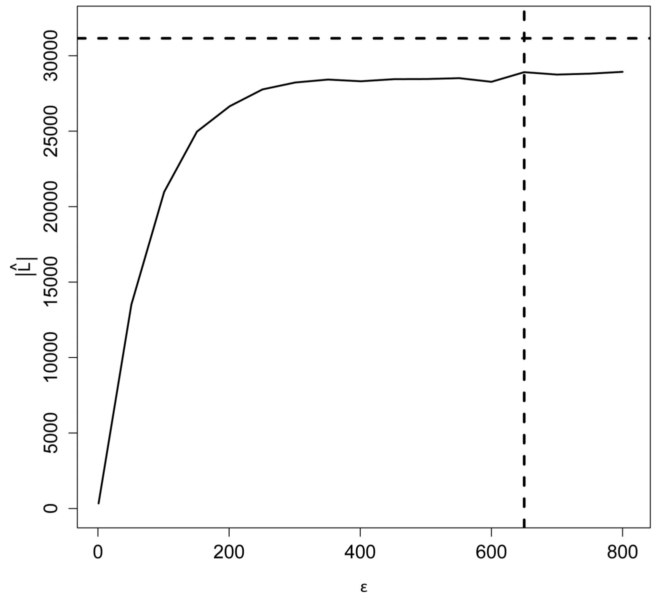


Figure 6. Bandwidth selection for Chicago crime data. $|\hat{L}| = \sum_{i=1}^{n} 1/\hat{\lambda}_{\epsilon}^{(D)}(\mathbf{u}_i)$ against a sequence of the bandwidth smoothing parameter ϵ (with feets as units). The horizontal dashed line shows the total length of the network.

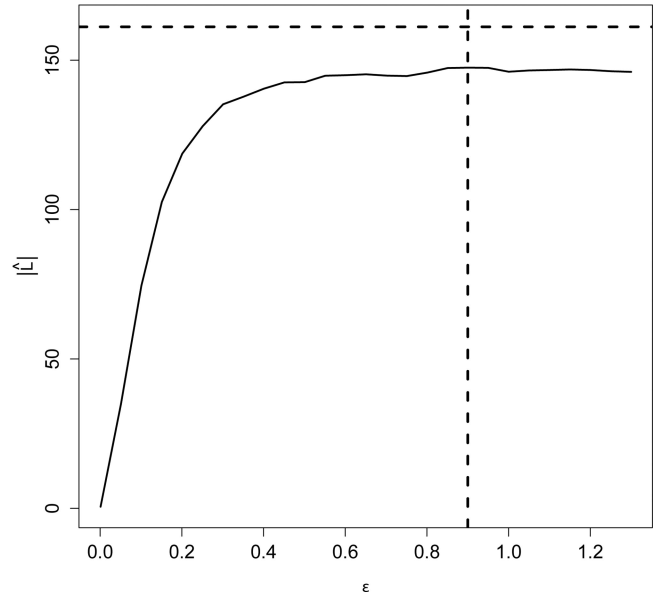


Figure 8. Bandwidth selection for Castellon anti-social behavior data. $|\hat{L}| = \sum_{i=1}^{n} 1/\hat{\lambda}_{\epsilon}^{(D)}(\mathbf{u}_i)$ against a sequence of the bandwidth smoothing parameter ϵ (with kilometers as units). The horizontal dashed line shows the total length of the network.

corrected estimator (see Figure 7) with this optimum bandwidth. The data are markedly inhomogenous, and though we are using a nonparametric estimation only based on the locations of crimes, we note that Figure 7(b) shows the same result as in Ang, Baddeley, and Nair (2012) who used a parametric model.

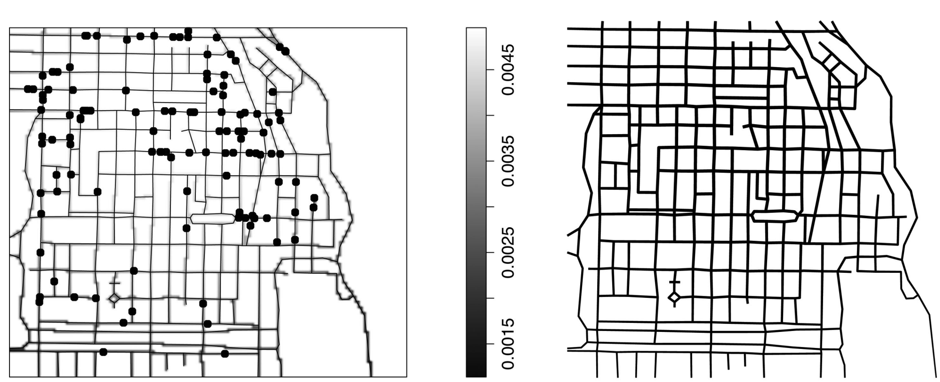
5.2. *Castellon Anti-Social Behavior Data*

The data report geo-referenced coordinates of phone calls received by the Police station in the city of Castellon (Spain) during January 2013. The listed calls were received at the local Police call center or transferred by 112 emergency service to the local Police call center. Geo-codification was performed indirectly by local officials based on precise address information provided by the callers. The calls comprise up to nine different types of crimes or anti-social behavior categories, but we here only focus on anti-social actions comprising a total number of 184 events in the streets of Castellon. The city of Castellon is divided into 108 census tracks with an overall surface of 108.6 km^2. The linear network contains 450 nodes and 2242 line segments, with a total length of 161.17 km. This linear network shows a fairly high degree of complexity given by the old-fashion design of the historical center of the city (see Figure 9).

Figure 8 provides an optimum bandwidth value of $\epsilon = 0.9$ km. The nonparametric kernel-based estimated intensity in Figure 9 provides a clear indication of the degree of inhomogeneity in the pattern. Figure 9(b) shows that most of the anti-social behavior problems occurred in the city center due to existing areas of night life with pubs and discos that generate noise and general disturbances among the population living in these areas. Adding covariate information would enrich the intensity estimation, but in this article we have preferred to focus only on the nonparametric estimation of the intensity.

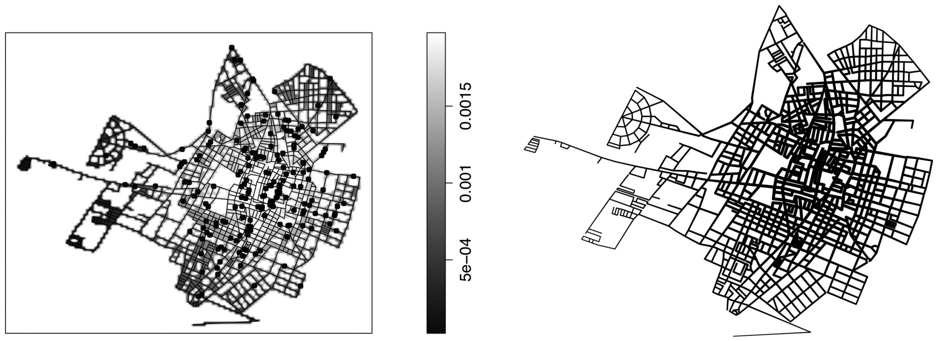
6. Discussion

The analysis of network related data has gained attention in numerous disciplines and, as a result, tremendous developments have taken place while several different network models have



(a) grayscale style (b) line-thickness style

Figure 7. Estimated intensity using adapted Diggle's corrected estimator for Chicago street crime data with a bandwidth parameter of $\epsilon = 650$ feet.



(a) grayscale style (b) line-thickness style

Figure 9. Estimated intensity using the adapted Diggle's corrected estimator for Castellon anti-social behavior data with a bandwidth parameter of $\epsilon = 0.9$ km.

been derived. However, only a very small proportion of articles have dealt with network structures with respect to spatial data. Even less work has been made restricted to the subfamily of planar point processes. In this context, the analysis of the first-order intensity function plays a fundamental role in analyzing the degree of inhomogeneity the data exhibit. In addition, the intensity function is also involved in exploratory analysis, modeling, and residual analysis of spatial point processes on linear networks.

Based on previous works by Okabe and Yamada (2001) and Okabe, Satoh, and Sugihara (2009), and in particular motivated by Okabe and Sugihara (2012), Ang, Baddeley, and Nair (2012), and Baddeley, Rubak, and Turner (2015), we have proposed a new intensity estimator, adapting the original Diggle's corrected estimator to the case of linear networks. In particular, we consider a nonparametric edge-corrected kernel-based estimator that improves the estimation accuracy in comparison with equal-split discontinuous estimator defined by Okabe and Sugihara (2012). The adapted Diggle's corrected estimator is used to highlight the high-intensity segments of an area of Chicago and Castellon street networks in terms of street crimes and anti-social behavior, respectively.

The adapted Diggle's corrected estimator is unbiased when the true intensity is constant, mass preserved, continuous, and easy to be computed, therefore it can be accommodated in other computations of point processes on linear networks such as summary statistics. Also, there is room for improvement in this context such as extensions to multitype and space-time point processes, and defining more complex models such as Cox point processes on linear networks.

We reinforce the fact that there are alternative, perhaps better, bandwidth selection procedures. In particular, we advocate the use of Cox processes on networks to develop an extension of Diggle's procedure.

Acknowledgments

The authors are grateful to O. Cronie and A. Baddeley for fruitful discussions when preparing and revising the manuscript. M. M. Moradi gratefully acknowledges funding from the European union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, *http:// www.geo-c.eu/*). Also, F. J. Rodríguez-Cortés and J. Mateu are supported by grant MTM2016-78917-R from the Spanish Ministry of Science and Education, and grant P1-1B2015-40.

Funding

Spanish Ministry of Science and Education [P1-1B2015-40]; GEO-C program [642332].

ORCID

M. Mehdi Moradi http://orcid.org/0000-0003-3905-4498
Francisco J. Rodríguez-Cortés http://orcid.org/0000-0002-2152-8619

References

Ang, Q. W., Baddeley, A., and Nair, G. (2012), "Geometrically Corrected Second Order Analysis of Events on a Linear Network, with Applications to Ecology and Criminology," *Scandinavian Journal of Statistics*, 39, 591–617. [302,303,304,305,308,310]

Baddeley, A., Jammalamadaka, A., and Nair, G. (2014), "Multitype Point Process Analysis of Spines on the Dendrite Network of a Neuron," *Journal of the Royal Statistical Society*, Series C, 63, 673–694. [303]

Baddeley, A., Rubak, E., and Turner, R. (2015), *Spatial Point Patterns: Methodology and Applications with R*, Boca Raton, FL: CRC Press. [303,304,305,306,308,310]

Borruso, G. (2005), "Network Density Estimation: Analysis of Point Patterns Over a Network," in *Computational Science and Its Applications–ICCSA 2005*, Springer, pp. 126–132. [302]

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013), *Stochastic Geometry and its Applications*, New York: Wiley. [303,305]

Cronie, O., and van Lieshout, M. (2017), "Bandwidth Selection for Kernel Estimators of the Spatial Intensity Function," arXiv preprint arXiv:1611.10221. [306]

Diggle, P. (1985), "A Kernel Method for Smoothing Point Process Data," *Applied Statistics*, 34, 138–147. [305,306]

Diggle, P. J. (2013), *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, Boca Raton, FL: CRC Press. [305]

Federer, H. (1996), *Geometric Measure Theory*, New York: Springer Verlag. [303,305]

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008), *Statistical Analysis and Modelling of Spatial Point Patterns*, New York: Wiley. [305]

Loader, C. (1999), *Local Regression and Likelihood*, New York: Springer. [306]

McSwiggan, G., Baddeley, A., and Nair, G. (2017), "Kernel Density Estimation on a Linear Network," *Scandinavian Journal of Statistics*, 44, 324–345. [303]

Møller, J., and Waagepetersen, R. P. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, Boca Raton, FL: CRC Press. [305]

O'Donnell, D., Rushworth, A., Bowman, A. W., Marian Scott, E., and Hallard, M. (2014), "Flexible Regression Models Over River Networks," *Journal of the Royal Statistical Society*, Series C, 63, 47–63. [303]

Okabe, A., and Satoh, T. (2006), "Uniform Network Transformation for Points Pattern Analysis on a Non-Uniform Network," *Journal of Geographical Systems*, 8, 25–37. [302,305]

Okabe, A., Satoh, T., and Sugihara, K. (2009), "A Kernel Density Estimation Method for Networks, its Computational Method and a GIS-Based Tool," *International Journal of Geographical Information Science*, 23, 7–32. [302,303,304,305,306,310]

Okabe, A., and Sugihara, K. (2012), *Spatial Analysis along Networks: Statistical and Computational Methods*, New York: Wiley. [302,304,305,306,310]

Okabe, A., and Yamada, I. (2001), "The *K*-Function Method on a Network and its Computational Implementation," *Geographical Analysis*, 33, 271–290. [302,310]

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis* (Vol. 26), Boca Raton, FL: Chapman & Hall/CRC press. [304,306]

Som, N. A., Monestiez, P., Hoef, J. M., Zimmerman, D. L., and Peterson, E. E. (2014), "Spatial Sampling on Streams: Principles for Inference on Aquatic Networks," *Environmetrics*, 25, 306–323. [303]

Ver Hoef, J. M., and Peterson, E. E. (2010), "A Moving Average Approach for Spatial Statistical Models of Stream Networks," *Journal of the American Statistical Association*, 105, 6–18. [302]

Ver Hoef, J. M., Peterson, E., and Theobald, D. (2006), "Spatial Statistical Models that use Flow and Stream Distance," *Environmental and Ecological Statistics*, 13, 449–464. [302]

Xie, Z., and Yan, J. (2008), "Kernel Density Estimation of Traffic Accidents in a Network Space," *Computers, Environment and Urban Systems*, 32, 396–406. [302]