ORIGINAL PAPER

Stochastic smoothing of point processes for wildlife-vehicle collisions on road networks

M. I. Borrajo[1] · C. Comas[2] · S. Costafreda-Aumedes[3] · J. Mateu[4]

Abstract

Wildlife-vehicle collisions on road networks represent a natural problem between human populations and the environment, that affects wildlife management and raise a risk to the life and safety of car drivers. We propose a statistically principled method for kernel smoothing of point pattern data on a linear network when the first-order intensity depends on covariates. In particular, we present a consistent kernel estimator for the first-order intensity function that uses a convenient relationship between the intensity and the density of events location over the network, which also exploits the theoretical relationship between the original point process on the network and its transformed process through the covariate. We derive the asymptotic bias and variance of the estimator, and adapt some data-driven bandwidth selectors to estimate the optimal bandwidth. The performance of the estimator is analysed through a simulation study under inhomogeneous scenarios. We present a real data analysis on wildlife-vehicle collisions in a region of North-East of Spain.

Keywords Bandwidth selection · Covariates · First-order intensity · Kernel estimation · Linear network · Spatial point pattern · Wildlife-vehicle accidents

Mathematics Subject Classification 62G05 · 62M30 · 62P12

1 Introduction

Spatial point processes are mathematical models that describe the geometrical structure of patterns formed by events, which are distributed randomly in number and space. In the last decades we have seen an explosion in the literature devoted to point processes, see Illian et al. (2008), Diggle (2013) and Baddeley et al. (2015), however, in most of the cases this literature has been devoted to spatial point processes defined on the euclidean plane.

✉ M. I. Borrajo
   mariaisabel.borrajo@usc.es

1   Department of Mathematics and Statistics and Data Science Institute, Lancaster University, Lancaster, UK

2   Department of Mathematics, Universitat de Lleida, Lleida, Spain

3   Department of Agri-food production and Environment sciences, University of Florence, Florence, Italy

4   Department of Mathematics, Universitat Jaume I de Castellón, Castelló de la Plana, Spain

In spatial statistics, there are real problems such as the location of traffic accidents in a geographical area or geo-coded locations of crimes in the streets whose domain is, by definition, restricted to a linear network. In recent years, researchers are making an effort to deal with this particular scenario and point patterns on linear networks and their associated statistical analysis have gained a considerable amount of interest. The study of points that occur, for example, on a road network has become increasingly popular during the last few decades; in particular street crimes, see Ang et al. (2012), traffic accidents, see Yamada and Thill (2004), Xie and Yan (2008), wildlife-vehicle collisions, see Díaz-Varela et al. (2011), Morelle et al. (2013) or invasive plant species, see Spooner et al. (2004), Deckers et al. (2005), amongst many others, are examples of events occurring on a network structure. Note that in all these examples, the events occur on line segments and they are not expected to be located outside the corresponding network. For instance, wildlife-vehicle collisions are always constrained to lie along a linear network, and as such the resulting point pattern depends on the spatial configuration of such linear structures.

The analysis of network data is challenging because of geometrical complexities, unique methodological problems derived from their particular structure, and also the huge volume of data. Estimates of the spatially-varying frequency of events on a network are needed for emergency response planning, accident research, urban modelling and other purposes.

In the analysis of spatial point patterns, see for example Van Lieshout (2000), Diggle (2013), Baddeley et al. (2015), exploratory investigation often starts with non-parametric analysis of the spatial intensity of points. The intensity function, which is a first-order moment characteristic of the point process assumed to have generated the data, reflects the abundance of points in different regions and may be seen as a heat map for the events. For most problems, it is more realistic to assume that the underlying point process is inhomogeneous, i.e., driven by a varying intensity function.

The technique which immediately comes to mind for intensity estimation is kernel density estimation, see Silverman (1986). For spatial point pattern data in two dimensions, kernel estimates are conceptually simple, see Diggle (1985), Bithell (1990), and very fast to compute using the Fast Fourier Transform (FFT), see Silverman (1982). However, for points on a network, kernel estimation is mathematically intricate and computationally expensive.

Thus far attention has mostly been paid to some non-parametric intensity estimators and second-order summary statistics, such as K- and pair correlation functions. Regarding intensity estimation, initially several poorly performing kernel-based intensity estimators were proposed, see Borruso (2005), Borruso (2008), Xie and Yan (2008). Later, other nonparametric kernel-based intensity estimators were defined, see for example Okabe et al. (2009), Okabe and Sugihara (2012), McSwiggan et al. (2017), Moradi et al. (2018) which, although being statistically well-defined, tend to be computationally expensive on large networks. As an alternative to kernel estimation, Moradi et al. (2019) introduced their so-called resample-smoothed Voronoi intensity estimator, which is defined for point processes on arbitrary spaces. Moreover, Rakshit et al. (2019) proposed a fast kernel intensity estimator based on a two-dimensional convolution which can be computed rapidly even on large networks.

However, none of these approaches take into account covariate information that is easily expected to have a direct effect on the intensity function. For example, considering underlying causes such as orography, demography and human mobility have an impact on the intensity and it is quite common to encounter sharp boundaries between high and low concentrations of events due to this covariate effect. The classical kernel estimation approach is often

unsuitable in such cases and echoing (Barr and Schoenberg 2010), we argue that kernel-based approaches may be unsatisfactory if they miss out covariate information. In this line, Borrajo et al. (2020) consider kernel estimation of the intensity under the presence of spatial covariates when the point pattern lives in the Euclidean plane. However, linear network point process versions have not yet appeared in the literature. In this paper we tackle this problem and propose a covariate-based kernel estimation for point processes on linear networks, showing its advantages on a wildlife-vehicle collision problem.

The paper is organised as follows. In Sect. 2 the problem and the data set that motivates the paper are presented. In Sect. 3 we provide some definitions and preliminaries of spatial point processes on linear networks needed for the new methodological approach presented later on. Section 4 shows some theoretical results on kernel estimation in the presence of spatial covariates related to the network structure. Optimal bandwidth selection is detailed in Sect. 5. Some simulated examples are presented in Sect. 6, and the real data is analysed in Sect. 7. The paper ends with a final discussion.

2 Wildlife-vehicle collisions on road networks

Among the variety of events and related problems that can occur on a linear structure, wildlife-vehicle collisions on road networks represent a good example of such type of data and a major safety issue. In particular, wildlife-vehicle collisions are one of the main coexistence problems that arise between human populations and the environment, affecting wildlife management, the building of road infrastructures and road safety in general terms. These accidents mean a risk to the life and safety of car drivers, property damage to vehicles, see Díaz-Varela et al. (2011), Bruinderink and Hazebroek (1996), and direct and indirect damage to wildlife populations, see Coffin (2007).

For instance, in 2017 in Spain, wildlife-vehicle collisions were the fourth external causes of death behind suicides, drownings and accidental falls (Press release of the INE, October 2018). Moreover, in 2018 there were 102299 traffic accidents with victims (1679 of them with fatalities), of which at least 403 were caused by wildlife-vehicle collisions (6 of them with fatalities), see Anuario Estadistica DGT (2018). These numbers highlight the importance and severity of wildlife-vehicle collisions. Sáez-de-Santa-María and Telleria (2015) established that 8.9% of the collisions that occurred in Spain between 2006 and 2011, 74600 collisions in total, are related to fauna, although their spatial distribution is very irregular; wild boar (*Sus scrofa*)

and roe deer (*Capreolus capreolus*), both with expanding populations, caused 79% of the collisions.

Consequently, the evaluation and description of the factors that affect these accidents on the road is a priority for determining effective mitigation measures and eradication of this cause of accidents for decades, see Lord and Mannering (2010). In this paper, we analyse a dataset containing 6590 wildlife-vehicle collisions occurred in the region of Catalonia, North-East of Spain, involving 11790 km of roads for three distinct road categories, namely, highways, paved and unpaved roads during the period 2010−2014, see Fig. 1. Two covariates were considered to analyse their effect on the spatial distribution: surface of forests and the surface of crop fields, which can affect the spatial distribution of the local wildlife and then also the spatial distribution of the wildlife-vehicle collisions. Visual inspection of this spatial structure reveals points forming aggregations on the road network, thereby suggesting the presence of hot-spots of wildlife-vehicle collisions probably due to a certain degree of inhomogeneity. The analysis of this motivating real data problem is fully detailed in Sect. 7.

3 Point processes on linear networks

This section provides a short overview of some concepts of point processes on linear networks following the developments in Ang et al. (2012), Moradi et al. (2018), Moradi et al. (2019). We need to introduce some notation and concepts: let $\mathbb{R}^2$ denote the two-dimensional Euclidean space, $\|\cdot\|$ the two-dimensional Euclidean norm, and all subsets under consideration will be Borel sets in the corresponding space. Moreover, $\int d_1 u$ will be used to denote integration with respect to arc length, and $\int dx$ will be used to denote integration with respect to Lebesgue measure.

Linear networks are convenient tools for approximating geometric graphs/spatial networks. The spatial statistical literature usually defines a linear network as a finite union of (non-disjoint) line segments. More specifically, we define a linear network as a union

$$L = \bigcup_{i=1}^{k} l_i, \qquad 1 \le k < \infty,$$

of k line segments $l_i = [u_i, v_i] = \{tu_i + (1-t)v_i : 0 \le t \le 1\} \subseteq \mathbb{R}^2$, $u_i \ne v_i \in \mathbb{R}^2$, with (arc) lengths $|l_i| = \|u_i - v_i\| \in (0, \infty)$, $i = 1, \ldots, k$, which are such that any intersection $l_i \cap l_j$, $j \ne i$, is either empty or given by line segment end points. We here restrict ourselves to connected networks since disconnected ones may simply be represented as unions of connected ones.

The end points of line segments are called *nodes* and the degree of a node is the number of line segments sharing this same node. A *path* between any two points $x, y \in L$ along L is a sequence $\pi = (x, p_1, \ldots, p_P, y)$ where p_i are nodes of the linear network such that $\exists i \, / \, l_i = [p_i, p_{i+1}]$. We can then use as metric the *shortest-path distance* between any two points $x, y \in L$, $d_L(x, y)$, defined as the length of the shortest path in L between x and y.

The Borel sets on L are given by $\mathcal{B}(L) = \{A \cap L : A \subseteq \mathbb{R}^2\}$ and they coincide with the σ-algebra generated by $\tau_L = \{A \cap L : A \text{ is an open subset of } \mathbb{R}^2\}$. Recall that $A \subseteq L$ will mean that A belongs to $\mathcal{B}(L)$. We further endow L with the Borel measure $|A| = v_L(A) = \int_A d_1 u$, $A \subseteq L$, which represents integration with respect to arc length. Note that the total network length is given by $|L| = \sum_{i=1}^{k} |l_i|$.

More formally, given some probability space $(\Omega, \mathcal{A}, \P)$, a finite simple point process $\mathbf{X} = \{x_i\}_{i=1}^{n}$, $0 \le n < \infty$, on a linear network L is a random element in the measurable space N_{lf} of finite point configurations
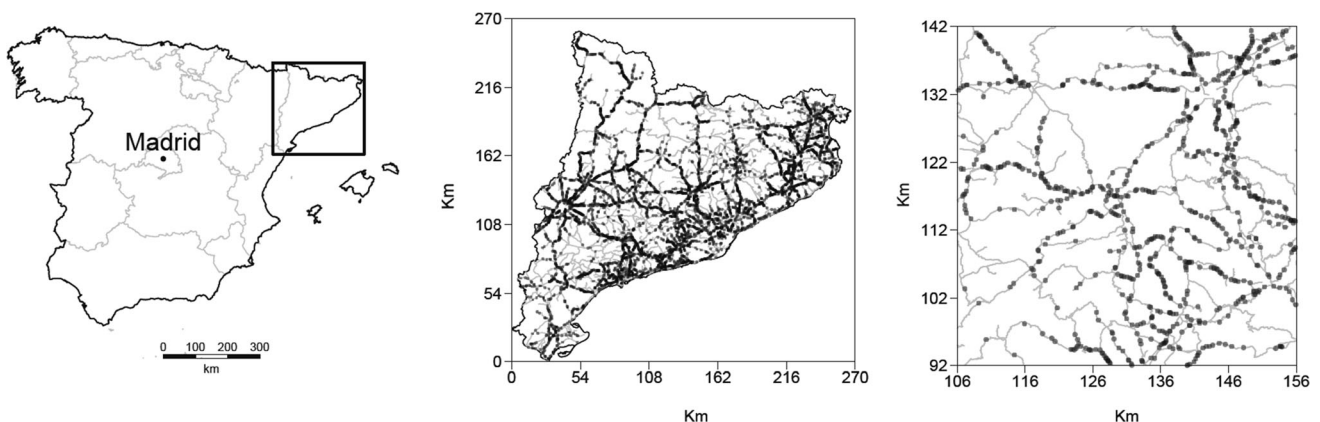
Fig. 1 *Left and centre*: location of the study area together with the location of 6590 road kills during the period 2010 − 2014 and the underlying road network in Catalonia (North-East of Spain), given in km. *Right*: a magnification of the central area of the study region (50 km × 50 km)

$\mathbf{x} = \{x_1, \ldots, x_n\} \subseteq L$, $0 \le n < \infty$; the associated σ-algebra is generated by the cardinality mappings $\mathbf{x} \mapsto N(\mathbf{x} \cap A) \in \{0, 1, \ldots\}$, $A \subseteq L$, $\mathbf{x} \in N_{lf}$, and coincides with the Borel σ-algebra generated by a certain metric on N_{lf}, see Daley and Vere-Jones (1988), p. 188 for details.

The *intensity function* $\lambda(u)$ of $\mathbf{X}$ gives the expected number of points per unit *length* of network in the vicinity of location u. Formally, $\mathbf{X}$ has intensity function $\lambda(u)$, $u \in L$, if

$$E[N(\mathbf{X} \cap B)] = \int_B \lambda(u) \, d_1 u, \tag{1}$$

for all measurable $B \subset L$, where $N(\mathbf{X} \cap B)$ is the number of points of $\mathbf{X}$ falling in B. We note that N stands for a random quantity coming from the counting random variable, while we denote by n the fixed number of points given a point pattern. *Campbell's formula* on a network states that

$$E\left[\sum_{x_i \in \mathbf{X}} h(x_i)\right] = \int_L h(u)\lambda(u) \, d_1 u, \tag{2}$$

where h is any Borel-measurable real function on L such that $\int_L |h(u)|\lambda(u) \, d_1 u < \infty$.

The literature on spatial point patterns on linear networks is not extensive yet, being this a relatively new topic. Different examples of spatial point patterns on linear networks can be found in Ang et al. (2012), Okabe and Sugihara (2012), McSwiggan et al. (2017), Moradi et al. (2018, 2019), Rakshit et al. (2019).

4 Covariate-dependent kernel-based intensity estimation

To analyse a point process we can take into account not only the spatial information given by the location of the events, but also some other external information that commonly appears in the form of covariates.

In this framework of point processes with covariates, let $Z : L \subset W \subset \mathbb{R}^2 \to \mathbb{R}$ be a spatial continuous covariate that is exactly known in every point of W, and particularly in every point of the network. Along this paper, and following Baddeley et al. (2012), we assume that the intensity can be described from the known covariate through the model

$$\lambda(u) = \rho(Z(u)), u \in L, \tag{3}$$

where ρ is an unknown function with no assumptions on it. As Z is known, the only target for intensity estimation is the function ρ.

Our aim here is to propose a kernel intensity estimator for processes on linear networks under model (3). Following previous literature in the field of spatial point processes with covariates, see Baddeley et al. (2012), Borrajo et al. (2020), we work with the transformed univariate point process, $Z(\mathbf{X})$, i.e., for any point pattern $\{x_1, \ldots, x_{N(L)}\} \subset L$ we compute $Z_i = Z(x_i)$.

To exploit and adapt the ideas in Borrajo et al. (2020) to the context of linear networks, we need to establish the theoretical relationship between the original point process $\mathbf{X}$ and the corresponding transformed one, $Z(\mathbf{X})$. First, we have to prove that the transformed point process, $Z(\mathbf{X})$, is indeed a point process. Second, we need to theoretically derive the expression of the intensity function of the transformed point process and its relationship with the original one so that we can still estimate λ.

The result establishing this relationship can not be directly transferred from the spatial context to the network domain, because of the different geometry of the support and in the metrics (shortest path distance instead of Euclidean one).

The following result characterises the transformed point process from the one on the network through a spatial covariate. The proof is included in the Appendix.

Theorem 1 *Let $\mathbf{X}$ be a point process on a linear network $L \subset \mathbb{R}^2$ with intensity function of the form $\lambda(u) = \rho(Z(u)) \; u \in L$, where $Z : L \subset W \subset \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function with non-zero gradient in every point of its domain. Then, the transformed point process $Z(\mathbf{X})$ is a one-dimensional point process with intensity function $\rho g^\star$, where $g^\star = (G^\star)'$ and $G^\star(z) = \int_L 1_{\{Z(u) \le z\}} d_1 u$ is the unnormalised version of the spatial cumulative distribution function of the covariate. Furthermore, if the original point process is Poisson, this property is inherited and the transformed one is also Poisson.*

Hence, we have shown that $Z(\mathbf{X})$ is a point process in $\mathbb{R}$ with intensity given by $\rho g^\star$. This characterisation of the intensity will be used to obtain a consistent kernel intensity estimator, jointly with the existing convenient relationship between the density and the intensity functions. The latter has been previously applied in slightly different contexts, see Cucala (2006), Fuentes-Santos et al. (2015), Borrajo et al. (2020), but not yet transferred to the network domain.

Let us define the relative density

$$f(z) = \frac{\rho(z)g^\star(z)}{m}, \tag{4}$$

where $m = \int_L \lambda(u) d_1 u$ using the integration with respect to arc length as explained in Sect. 3.

The kernel density estimate is structurally the same as the one proposed by Borrajo et al. (2020), and takes the form

$$\hat{f}_h(z) = g^{\star}(z)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{g^{\star}(Z_i)}K_h(z - Z_i)1_{\{n\neq 0\}}. \tag{5}$$

However the nature of the elements involved is quite different because of the linear network domain and the use of the shortest path distance replacing the Euclidean norm. This fact requires a different theoretical treatment and the use of tools adapted to the network domain.

The global idea is to plug-in (5) and an estimate of m into (4), and then obtain an estimate of ρ which can be used in (3) to build an estimator of λ as follows

$$\hat{\lambda}_h(u) = \sum_{i=1}^{n}\frac{1}{g^{\star}(Z_i)}K_h(Z(u) - Z_i). \tag{6}$$

We note that $g^{\star}(\cdot)$ is obtained using a classical one-dimensional kernel estimator over the transformed data, Z_i, with $i = 1,\ldots,n$. Following Borrajo et al. (2020), under a Poisson assumption and using an infill structure asymptotic framework (which means that the observation region remains fixed while the sample size increases), we can compute the mean squared error $MSE(h, z) = E[\{\hat{f}_h(z) - f(z)\}^2]$ of (5). Remark that in this scenario the bandwidth h is considered as a function of the expected sample size, this is, $h \equiv h(m)$ and its properties are shown when $m \to \infty$. The following result, which is an adaptation of Borrajo et al. (2020) to the network scenario, provides a close form for the $MSE(h, z)$.

Theorem 2 *Let us assume some regularity conditions:*

(A.1) $\int_{\mathbb{R}} K(z)dz = 1;$ $\int_{\mathbb{R}} zK(z)dz = 0$ *and* $\mu_2(K) := \int_{\mathbb{R}} z^2K(z)dz < \infty.$

(A.2) $\lim_{m\to\infty} h = 0$ *and* $\lim_{m\to\infty}\frac{A(m)}{h} = 0,$ *where* $A(m) := E\left[\frac{1}{N(\mathbf{X})}1_{\{N(\mathbf{X})\neq 0\}}\right].$

(A.3) G *is three times differentiable.*

(A.4) z *is a continuity point of* ρ.

Then under (A.1) to (A.4) we have that

$$E\left[\hat{f}_h(z)\right] = \frac{g^{\star}(z)(K_h \circ \rho)(z)}{m}(1 - e^{-m}) \text{ and}$$

$$Var\left[\hat{f}_h(z)\right] = A(m)\frac{(g^{\star}(z))^2}{m}\left(K_h^2 \circ \frac{\rho}{g^{\star}}\right)(z)$$
$$- (A(m) + e^{-2m} - e^{-m})(g^{\star}(z))^2(K_h \circ \rho)^2(z),$$

where $\circ$ denotes the convolution between two functions.
Moreover, adding condition (A.5)

(A.5) ρ *is three times differentiable*

we have

$$MSE(h, z) = e^{-2m}f^2(z) + (1 - e^{-m})^2\frac{h^4}{4}\left(\frac{\rho''(z)g^{\star}(z)}{m}\right)^2\mu_2^2(K)$$
$$- e^{-m}(1 - e^{-m})h^2\mu_2(K)\frac{(g^{\star}(z))^2\rho(z)\rho''(z)}{m^2}$$
$$+ \frac{A(m)}{h}f(z)R(K) + o(h^2(1 - e^{-m})e^{-m})$$
$$+ o(h^4(1 - e^{-m})^2) + o\left(\frac{A(m)}{mh}\right),$$

where $R(K) = \int_{\mathbb{R}} K^2(z)dz.$

The proof is omitted as follows the same arguments as the ones used in Borrajo et al. (2020). A direct consequence of this result is that the mean integrated square error $MISE(h) = E\int\{\hat{f}_h(z) - f(z)\}^2 dz$, as well as its asymptotic version, denoted by $AMISE$, can both be derived

$$MISE(h) = e^{-2m}R(f) + (1 - e^{-m})^2\frac{h^4}{4}R$$
$$\left(\frac{\rho''g^{\star}}{m}\right)\mu_2^2(K)$$
$$- e^{-m}(1 - e^{-m})h^2\mu_2(K)$$
$$\int_{\mathbb{R}}\frac{g^{\star}(z)\rho''(z)f(z)}{m}dz$$
$$+ \frac{A(m)}{h}R(K) + o(h^2(1 - e^{-m})e^{-m})$$
$$+ o(h^4(1 - e^{-m})^2) + o\left(\frac{A(m)}{mh}\right) \text{ and} \tag{7}$$

$$AMISE(h) = (1 - e^{-m})^2\frac{h^4}{4}R\left(\frac{\rho''g^{\star}}{m}\right)\mu_2^2(K)$$
$$+ \frac{A(m)}{h}R(K).$$

Based on the asymptotic expression obtained in (7), the optimal bandwidth value in this sense, which minimises the $AMISE$, is given by

$$h_{AMISE} = \left(\frac{A(m)R(K)}{\mu_2^2(K)(1 - e^{-m})^2R\left(\frac{\rho''g^{\star}}{m}\right)}\right)^{1/5}. \tag{8}$$

4.1 A note on covariates on networks

The use of the intensity estimator (6) on a linear network needs a direct evaluation of a covariate over the linear

network. In the linear network framework the inclusion of covariates is not straightforward, indeed a distinction between two different types of covariates needs to be done.

Assuming the linear network, included in the Euclidean plane, is our current domain, and hence the "region" of interest, a first approach is to use covariates which are just defined on the linear network itself, i.e., their domain is this union of linear segments. A second setup is to take into account covariates that are defined in a spatial region of the Euclidean plane containing the network and having an impact on it (understanding the network as a subset of the Euclidean plane, any spatial region containing its convex hull).

We can have two types of covariates that can be related to a linear network: internal and external. In this work we only deal with the latter. This distinction affects the distribution of points on this structure, as well as the tools required to analyse them. For instance, the percentage of forest coverage is an external covariate that can affect wildlife-vehicle collisions over a road crossing this region. Moreover, examples of internal covariates include road slope, road visibility and traffic road intensity among others. For external covariates, as they are not defined over the linear network, we need to approximate its effects on the spatial distribution of points over this linear structure. A tentative way to do so is to take the average value of this covariate in a cirle of radius r centrered at every point of the linear network. As such, the average effect of this covariate is considered to analyse its effects on the distribution of points on this linear structure.

5 Bandwidth selection methods

We note that there is no reference so far in the literature of selectors adapted to the network case under the presence of covariates. We thus recall here several bandwidth selection procedures that have been used under model (3) for planar spatial point processes in Borrajo et al. (2020) and adapt them to the context of point processes on linear networks. This adaptation is possible due to the inclusion of covariates. In Borrajo et al. (2020) the authors show that the bootstrap bandwidth selector generally outperforms the others, however this is not necessary the case for linear networks, whose specific structure may affect the performance of the bandwidth selectors. Hence we include all the available possibilities.

5.1 Rule-of-thumb

In the literature two different bandwidths assuming normality have been used. In Baddeley et al. (2012), the authors propose to use Silverman's bandwidth defined for

the classical kernel density estimator, see Silverman (1986), directly applied to the transformed point pattern, $Z_1, \ldots, Z_n$, where n is the observed sample size for a specific realisation. This bandwidth selector will be denoted from now on as $\hat{h}_{Silv}$.

A more elaborated approach has been proposed in Borrajo et al. (2020), based on the optimal bandwidth (8) and assuming normality of the relative density. We adapt this procedure by using our relative density 4 and by estimating the corresponding quantities on the network domain; the resulting selector will be denoted by $\hat{h}_{RT}$.

5.2 Bootstrap bandwidth

The bootstrap bandwidth presented in Borrajo et al. (2020) is based on a consistent resampling bootstrap procedure that the authors have defined for the transformed point process under the Poisson assumption. This idea can be directly applied to our transformed point process, and then we can obtain the following bandwidth selector

$$\hat{h}_{Boot} = \left(\frac{A(\hat{m})R(K)}{\mu_2^2(K)(1 - e^{-\hat{m}})^2 R(\frac{\hat{\rho}_b'' g^\star}{\hat{m}})} \right)^{1/5},$$

where the pilot bandwidth b is computed as an appropriate rescaled version of the previously presented rule-of-thumb, $\hat{h}_{RT}$. Numerical integration is required to compute the values $\hat{m}$, $A(\hat{m})$ and $R(\frac{\hat{\rho}_b'' g^\star}{m})$.

5.3 Non-model-based approach

This is a recent bandwidth selector that has been initially proposed in Cronie and van Lieshout (2018) for spatial point processes and later in Borrajo et al. (2020) for spatial point processes with covariates. The initial idea relies on the fact that $\int_D \frac{1}{\hat{\lambda}(x)} \lambda(x) dx = |D|$, which allows building a discrepancy measure between the inverse of the kernel intensity estimator by Diggle (1985) and the area of the observation region, $|D|$. Minimising this discrepancy measure, a data-driven bandwidth selection procedure is obtained. The adaptation to the context of network point processes with covariates involves replacing Diggle's estimator by (6) and computing the same minimising criteria, where now the measure of the region is the length of the network, $|L|$

$$\hat{h}_{NM} = \arg \min_{h > 0}(T(h) - |L|),$$

where $T(h) = \sum_{i=1}^N \hat{\rho}_h(Z_i)^{-1}$ inside L, and $|L|$ elsewhere. Remark that this criterion does not aim to optimise the bias-variance trade-off of the kernel intensity estimator and

therefore does not guarantee to provide good intensity estimates.

6 Simulation study

6.1 Setup

We present an illustration to show the use of the new kernel intensity estimator and the bandwidth selection methods under several point configurations defined over a linear network in a square area of 50×50 km² in the centre of Catalonia (North-East of Spain), see Fig. 1. This linear network involves 1267 km of roads for three distinct road categories, namely, highways, paved and unpaved roads. We consider inhomogeneous Poisson point processes with intensity function given by

$$\lambda(u) = \exp(\beta_0 + \beta_1 Z(u)), \ u \in W, \tag{9}$$

where β_0 and β_1 are known parameters, Z denotes a covariate, and W is the planar region of study. The covariate comes from a realisation of a Gaussian random field, with mean zero and an exponential covariance structure with parameters $\sigma = 0.316$ and $s = 150$. Thus the covariance function is given by $C(r) = \sigma^2 \exp(-r/s)$, together with $\beta_0 = 3$ and $\beta_1 = 1$. This is an external covariate, see Sect. 4.1, hence to calculate the value defined in the Euclidean plane over the linear network, we consider the average value of this covariate in a circle of radius $r = 0.5$ km. centred at points of the linear network. Once the covariate is defined on the linear network, we construct an intensity function following equation (9). We further obtain patterns from an inhomogeneous Poisson point process with this intensity to evaluate the performance of our proposals. Figure 2 shows the intensity function and a realisation of this inhomogeneous Poisson point process over the linear network.

6.2 Simulated examples

We conduct a simulation study to estimate the intensity function based on this external covariate to show the performance of the resulting intensity estimators for each bandwidth selector. We simulate 500 realisations for four expected sample sizes $m = 100, 300, 700, 1000$ points, to analyse its effect. Note that for $m = 700$ the scenario is similar to the one of our real data problem of wildlife-vehicle collisions, both with around 0.56 events per linear km. To guarantee the expected sample size on each scenario, we need to appropriately scale the intensity function given in (9). From the simulated samples, we evaluate the performance of our intensity estimator (6) with different bandwidth choices through three different measures. We first consider the relative integrated squared error

$$\text{ISE}_{\text{rel}}(\hat{h}) = \int_L \left(\frac{\hat{\lambda}(u) - \lambda(u)}{\lambda(u)} \right)^2 du,$$

where recall that L denotes the network, and define the first two performance measures as

$$e_1 = \text{mean}(\text{ISE}_{\text{rel}}(\hat{h})) \ \text{and} \ e_2 = \text{std}(\text{ISE}_{\text{rel}}(\hat{h})),$$

the average relative error and its variability. On the other hand, most of the bandwidth selectors adapted in this paper aim to estimate the infeasible optimal bandwidth that minimises the $MISE(h)$ criterion. So it is natural to consider such infeasible value as a benchmark in our simulations, and measure how close our estimates are from such value. This motivates our third performance measure that is the relative bias of the bandwidth selectors, defined as

$$e_3 = \text{mean}\big((\hat{h} - \hat{h}_{\text{MISE}})/\hat{h}_{\text{MISE}}\big)$$

where $\hat{h}_{\text{MISE}}$ is the minimiser of the Monte Carlo approximation (based on the 500 simulated samples) of the $MISE(h)$ criterion.

Table 1 shows the performance of the resulting bandwidth selectors $\hat{h}_{MISE}$, $\hat{h}_{RT}$, $\hat{h}_{Boot}$, $\hat{h}_{NM}$, together with the selector proposed by Baddeley et al. (2012), $\hat{h}_{Silv}$. We compute them for inhomogeneous Poisson point processes with intensity given in (9) and the four expected sample sizes. This table highlights that independently of the point intensity, the bootstrap bandwidth selector is the method that outperforms the rest, followed by Silverman's bandwidth selector, the non-model-based approach, and finally the rule-of-thumb. This result gains in strength when the point intensity increases. Moreover, any bandwidth selector increasing the expected sample size, the corresponding error decreases approaching the values of the optimal bandwidth minimising the $MISE(h)$ criterion. In terms of variability, the four proposals show a similar behaviour as it is reflected in measure $e2$.

Figure 3 shows boxplots of the resulting four bandwidth selectors based on 500 point pattern realisations. We note that, independently of the point intensity, these selectors are always smaller than the optimal bandwidth value that minimises the $MISE(h)$ criterion (h_{MISE}). The resulting average values of the four selectors show the same sort of behaviour as that observed for the $e1$ measure, i.e. the bootstrap method is the bandwidth selector that is closer to the $\hat{h}_{MISE}$, followed by Silverman's method, the non-model-based approach, and finally the rule-of-thumb. In terms of variability (see also criteria $e2$ in Table 1), the four
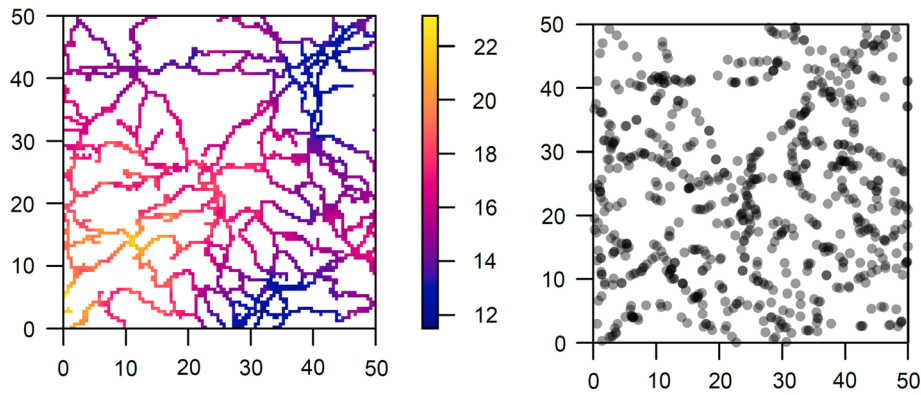
Fig. 2 Left: intensity function over the linear network corresponding to Fig. 1 (right panel) based on a realisation of a zero-mean Gaussian random field with parameters $\sigma = 0.316$, $s = 150$, $\beta_0 = 3$ and $\beta_1 = 1$, assuming the average value of this resulting covariate in a circle of radius $r = 0.5$ km centrered at points on the linear network. Right: a random realisation (709 points) of a point pattern from an inhomogeneous Poisson with intensity (9)

Table 1 Measure values $e1$, $e2$ and $e3$ based on 500 point patterns realisations for bandwidth selectors $\hat{h}_{MISE}$, $\hat{h}_{RT}$, $\hat{h}_{Boot}$, $\hat{h}_{NM}$, and $\hat{h}_{Silv}$ (see descriptions in Sect. 4), under four point intensities and the intensity function corresponding to Fig. 2, generated by the proposed inhomogeneous Poisson process

	$\hat{h}_{MISE}$	$\hat{h}_{RT}$	$\hat{h}_{Boot}$	$\hat{h}_{NM}$	$\hat{h}_{Silv}$
$m = 100$					
$e1$	28.982	54.897	39.818	44.909	41.978
$e2$	16.595	29.654	24.800	23.748	25.313
$e3$	–	− 0.605	− 0.410	− 0.525	− 0.488
$m = 300$					
$e1$	15.485	25.405	18.280	24.719	19.110
$e2$	7.402	10.909	9.615	10.463	9.155
$e3$	–	− 0.533	− 0.318	− 0.518	− 0.392
$m = 700$					
$e1$	10.308	16.609	11.430	16.126	12.242
$e2$	4.860	6.291	4.998	6.079	5.156
$e3$	–	− 0.457	− 0.197	− 0.445	− 0.292
$m = 1000$					
$e1$	8.610	14.339	9.761	12.885	10.551
$e2$	3.812	4.938	3.891	4.481	3.942
$e3$	–	− 0.466	− 0.199	− 0.417	− 0.303

compared methods are similar, although the selectors $\hat{h}_{Boot}$ and $\hat{h}_{NM}$ present more heterogeneity for smaller point intensities.

As a final assessment to reinforce our approach for networks against the alternative existing procedure in Borrajo et al. (2020), we evaluated the measures $e1$, $e2$ under the same scenarios as in Table 1, but now just missing out the existence of the network, i.e., assuming the realisations of the point process lying on the Euclidean plane. We thus used the estimators proposed in Borrajo

et al. (2020) for the plane with point patterns simulated on the network, but assuming that the plane is their support.

In Table 2 we sum up the corresponding results, including measures $e1$ and $e2$. Clearly the magnitudes of this two performance measures are far much larger than those shown in Table 1 under our proposal. This reflects the need of considering our adaptation to the network and reinforces the fact that the network effect can not be missed out. Figure 4 shows the corresponding spatial intensity used in the simulations.

7 Case study: wildlife-vehicle collisions on a road network

We now illustrate the use of our kernel intensity estimator and the bandwidth selection methods by analysing the real data set involving wildlife-vehicle collisions on a road network, initially introduced in Sect. 2. Let us recall that the data set contains the whole road network of Catalonia (North-East of Spain) involving 11790 km of roads for three distinct road categories, namely, highways, paved and unpaved roads, and provides the locations of 6590 wildlife-vehicle collision points occurred during the period $2010 − 2014$. Most of the roadkills involve ungulates and other non-identified mammals.

Inspection of Fig. 1 reveals points forming aggregations along some of the roads, suggesting the presence of a cluster structure of roadkills along the roads. Two covariates were considered to analyse their effects on the spatial distribution of roadkills: surface of forests and surface of crop fields, given as a percentage based on a buffer area of 0.5 km around the roads. Several authors have considered these two covariates as possible predictors of wildlife-vehicle collisions, see for instance, Ha and Shilling (2018), Hegland and Hamre (2018), Tatewaki and Kioke (2018)
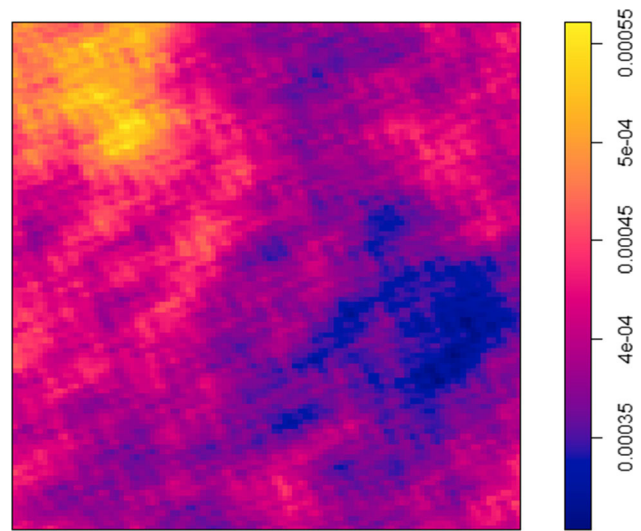
Fig. 3 Boxplots of the four bandwidth selectors based on 500 point pattern realisations of the intensity function corresponding to Fig. 2, for the four point intensities (from top left to down right, the expected sizes are 100, 300, 700 and 1000 points). The horizontal red line is the $\hat{h}_{MISE}$ value
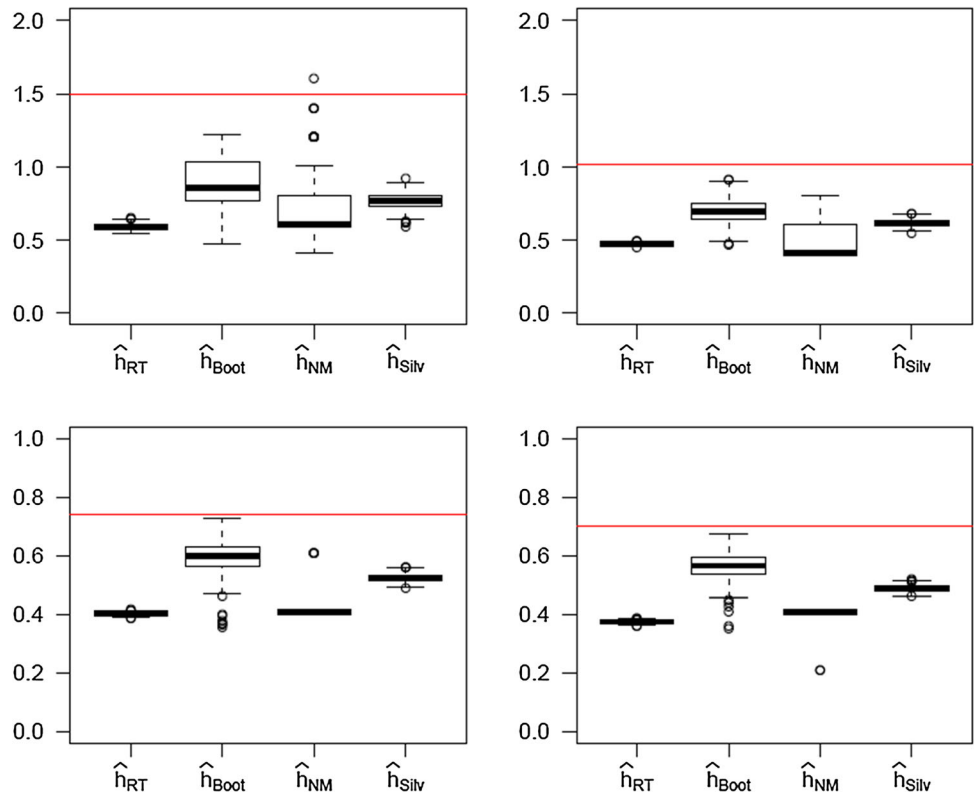
Table 2 Measure values $e1$ and $e2$ based on 500 point patterns realisations for bandwidth selectors $\hat{h}_{RT}$, $\hat{h}_{Boot}$, $\hat{h}_{NM}$, and $\hat{h}_{Silv}$ (see descriptions in Sect. 4), under four point intensities and the intensity function corresponding to Fig. 4, generated by the proposed inhomogeneous Poisson process, where the network structure is omitted and $W \subset \mathbb{R}^2$ is the considered support

	$\hat{h}_{RT}$	$\hat{h}_{Boot}$	$\hat{h}_{NM}$	$\hat{h}_{Silv}$
$m = 100$				
$e1$	200.776	174.060	155.336	337.542
$e2$	111.962	104.191	103.467	1045.431
$m = 300$				
$e1$	121.865	109.318	98.966	523.572
$e2$	52.005	49.681	47.131	811.909
$m = 700$				
$e1$	95.679	87.436	80.681	813.417
$e2$	32.014	30.309	29.239	708.880
$m = 1000$				
$e1$	92.675	85.129	78.817	997.838
$e2$	30.605	29.694	28.339	721.105

Fig. 4 Representation of the theoretical spatial intensity used in our simulation study

(2016) analysed the effect of the surface of crop fields as a predictor for these type of traffic collisions.

Figure 5 shows the effect of these two covariates on the network, and it highlights that where the percentage of forest is high, the corresponding percentage of crop fields is low, which is an expected result. Informal visual inspection of these two covariates and wildlife-vehicle collision locations seem to show that roads with a high percentage of

who have considered forests as an explanatory covariate to explain the distribution of roadkills, whilst Hubbard et al. (2000), Acevedo et al. (2014), Colino-Rabanal and Peris
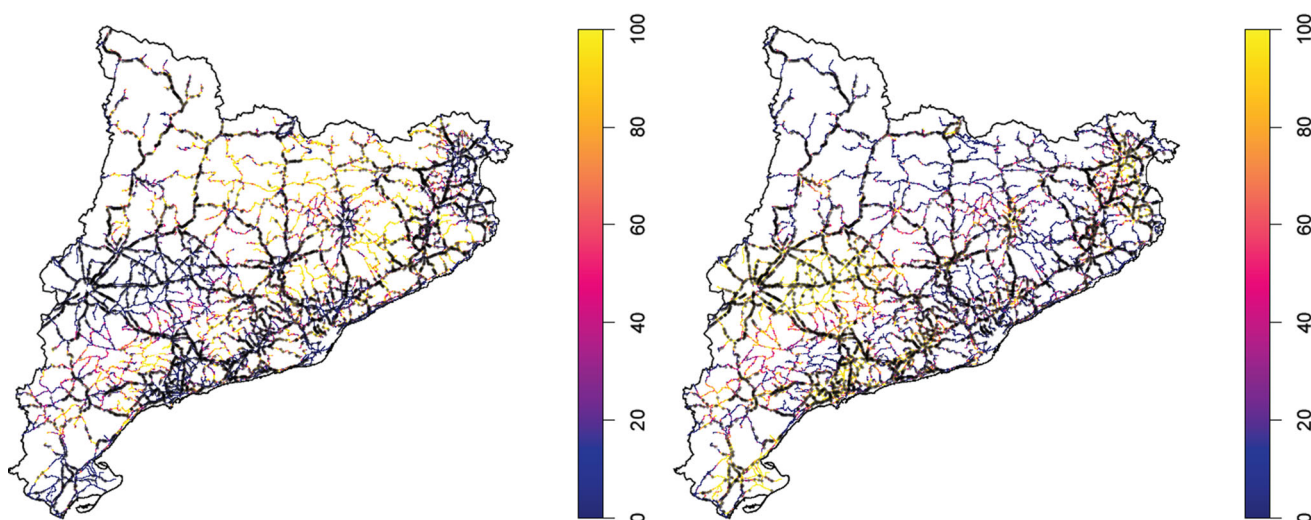
Fig. 5 Wildlife-vehicle collision point pattern together with the percentage of forests (left panel) and crop fields (right panel) around the road network of Catalonia (see Fig. 1)

crop fields have a larger number of roadkills than roads with a high percentage of forest coverage. This has to be formally proved, and we here perform such an analysis using our procedure. We consider our new kernel intensity estimation to investigate if these two covariates are relevant for the estimation of the location and number of roadkills.

Table 3 shows the values of the four bandwidth selectors for the two explanatory covariates, and we note that the resulting bandwidth values are quite distinct, being Silverman's the method with the largest bandwidth value for both covariates. Note that the resulting bandwidth values obtained under the percentage of forest and the percentage of crop fields are very similar, thereby suggesting that both covariates need the same amount of smoothing when estimating the intensity.

The resulting intensity estimations based on these four bandwidth selectors are shown in Figs. 6 and 7 for the percentage of forests and field crops, respectively. Visual inspection of these two figures shows that the largest bandwidth values, i.e. the Silverman's rule-of-thumb and the non-model-based approach, identify better the roads

with a larger number of roadkills than the other bandwidth selectors.

Moreover, Figs. 8 and 9 show the density of roadkills as a function of the percentage of forests and field crops, respectively. In terms of the density function, the results obtained for the four bandwidth are similar, although both Silverman's rule-of-thumb and the non-model-based approaches result in smoother curves than those for the rule-of-thumb approach and the bootstrap bandwidth methods. For the four bandwidth values the resulting density functions suggest a structure between the number of roadkills and the covariate values. In particular, the presence of forest around roads apparently reduces the number of wildlife-vehicle collisions, whilst crop fields around roads increase this type of traffic collisions. Note that for the rule-of-thumb proposed and the bootstrap bandwidth selector this function shows a "saw-tooth" pattern peaking at small and large percentages of both covariates, whilst for the other two bandwidth selectors this density is a smooth curve. These density patterns are expected since the bandwidth values of Silverman's rule-of-thumb and the non-model-based approaches are larger than those under the other bandwidth approaches (see Table 3). As expected, both covariates have distinct effects on the presence of road kills, though they seem to be complementary.

8 Discussion

We have presented a kernel intensity estimator in the context of spatial point processes defined on a linear network with covariates. The literature on kernel smoothing for general spatial point processes is well established, but

Table 3 Resulting values for the four bandwidth selectors corresponding to the wildlife-vehicle collision point pattern and the road network of Catalonia (see Fig. 1), under two explanatory covariates, the percentage of forests and crop fields around roads (see Fig. 5)
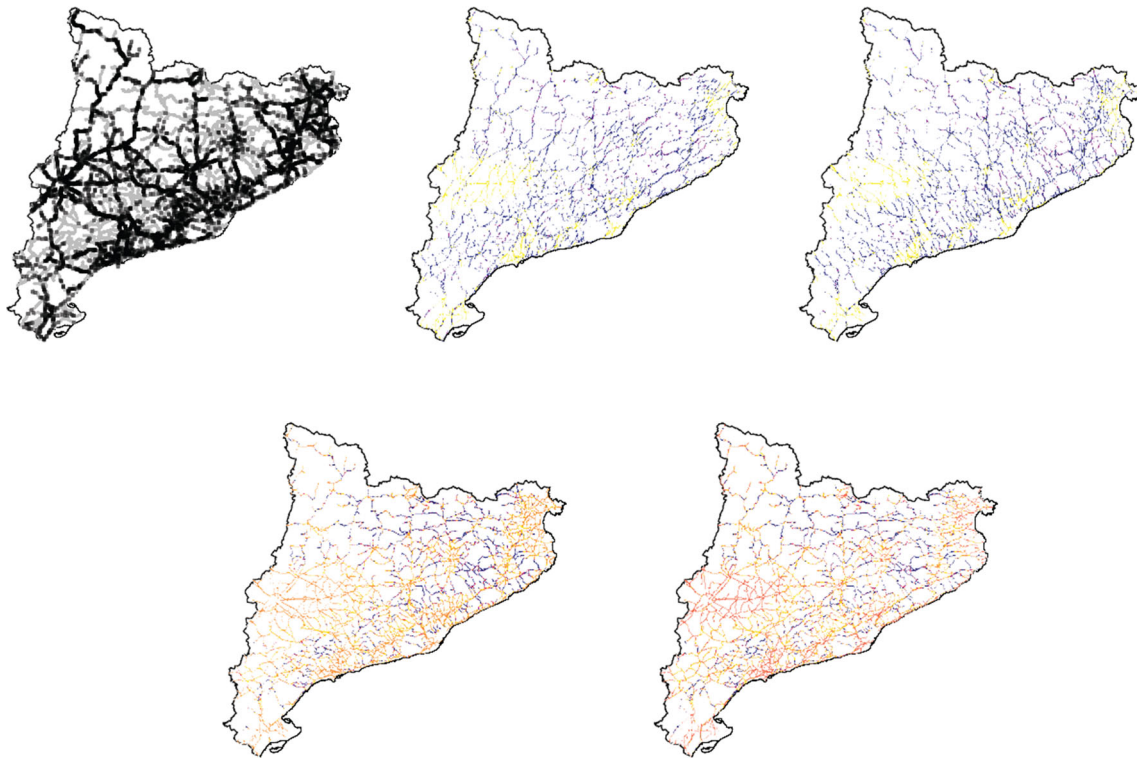
	$\hat{h}_{RT}$	$\hat{h}_{Boot}$	$\hat{h}_{NM}$	$\hat{h}_{Silv}$
Forest	0.146	0.057	4.840	5.392
Crop fields	0.146	0.070	4.840	5.747

Fig. 6 Wildlife-vehicle collision point pattern (*top, left panel*) together with the resulting intensity estimation for the percentage of forests based on bandwidth selector methods $\hat{h}_{RT}$ (*top, middle panel*), $\hat{h}_{Boot}$ (*top, right panel*), $\hat{h}_{NM}$ (*bottom, left panel*) and $\hat{h}_{Silv}$ (*bottom, right panel*). Yellow, red and blue colours represent, respectively, high, medium and low intensities

this is not the case when the intensity depends on covariates that have an impact of the spatial structure of the events, see Baddeley et al. (2012), Borrajo et al. (2020). In particular, our work is the first attempt to deal with such problem when the point pattern has as support a linear network, making available the use of covariates in this new context of high interest in several disciplines.

We have thus proposed a statistically principled method for kernel smoothing of point pattern data on a linear network when the first-order intensity depends on covariates. Our estimator relies on the relationship between the original point process on the network and its transformed process through the covariate. We have derived the asymptotic bias and variance of the estimator, and we have adapted some data-driven bandwidth selectors, previously used in the Euclidean plane context, to estimate the optimal bandwidth for linear networks.

The multivariate extension is not considered here, but it would be an interesting problem extending our estimator to the spatio-temporal case and to the case of more than one covariate. The theoretical developments in the multivariate framework require further work, and even if in Borrajo et al. (2020) the authors have skimmed some ideas about

the multivariate problem, its adaptation to the linear network domain require an extra effort. A complementary idea would be to go parametrically, and consider a fully parametric model for the first-order intensity on networks depending on covariates.

Another possible extension of this work is the use of not only external but also internal covariates (those that are not defined in the Euclidean plane but only over the network). The theoretical developments required to use these covariates are not straightforward and the authors are already working on them.

As noted in the paper, the literature on kernel smoothing depending on covariates for spatial point processes is quite limited and this paper contributes in this line for the particular case of point processes on networks. It is important to underline that estimating the first-order intensity function is also crucial for second order characteristics in inhomogeneous point processes. Network inhomogeneous K- or pair-correlation functions, as main second order tools, need an estimator of the intensity. Hence, this paper would also be useful to tackle second order problems on linear networks from a nonparametric perspective.
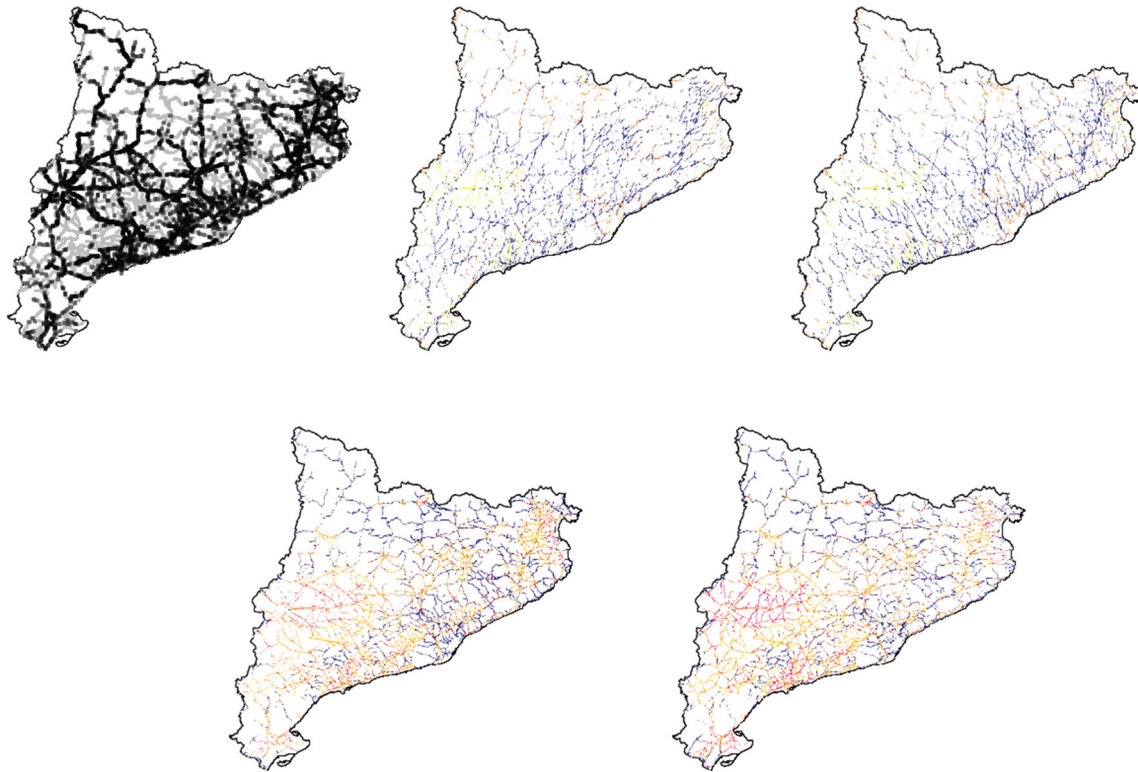
Fig. 7 Wildlife-vehicle collision point pattern (*top, left panel*) together with the resulting intensity estimation for the percentage of crop fields based on bandwidth selector methods $\hat{h}_{RT}$ (*top, middle panel*), $\hat{h}_{Boot}$ (*top, right panel*), $\hat{h}_{NM}$ (*bottom, left panel*) and $\hat{h}_{Silv}$ (*bottom, right panel*). Yellow, red and blue colours represent, respectively, high, medium and low intensities

Proof of theorem 3.1

We consider two steps in the proof: we first prove that $Z(\mathbf{X})$ is itself a point process, and then we show that its intensity is given by $\rho g^{\star}$.

Let us denote by $L_d = (L, d)$ the metric space formed by the linear network L and the shortest-path metric, d. As $\mathbf{X}$ is a point process in L_d, following Møller and Waagepetersen (2003), there is a measurable mapping $N : (\Omega, \mathcal{A}, \P) \longrightarrow (N_{lf}, \mathcal{N}_{lf})$ defined in some probability space $(\Omega, \mathcal{A}, \P)$ with N_{lf}, defined before, the space of locally finite point configurations in L_d, i.e., $N_{lf} = \{\mathbf{x} \subset L_d \text{ with } n(x_B) < \infty \text{ for all bounded } B \subset L_d\}$. Here, $x_B = \mathbf{X} \cap B$ and $n(\cdot)$ denotes the cardinality. Finally, $\mathcal{N}_{lf}$ is a σ−algebra on N_{lf}, i.e., $\mathcal{N}_{lf} = \sigma\big(\{\mathbf{x} \in N_{lf}/n(x_B) = m\} : B \in \mathcal{B}_0, m \in \mathbb{N}_0\big)$, with $\mathcal{B}_0$ the class of bounded Borel sets and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

In order to show that $Z(\mathbf{X})$ is a one-dimensional point process, with Z a real-valued covariate, we need to construct another measurable mapping associated with $Z(\mathbf{X})$, $N^Z : (\Omega^Z, \mathcal{A}^Z, \P^Z) \longrightarrow (N_{lf}^Z, \mathcal{N}_{lf}^Z)$, where

– $(\Omega^Z, \mathcal{A}^Z, \P^Z)$ is the probability space induced from $(\Omega, \mathcal{A}, \P)$ with Z, meaning that $\Omega^Z = Z(\Omega) = \{Z(x) :$

$x \in \Omega\}$, $\mathcal{A}^Z = \sigma\{Z(B) : B \in \mathcal{A}\}$, and for every $A \in \mathcal{A}^Z, \exists B \in \mathcal{A}$ such that $A = Z(B)$, then, $\P^Z(A) = \P(B)$.
– $N_{lf}^Z = \{z \subset Z(L_d), \text{ such that } n(z_A) < +\infty \text{ for } hboxall \text{ bonded } A \subset Z(L_d)\}$.
– $\mathcal{N}_{lf}^Z$ is a σ−algebra in N_{lf}^Z.

Hence, we define for every $A \in \Omega^Z$, $N^Z(A) := N(Z^{-1}(A)) = N(B)$ for a certain $B \in \Omega$ whose existence is guaranteed by construction of the induced spaces. And, as N is measurable, then N^Z is measurable.

For the second argument, i.e., to determine the intensity of the transformed process $Z(\mathbf{X})$, we make use of the result in Federer 1969, Th. 3.2.22). We first rewrite the expression of the unnormalised version of the spatial cumulative distribution function of Z

$$G^{\star}(z) = \int_L 1_{\{Z(u) \le z\}} d_1 u = \int_L 1_{\{Z(u) \le z\}} ||\nabla Z(u)||^{-1} ||\nabla Z(u)|| d_1 u$$

where $||\nabla Z(u)||$ denotes the two-dimensional Euclidean norm of the gradient of Z, that can be defined as $L \subset \mathbb{R}^2$.

Moreover, as L is assumed to be a piecewise regular path, there exists a parametrisation $\alpha : [a, b] \longrightarrow L$ that can be split in a finite number of pieces, and in each of them α is regular, i.e., it is first-order derivable and its derivative is
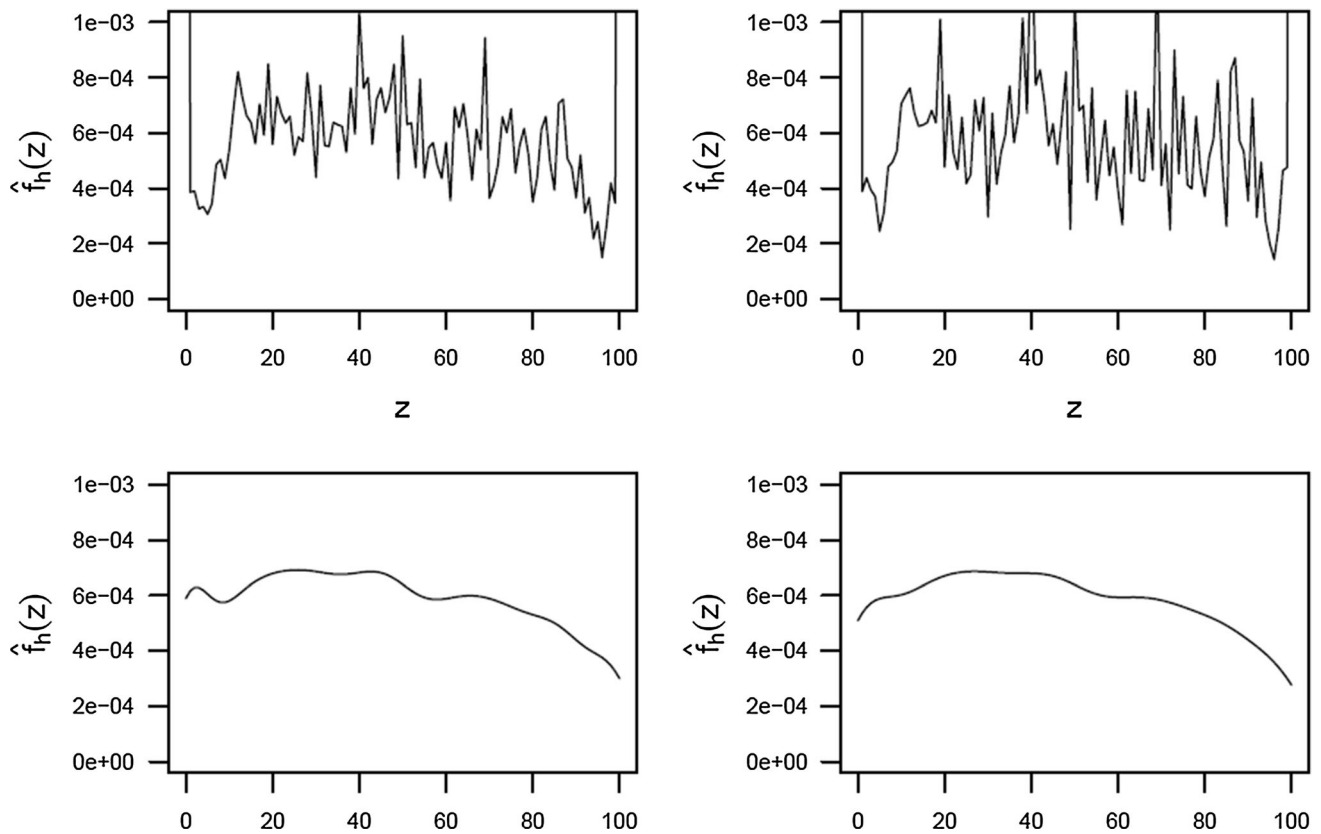
Fig. 8 Estimates of the intensity of wildlife-vehicle collisions as a function of percentage of forest around roads for the estimated bandwidth selectors $\hat{h}_{RT}$ (*top, left panel*), $\hat{h}_{Boot}$ (*top, right panel*), $\hat{h}_{NM}$ (*bottom, left panel*) and $\hat{h}_{Silv}$ (*bottom, right panel*) (see Table 3 for specific values of these bandwidth selectors)

continuous. Hence, using (Federer 1969, Th 3.2.22), we can write

$$G^{\star}(z) = \int_{a}^{b} 1_{\{Z(\alpha(t)) \leq z\}} ||\nabla Z(\alpha(t))||^{-1} ||\nabla Z(\alpha(t))|| \ ||\alpha'(t)|| dt$$
$$= \int_{-\infty}^{z} \int_{(Z \circ \alpha)^{-1}(s)} ||\nabla Z(\alpha(t))||^{-1} ||\alpha'(t)|| dH(t) ds,$$

(10)

with $dH(t)$ the one-dimensional Hausdorff measure.

Differentiation of (10) with respect to z gives

$$g^{\star}(z) = (G^{\star})'(z) = \int_{(Z \circ \alpha)^{-1}(z)} ||\nabla Z(\alpha(t))||^{-1} ||\alpha'(t)|| dH(t).$$

(11)

Additionally, we have the following expression for the intensity function

$$\int_{L} \lambda(u) d_{1}u = \int_{\mathbb{R}} \int_{(Z \circ \alpha)^{-1}(s)} \lambda(\alpha(t)) ||\nabla Z(\alpha(t))||^{-1} ||\alpha'(t)|| dH(t) ds$$
$$= \int_{\mathbb{R}} \rho(s) \int_{(Z \circ \alpha)^{-1}(s)} ||\nabla Z(\alpha(t))||^{-1} ||\alpha'(t)|| dH(t) ds$$
$$= \int_{\mathbb{R}} \rho(s) g^{\star}(s) ds$$

(12)

Finally, as **X** has intensity λ, we have that $E[N(\mathbf{X})] = \int_{L} \lambda(u) d_{1}u$, and for construction we know that $E[N(Z(\mathbf{X}))] = E[N(\mathbf{X})]$. Then, using the above result, it is straightforward to see that

$$E[N(Z(\mathbf{X}))] = E[N(\mathbf{X})] = \int_{L} \lambda(u) d_{1}u = \int_{\mathbb{R}} \rho(s) g^{\star}(s) ds.$$

Thus, we have shown that $Z(\mathbf{X})$ is a point process in $\mathbb{R}$ with intensity function $\rho g^{\star}$. The last statement of the theorem about the inheritance of the Poisson property is trivial by construction, because the expected number of points in the transformed process is exactly the same as the expected number in the original one.
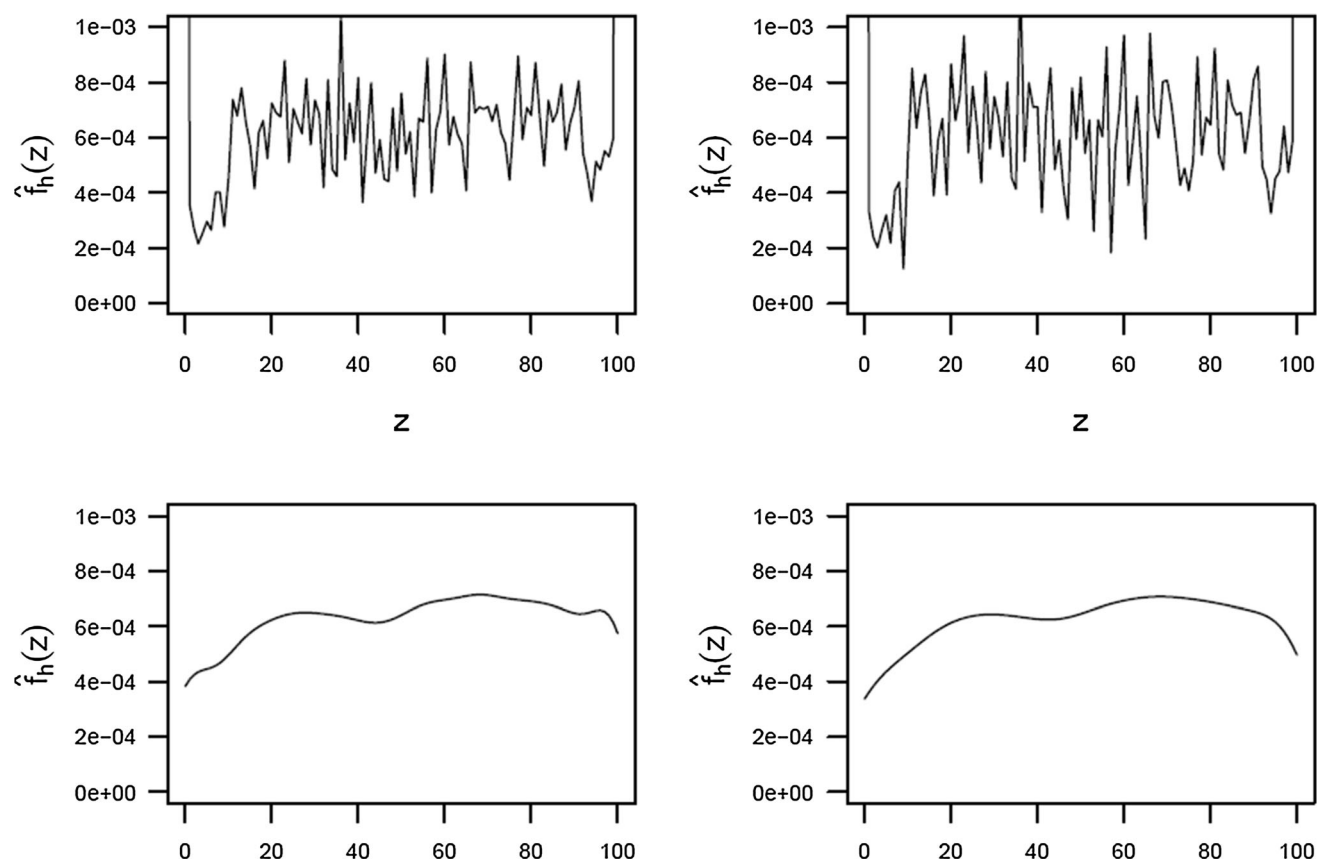
Fig. 9 Estimates of the intensity of wildlife-vehicle collisions as a function of percentage of crop fields around roads for the estimated bandwidth selectors $\hat{h}_{RT}$ (*top, left panel*), $\hat{h}_{Boot}$ (*top, right panel*), $\hat{h}_{NM}$ (*bottom, left panel*) and $\hat{h}_{Silv}$ (*bottom, right panel*) (see Table 3 for specific values of these bandwidth selectors)

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Acevedo P, Quirós-Fernández F, Casal J, Vicente J (2014) Spatial distribution of wild boar population abundance: Basic information for spatial epidemiology and wildlife management. Ecol Ind 36:594–600

Ang QW, Baddeley A, Nair G (2012) Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. Scand J Stat 39:591–617

Anuario Estadistica DGT (2018) Anuario estadistica de la Dirección General de tráfico (DGT) para el año 2018. Ministerio del Interior, Gobierno de España

Baddeley A, Chang YM, Song Y, Turner R (2012) Nonparametric estimation of the dependence of a spatial point process on spatial covariates. Stat Interface 5:221–236

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. CRC Press, Amsterdam

Barr CD, Schoenberg FP (2010) On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. Biometrika 97:977–984

Bithell JF (1990) An application of density estimation to geographical epidemiology. Stat Med 9:691–701

Borrajo MI, González-Manteiga W, Martínez-Miranda MD (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes with spatial covariates. Comput Stat Data Anal 144

Borruso G (2005) Network density estimation: analysis of point patterns over a network. In: International conference on

computational science and its applications. Springer, Berlin, pp 126–132

Borruso G (2008) Network density estimation: a GIS approach for analysing point patterns in a network space. Trans GIS 12:377–402

Bruinderink GG, Hazebroek E (1996) Ungulate traffic collision in Europe. Conserv Biol 26:1059–1067

Coffin AW (2007) From roadkill to road ecology: a review of the ecological effects of roads. J Transp Geogr 15:396–406

Colino-Rabanal VJ, Peris SJ (2016) Wildlife roadkills: improving knowledge about ungulate distributions? Hystrix 27:91–98

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

Cucala L (2006) Espacements bidimensionnels et données entachés d'erreurs dans l'analyse des procesus ponctuels spatiaux. PhD thesis, Université des Sciences de Toulouse I

Daley DJ, Vere-Jones D ((1988) An introduction to the theory of point processes. Springer, Berlin

Deckers B, Verheyen K, Hermy M, Muys B (2005) Effects of landscape structure on the invasive spread of black cherry (Prunus serotina) in an agricultural landscape in Flanders, Belgium. Ecography 28:99–109

Díaz-Varela ER, Vázquez-Gonzalez I, Marey-Pérez MF (2011) Assessing methods of mitigating wildlife-vehicle collisions by accident characterization and spatial analysis. Transp Res Part D 16:281–287

Diggle PJ (1985) A kernel method for smoothing point process data. J R Stat Soc Ser C (Appl Stat) 34:138–147

Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns. CRC Press, Amsterdam

Federer H (1969) Geometric measure theory. Vol. 153 of Die Grundlehren der mathematischen Wissenschaften, Springer, Berlin

Fuentes-Santos I, González-Manteiga W, Mateu J (2015) Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial poisson point processes. Scand J Stat 43:416–435

Ha H, Shilling F (2018) Modelling potential wildlife-vehicle collisions (WVC) locations using environmental factors and human population density: A case-study from 3 state highways in Central California. Eco Inform 48:212–221

Hegland SJ, Hamre N (2018) Scale-dependent effects of landscape composition and configuration on deer-vehicle collisions and their relevance to mitigation and planning options. Landsc Urban Plan 169:178–184

Hubbard MW, Danielson BJ, Schmitz RA (2000) Factors influencing the location of deer-vehicle accidents in Iowa. J Wildl Manag 64:707–712

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

Lord D, Mannering F (2010) The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp Res Part A Policy and Practice 44(5):291–305

McSwiggan G, Baddeley A, Nair G (2017) Kernel density estimation on a linear network. Scand J Stat 44:324–345

Møller J, Waagepetersen RP (2003) Statistical inference and simulation for spatial point processes. Chapman and Hall/CRC

Moradi MM, Rodríguez-Cortés FJ, Mateu J (2018) On kernel-based intensity estimation of spatial point patterns on linear networks. J Comput Graph Stat 27:302–311

Moradi MM, Cronie O, Rubak E, Lachieze-Rey R, Mateu J, Baddeley A (2019) Resample-smoothing of Voronoi intensity estimators. Stat Comput 29:995–1010

Morelle K, Lehaire F, Lejeun P (2013) Spatio-temporal patterns of wildlife-vehicle collisions in a region with a high-density road network. Nat Conserv 5:53–73

Okabe A, Satoh T, Sugihara K (2009) A kernel density estimation method for networks, its computational method and a GIS-based tool. Int J Geogr Inf Sci 23:7–32

Okabe A, Sugihara K (2012) Spatial analysis along networks: statistical and computational methods. Wiley, New York

Rakshit S, Davies T, Moradi MM, McSwiggan G, Nair G, Mateu J, Baddeley A (2019) Fast Kernel smoothing of point patterns on a large network using two-dimensional convolution. Int Stat Rev 87:531–556

Sáez-de-Santa-María A, Telleria JL (2015) Wildlife-vehicle collisions in Spain. Wildlife Res 61:399–406

Silverman BW (1982) Algorithm AS 176: Kernel density estimation using the fast Fourier transform. J R Stat Soc Ser C (Appl Stat) 31:93–99

Silverman BW (1986) Density estimation for statistics and data analysis. CRC Press, Amsterdam

Spooner P, Lunt I, Okabe A, Shiode S (2004) Spatial analysis of roadside Acacia populations on a road network using the network K-function. Landscape Ecol 19:491–499

Tatewaki T, Kioke F (2018) Synoptic scale mammal density index map based on roadkill records. Ecol Ind 85:468–478

Van Lieshout MNM (2000) Markov point processes and their applications. World Scientific, Singapore

Xie Z, Yan J (2008) Kernel density estimation of traffic accidents in a network space. Comput Environ Urban Syst 2:396–406

Yamada I, Thill JC (2004) Comparison of planar and network K-functions in traffic accident analysis. J Transp Geogr 12:149–158