An adapted intensity estimator for linear networks with an application to modelling anti-social behaviour in an urban environment

M. M. Moradi[1,2,*], F. J. Rodríguez-Cortés[2] and J. Mateu[2]

[1] *Institute of New Imaging Technologies (INIT), University Jaume I, Castellón, Spain; moradi@uji.es*
[2] *Department of Mathematics, University Jaume I, Castellón, Spain; cortesf@uji.es, mateu@uji.es*
[*] *Corresponding author*

Abstract. *We consider an inhomogeneous point process on a linear network, and extend Diggle's non-parametric kernel-based edge-corrected intensity estimator in this context. We compare the accuracy and performance of Diggle's method with respect to the equal-split discontinuous kernel estimator. We show the improvement on the estimation of second-order summary statistics, in particular on the inhomogeneous geometrically corrected network K- and pair correlation functions. Finally, we analyse a spatial point pattern of calls related to anti-social behaviour in the streets of a Spanish town.*

Keywords. *Anti-social behaviour; Intensity; Linear network; Second-order characteristics; Shortest path distance.*

---

1 Point processes on a linear network

In the spatial statistics context, there are numerous real problems such as the location of traffic accidents in a geographical area, or geo-coded locations of anti-social behaviour in the streets of cities that need to restrict the support of the process over networks to face a more realistic scenario. 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 [5]). We consider the events which directly occur on a linear network as a point pattern, that is, a set of locations, irregularly distributed within a linear network and assumed to have been generated by some form of stochastic mechanism. In order to propose statistical methods for analysing point patterns on linear networks, it is necessary to define some fundamental concepts in this new support.

A line segment in the plane with endpoints $\mathbf{u}$ and $\mathbf{v}$ can be written in parametric form as $[\mathbf{u},\mathbf{v}] = \{t\mathbf{u}+(1-t)\mathbf{v} : 0 \leq t \leq 1\}$ with $\mathbf{u},\mathbf{v} \in \mathbb{R}^2$. Following [2], a linear network L can be defined as the union of a finite collection of line segments embedded in a plane, the endpoints of the segments are called nodes, and the degree of a node is the number of segments that share the same node. Denote by $|L|$ the total sum of the lengths of line segments 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 lengths of all paths linking $\mathbf{u}$ and $\mathbf{v}$. The disc of radius $r > 0$ centered at the point $\mathbf{u} \in L$ is given by $b_L(\mathbf{u},r) = \{\mathbf{v} : \mathbf{v} \in L : d_L(\mathbf{u},\mathbf{v}) \leq r\}$ and its relative boundary $\partial b_L(\mathbf{u},r)$ is the set of points lying exactly r units away from $\mathbf{u}$. The circumference $m(\mathbf{u},r)$ is the number of points in the relative boundary $\partial b_L(\mathbf{u},r)$ (for more details see [2]).

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 events $\{\mathbf{u}_i\}_{i=1}^n$ of X within a bounded linear network $L \subset \mathbb{R}^2$, and let $N(L)$ denote the number of events falling in L. We assume that X has intensity function $\lambda(\cdot)$ and pair correlation function $g(\cdot)$ (see [2]). Then, for any Borel measurable real function h on L

$$\mathbb{E}\left[\sum_{\mathbf{u}_i \in X} h(\mathbf{u}_i)\right] = \int_L h(u)\lambda(u)\mathrm{d}_1 u, \tag{1}$$

where $\mathrm{d}_1 u$ denotes one-dimensional integration over the line segment. Generally, $\lambda(\mathbf{u})$ is interpreted as the expected number of events per unit length of linear network in a neighbourhood of $\mathbf{u}$, and under the assumption of homogeneity it can be estimated by the quantity $n/|L|$.

2 Adapted Diggle's method for the intensity on a linear network

In the inhomogeneous case, several kernel smoothing methods for intensity estimation on linear networks have been proposed in the literature (see [5]). These methods are straightforward extensions of the planar case to the linear networks, but without the consideration of important facts such as edge-effects and the degree of the nodes. However, the equal-split kernel function proposed by [4] tries to remedy these problems by dividing the weight assigned by the kernel by a quantity that depends on the degree of the nodes which are equally distributed over the segments that share the node. Nevertheless, this estimation does not seem to provide quite good results when it is used in the estimation of second-order summary statistics. This method has been implemented in the R package spatstat [2].

We extend the non-parametric kernel-based edge-corrected intensity estimator of [3] to the linear network support following the mathematical arguments in [1]. We thus define the estimator of the intensity function as

$$\hat{\lambda}_{\varepsilon}^{(\mathrm{D})}(\mathbf{u}) = \sum_{i=1}^{n} \frac{\kappa_{\varepsilon}(d_L(\mathbf{u}, \mathbf{v}_i))}{\int_0^{\infty} \kappa(t) m(\mathbf{v}_i, t)\mathrm{d}t}, \qquad \mathbf{u} \in L, \tag{2}$$

where κ is a one-dimensional kernel function with smoothing parameter ε. Note that the edge-correction in the denominator of (2) involves the circumference function, and therefore we are also considering the degree of the nodes, while removing the overestimation at the nodes and the edges. We highlight that this adaptation on linear networks of the classical techniques for the intensity estimation are appropriate, and the circumference function is an necessary tool that provides information of the structure of the network on every point.

3 Simulation study

We conduct a simulation experiment to compare the accuracy and performance of adapted Diggle's method with respect to the equal-split discontinuous kernel estimator under the assumption of inhomogeneous Poisson processes. We set a simple linear network composed by 19 vertices, 26 lines with total length 39.38 units, within a window $[0,5] \times [0,5]$ units^2, as it is shown in Figure (1). We consider the intensity function $\lambda(x,y) = 0.2\,e^{0.4(x+y)}$ and use this to simulate 1000 point patterns with expected number of points $\mathbb{E}[N(L)] = 95$. For each one of the patterns we estimate the intensity based on both techniques (adapted Diggle's method and equal-split discontinuous kernel) and compare them through a mean error (ME) approach. Figures (1b) and (1c) show the intensity estimation based on the adapted Diggle's method using the Epanechnikov kernel, and on the equal-split discontinuous method using a Gaussian kernel, respectively, with bandwidth $\varepsilon = 1.25$ units for both cases.

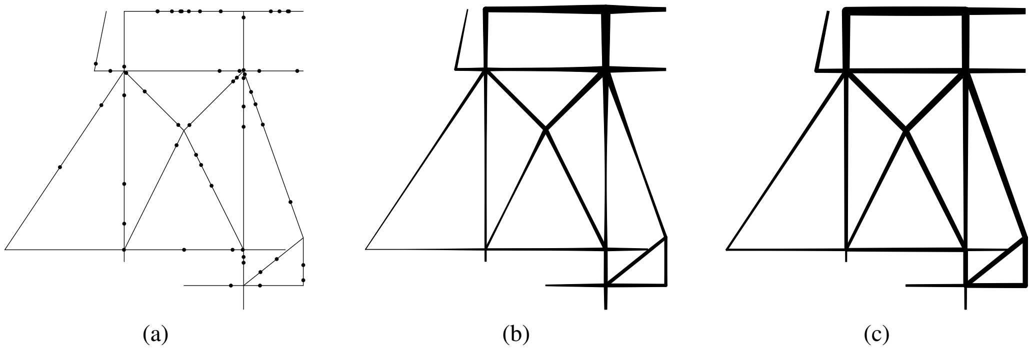

Figure 1: Simulation and smoothing kernel density estimates of an inhomogeneous Poisson process on the simple network: (1a) Realisation of an inhomogeneous pattern on a simple network. (1b) Line-thickness display for adapted Diggle's estimator. (1c) Line-thickness display for the equal-split discontinuous method.

From a visual comparison in Figure (1a) we note that the adapted Diggle's method improves the estimation particularly at those segments with a low number of points because it does not overestimate the intensity at these segments, while the equal-split discontinuous method shows a constant high value along the segments regardless the distribution of the points. Furthermore, adapted Diggle's estimator proved to be robust showing small variations against smooth variations in the bandwidths. Figure (2) shows that the ME for the adapted Diggle's estimator is around 0, whereas that for the equal-split discontinuous method is around -2.5, indicating that the former estimator shows a smaller error than the latter with respect to the theoretical intensity.

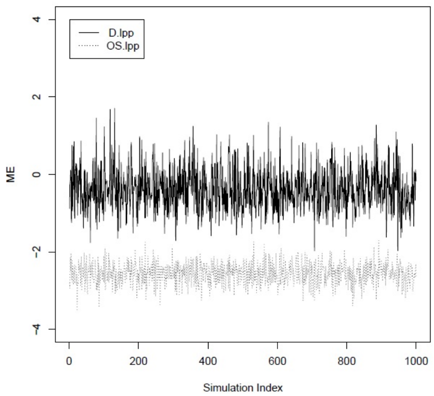

Figure 2: Mean error (ME) for 1000 simulations of an inhomogeneous Poisson point process on a simple linear network.

4 Anti-social behaviour analysis

We analyse the spatial point pattern of calls related to anti-social behaviour in the streets of Castellón, Spain in January 2013. The city of Castellón is divided into 108 census tracks with an overall surface of 108.659 km^2. The linear network associated with this city contains 450 nodes and 2242 line segments, with a total length of 1611.688 km, and it is shown in Figure 3. The dataset is formed by 189 locations of calls related to anti-social behaviour with a circumradius of 161.168 km.

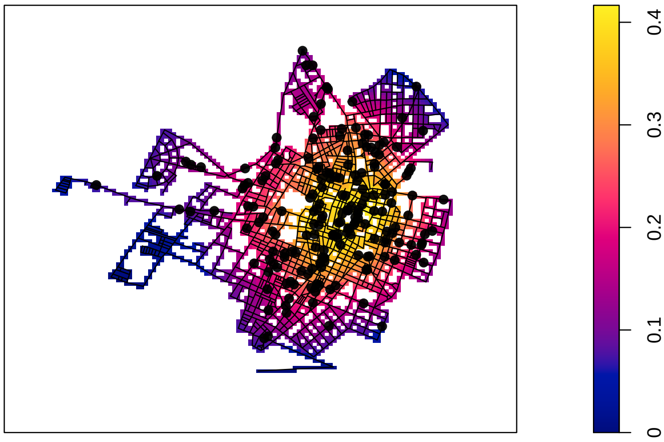

Figure 3: Estimated intensity for anti-social behaviour on the streets of Castellón, Spain in January 2013 with $\varepsilon = 1.333$ km.

The pattern in Figure 3 looks quite regular and does not exhibit any apparent clustering. However, the streets leading to the city center show higher intensity than those in the boundary which may be related to night-life and discotheques. Second-order summary statistics on the linear network will be further used to support these findings.

Acknowledgments. The research of M.M. Moradi is supported by funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, http://www.geo-c.eu/). F.J. Rodríguez-Cortés and J. Mateu are supported by Grant MTM2013-43917-P from the Spanish Ministry of Science and Education, and Grant P1-1B2015-40.

References

[1] 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.

[2] Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with* R. Chapman and Hall/CRC Press, London.

[3] Diggle, P. J. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **34** 138–147.

[4] 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.

[5] Okabe, A., and Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods*. John Wiley & Sons, London.