Nearest neighbour clutter removal for estimating features in point process on linear networks

Juan F. Díaz-Sepúlveda*1, Nicoletta D'Angelo2, Giada Adelfio2, Jonatan A. González3 and Francisco J. Rodríguez-Cortés1

1Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia
2Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy
3Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Abstract

We consider the problem of features detection in the presence of clutter in point processes on a linear network. For the purely spatial case, previous studies addressed the issue of nearest-neighbour clutter removal. We extend this classification methodology to a more complex geometric context, where the classical properties of a point process change and data visualization is not intuitive. As a result, the method is suitable for a feature with clutter as two superimposed Poisson processes on the same linear network, without assumptions about the feature shapes. We present simulations and examples of road traffic accidents that resulted in injuries or deaths in two cities of Colombia to illustrate the method.

Keywords– Clutter, EM-Algorithm, Feature, Kth nearest-neighbor, Linear network, Spatial point pattern.

1 Introduction

In the last decade, spatial statistics has undergone an extraordinary methodological and computational advancement focused on generalising and extending the foundations of the theories to more complex geometric spaces that allow fairer statistical analysis of new kinds of spatial data. In particular, spatial point process methodologies have been addressed to consider non-classical geometric supports for analysing events on linear networks, such as

*Corresponding author: jfdiazs0@unal.edu.co (Juan F. Díaz-Sepúlveda).
traffic accidents in a city (Baddeley et al., 2021a). One of the most relevant contributions to the statistical analysis of network data is the definition of the geometrically corrected first- and second-order characteristics. These descriptors can be used to identify spatial configurations and discriminate between point patterns according to their aggregation structure (Ang et al., 2012; McSwiggan et al., 2017; Moradi et al., 2018; Rakshit et al., 2019a,b; D’Angelo et al., 2021, 2022), select or reduce a set of variables (Rakshit et al., 2021) and estimate relative risks (McSwiggan et al., 2020).

One of the most frequent research fields for spatial data analysis is identifying features (clusters) of events in the presence of clutter. For instance, detecting surface minefields from an image from a reconnaissance aircraft can be processed to obtain a list of objects, some of which may be mines and others any other type of object (Allard and Fraley, 1997; Byers and Raftery, 1998). For spatial point processes, the problem has been addressed in different manners. Allard and Fraley (1997) developed a method to find the maximum likelihood solution using Voronoi polygons. Dasgupta and Raftery (1998) used model-based clustering to extend the methodology proposed by Banfield and Raftery (1993). While these methods are based on some limiting assumptions, Byers and Raftery (1998) adopted a different approach in which they estimated and removed the clutter without making any assumptions about the shape or number of features.

However, when minefields are distributed over the streets of a city or a neighbourhood, i.e., over a linear network, this problem cannot be analysed similarly because the events do not occur in the whole plane. Indeed, Baddeley et al. (2021a) states that statistical analysis of point data on linear networks presents severe problems because the associated graph is not spatially homogeneous, creating geometric and computational complexities and leading to new problems with a high risk of methodological error. In addition, data on networks can have a wide range of spatial scales. Baddeley et al. (2021a) emphasise that the point processes on linear networks also challenge the classical methodology of spatial statistics based on stationary processes, which is mainly inapplicable to data in networks. Therefore, we extend the spatial technique proposed by Byers and Raftery (1998) from the planar case to the linear network’s one. We use the observed $K$-th nearest neighbour distances modelled as a mixture distribution and estimate the parameters by an $EM$ algorithm to classify data points as clutter or otherwise.

The structure of the paper is as follows. Section 2 gives some basic techniques for analysing point patterns on linear networks. Section 3 presents the proposed method for feature detection on linear networks. Section 4 shows an extended simulation study carried out with different linear networks. Section 5 shows two applications of the classification method for traffic accidents in Bogota and Medellin (Colombia). Section 6 presents some conclusions.
2 Mathematical framework

Following Ang et al. (2012); Baddeley et al. (2021a), we consider a linear network as the union of a finite number of line segments on the plane, namely \( L = \bigcup_{i=1}^{n} l_i \), such that each \( l_i = [u_i, v_i] = \{ w : w = tu_i + (1-t)v_i, 0 \leq t \leq 1 \} \), and \( u_i, v_i \) are the endpoints of the segment \( l_i \). The total length of the network \( L \) is denoted by \( |L| \). Additionally, a path between two points \( u \) and \( v \) in a linear network \( L \) is a sequence \( x_0, x_1, \ldots, x_m \) of points in \( L \) so that \( x_0 = u \), \( x_m = v \) and \( [x_i, x_{i+1}] \subset L \) for each \( i = 0, \ldots, m - 1 \). The length of the path is the sum of its segments. The shortest path distance \( d_L(u, v) \) between \( u \) and \( v \) in \( L \) is the minimum of the lengths of all paths from \( u \) to \( v \). If there are no paths from \( u \) to \( v \) (implying that the network is not connected), then the distance is defined by \( d_L(u, v) = \infty \).

We also assume that the disc of radius \( r > 0 \) and centre point \( u \in L \) is the set of all points \( v \) in the network that lie no more than a distance \( r \) from \( u \), that is \( b_L(u, r) = \{ v \in L : d_L(u, v) \leq r \} \). The relative boundary of the disc is the set of points lying exactly \( r \) units away from \( u \), \( \partial b_L(u, r) = \{ v \in L : d_L(u, v) = r \} \). The circumference \( m(u, r) \) is the number of points of \( L \) there is in \( \partial b_L(u, r) \). The circumference is finite for all \( r < \infty \), and we set \( m(u, \infty) = \infty \) by convention. The circumradius of the network is the radius of the smallest disc that contains the entire network, \( R = R(L) = \inf \{ r : m(u, r) = 0 \}, \) for some \( u \in L \) = \( \min_{u \in L} \max_{v \in L} d_L(u, v) \). If \( L \) is not path-connected, then \( R \) is the minimum of the circumradius of the connected components of \( L \); for more details see Ang et al. (2012); Baddeley et al. (2021a).

We define a (finite, simple) point pattern \( X \) on \( L \) as a finite set \( X = \{ x_1, x_2, \ldots, x_n \} \) of distinct points \( x_i \in L \), where \( n \geq 0 \) (see e.g., Rakshit et al. 2021). Further, \( N_X(B) = N(X \cap B) \) is the number of points on \( X \) lying in \( B \) for any set \( B \subset L \). Thus, for a finite and simple point possess with intensity function \( \lambda(u) \), \( u \in L \), we have \( \mathbb{E}[N_X(B)] = \Lambda(B) = \int_B \lambda(u) du \), for all measurable \( B \) in \( L \), where \( d_1 \) denotes integration with respect to arc length (Federer 1969). If \( \lambda \) is constant, then \( X \) is called homogeneous; otherwise, \( X \) is called inhomogeneous. \( \lambda \) is the mean number of points per unit length for a homogeneous point process. A Poisson point process on \( L \) with intensity \( \lambda > 0 \) is characterized by the properties that, for any line segment \( B \subset L \), the number of points falling in \( B \) has a Poisson distribution with mean \( \lambda |B| \), while events occurring in disjoint line segments \( B_1, \ldots, B_m \subset L \) are independent (Ang et al. 2012). Finally, a sub-network of \( L \) is a linear network which is a subset of \( L \).

3 Approach for clutter removal on linear networks

One of the main interests of spatial point pattern analysis is identifying features surrounded by clutter. Unlike the planar case, which is primarily straightforward, identifying features on networks through visual inspection is far from direct since the complexity of this domain. We consider the extension of the approach in Byers and Raftery (1998) and assume that the clutter is distributed as a homogeneous Poisson point process on a linear network. The
features are also homogeneous Poisson point processes but with different intensities, restricted to a certain sub-network and overlaid on the clutter. Therefore, the resulting point process is Poisson with piece-wise constant intensity on the linear network, and there are no assumptions about the features’ shape or densities.

![Feature sub-network of Chicago.](image)

![Clutter point pattern.](image)

![Feature point pattern.](image)

![Superposition of clutter and feature.](image)

Figure 1: Linear network **chicago** at four different stages. A clutter point pattern simulated from a homogeneous Poisson point process with $\lambda_c = 0.064$ and $\mathbb{E}[n_c] = 2000$. The feature is an observation of a homogeneous Poisson point process with $\lambda_f = 0.017$ and $\mathbb{E}[n_f] = 50$.

In order to illustrate the geometric context for our developments, we consider the linear network **chicago** of the package **spatstat.data** (Baddeley et al. 2021b), which represents the street network of some area of Chicago (Illinois, USA) close to the University of Chicago; a particular case of simulated data on this linear network is shown in Figure 1 at four different stages (Ang et al. 2012). The sub-network where the features are generated is shown in Figure 1a. The street network has 338 intersections, 503 uninterrupted segments and a total length of 31150 feet. The sub-network has 39 intersections, 53 uninterrupted segments and
a total length of 2991 feet. Figure 1b shows the clutter point pattern simulated from a homogeneous Poisson point process with intensity \( \lambda_c = 0.013 \) and expected number of points \( \mathbb{E}[n_c] = 2000 \) across the entire linear network. Figure 1c shows the feature point pattern simulated from a Poisson process with intensity \( \lambda_f = 0.068 \) and expected number of points \( \mathbb{E}[n_f] = 50 \) in the sub-network. The superposition of the clutter and feature point patterns in Figures 1b and 1c are displayed in Figure 1d which is the resulting point pattern.

3.1 K-th nearest neighbor distribution of distances

In the linear networks geometry, the disc \( b_L(u,r) \) has variable volume for a given \( r \) and different locations \( u \) (i.e. the disc volume depends on \( u \)), as mentioned in Cronie et al. (2020). Therefore, for \( u \in L \) and \( 0 < x < R \), there is an \( r_u > 0 \) such that \( |b_L(u,r_u)| = x \). Let \( D_K^L \) be the distance of the \( K \)-th nearest neighbour of \( u \); if \( D_K^L \) is greater than \( r_u \), then there must be one of the \( 0,1,\ldots,K-1 \) points at a distance less than \( r_u \). Moreover, the shortest path length from \( u \) to any other point on disc \( b_L(u,r_u) \) is less or equal to the disc volume \( x \), then \( r_u \leq x \). Furthermore, the largest possible radius for a disc of volume \( x \) in a linear network is \( x \). Therefore, for all \( u \in L \) and \( x \in [0,\infty) \), the \( K \)-th nearest neighbour distribution approximation in the linear network is given by

\[
\mathbb{P}(D_K^L \geq x) = \sum_{j=0}^{K-1} \frac{e^{-\lambda x} (\lambda x)^j}{j!} = 1 - F_{D_K^L}(x),
\]

where \( \mathbb{P}(D_K^L \geq x) \) is the probability that the \( K \)-th nearest neighbour point falls out of \( b_L(u,x) \) with \( |b_L(u,x)| = x \), assuming that this disk exists around \( u \). If the \( K \)-th nearest neighbour point of \( u \) is outside \( b_L(u,x) \), it is also outside \( b_L(u,r_u) \).

Accordingly, the density \( f_{D_K^L}(x) \) can be developed as

\[
f_{D_K^L}(x) = \frac{d}{dx} F_{D_K^L}(x) = \frac{d}{dx} \left[ 1 - \sum_{j=0}^{K-1} \frac{e^{-\lambda x} (\lambda x)^j}{j!} \right] = -\lambda e^{-\lambda x} \sum_{j=0}^{K-1} \frac{(\lambda x)^j}{(j-1)!} - \frac{(\lambda x)^j}{j!}
\]

\[
= -\lambda e^{-\lambda x} \left[ (\lambda x)^{K-1} (K-1)! - (\lambda x)^K (K-1)! \right] = \frac{\lambda (\lambda x)^{K-1} e^{-\lambda x}}{\Gamma(K)}
\]

and therefore \( D_K^L \sim \Gamma(K, \lambda) \). Having a closed-form and the Gamma distribution properties, the maximum likelihood estimation of the rate given the observed values of \( D_K^L \) is straightforward. Indeed, the maximum likelihood estimate of \( \lambda \) is

\[
\hat{\lambda} = \frac{K}{\sum_{i=1}^{n} d_i},
\]

where \( d_i \) is the \( i \)-th observed \( K \)-th nearest neighbour distance.
3.2 Mixture modeling and estimation

Six estimated distance distributions $D^L_K$ for the mixture of a couple of homogeneous Poisson point patterns simulated in the Chicago network (see Figure 1) from the 27-th to the 32-nd nearest neighbour are displayed as histograms in Figure 2, where a clear bimodality is evident. They can be used as exploratory tools and motivation to propose a model (Byers and Raftery, 1998). We assume two types of processes to be classified through a mixture of the corresponding $K$-th nearest neighbour distances coming from the clutter and feature, that is, two superimposed Poisson processes. Therefore, based on equation (1), we assume that

$$D^L_K \sim p \Gamma(K, \lambda_1) + (1 - p) \Gamma(K, \lambda_2),$$  \hspace{1cm} (2)

where $\lambda_1$ and $\lambda_2$ are the intensities of the two homogeneous Poisson point processes (clutter and feature, respectively) on the linear network, and $p$ is the constant that characterizes the postulated distribution of the distances $D^L_K$.

![Figure 2: Distributions of the distances $D^L_K$ for the mixture of clutter and feature from the 27-th to the 32-nd nearest neighbours simulated on the Chicago network in Figure 1.](image_url)

The parameters $\lambda_1$, $\lambda_2$ and $p$ associated with the mixture are estimated using an EM algorithm (Dempster et al., 1977), wherein we use the closed-form of a Gamma distribution in the expectation step. On the other hand, let $\delta_i \in \{0, 1\}$ be the two classification components.
for each data point, where $\delta_i = 1$ if the $i$-th point belongs to the feature and $\delta_i = 0$ otherwise. Thus each data point has an observed distance $d_i$ of $D^L_K$ and an unknown classification component $\delta_i$. Hence, the $E$ step of the algorithm consists of

$$
\mathbb{E}[\hat{\delta}_i^{(t+1)}] = \frac{\hat{p}^{(t)} f_{D^L_K}(d_i; \hat{\lambda}_1^{(t)})}{\hat{p}^{(t)} f_{D^L_K}(d_i; \hat{\lambda}_1^{(t)}) + (1 - \hat{p}^{(t)}) f_{D^L_K}(d_i; \hat{\lambda}_2^{(t)})},
$$

and the maximization $M$ step consists of

$$
\hat{\lambda}_1^{(t+1)} = \frac{K \sum_{i=1}^{n} \hat{\delta}_i^{(t+1)}}{\sum_{i=1}^{n} d_i \hat{\delta}_i^{(t+1)}}, \quad \hat{\lambda}_2^{(t+1)} = \frac{K \sum_{i=1}^{n} (1 - \hat{\delta}_i^{(t+1)})}{\sum_{i=1}^{n} d_i (1 - \hat{\delta}_i^{(t+1)})} \quad \text{and} \quad \hat{p}^{(t+1)} = \frac{\sum_{i=1}^{n} \hat{\delta}_i^{(t+1)}}{n}.
$$

We follow an intuitive classification test criterion which classifies the points according to the mixture component where the distances have the highest densities. We are mainly interested in identifying the feature points in this proposed classification approach; consequently, we do not consider the edge effect because feature points are predominantly far from the edges in practice. Additionally, for large $n$, the convergence of the $EM$ algorithm is good since it takes less time to arrive at an approximately acceptable solution, and it does so with the fewest number of iterations.

### 3.3 Choosing a $K$-th nearest neighbor

The development mentioned above assumes a proper value of $K$ priorly chosen. The natural way to choose the suitable $K$-th neighbour is by analysing several increasing values of $K$ and then selecting that $K$, after which no improvement is found. However, in the literature, there are several methodological proposals for this target; in this work, we use an entropy-type measure of separation introduced in Celeux and Soromenho (1996) given by

$$
S = -\sum_{i=1}^{n} \delta_i \log_2(\delta_i),
$$

where $\delta_i$ are the probabilities of being in the first component of the mixture in equation (2), that is, the feature. As stated by Byers and Raftery (1998), plotting the entropies sequentially and looking for a levelling-off changepoint in the graph is an easy way to choose $K$. An example of this procedure is shown in Figure 3, where the classification entropies for values of $K$ up to 30 are plotted for the simulated point pattern on the chicago network in Figure 1.

The optimal $K$ can be automatically selected by fitting a segmented regression model as

$$
\mathbb{E}[Y|x_i] = \beta + \delta(x_i - \psi)I(x_i > \psi),
$$

where our interest is estimating a unique changepoint $\psi$, after which the slope $\beta + \delta$ is constrained to be equal to zero. As depicted in Figure 3, the observed response variable
\( Y \) is the entropy level, modelled as a function of the number of nearest neighbours \( x \). We implemented this automatic option using the function \texttt{segmented} of the package \texttt{segmented} (Muggeo, 2008). We refer to Muggeo (2003) for inferential details on segmented regression models.

The following steps implement the classification procedure:

1. Choose a value of \( K \) either by imputing a sought value or automatically applying a segmented regression model. Note that an upper bound for the \( K \) domain should be fixed manually to a reasonable maximum number of neighbors.

2. Find the \( K \)-th nearest neighbours distance for each point in the point pattern using the shortest path distances computed on the given linear network.

3. Apply the \( EM \) algorithm for estimating \( \lambda_1, \lambda_2, \) and \( p \).

4. Classify the points according to whether they have a higher density under the feature or clutter component of the mixture.

![Figure 3: Classification entropies for \( K \) values up to 30 for the simulated point pattern on the \texttt{chicago} network in Figure \[1\]. The black line represents the observed entropies, and the dotted line represents the estimated segmented model. The vertical line indicates the estimated changepoint with \( \hat{K} = 4 \).](image)

4 Simulation study

This section aims to study the proposed method’s performance in terms of classification rates, considering different scenarios and feature shapes. We generate clutter in the whole network...
in all scenarios, and the feature is simulated constrained to a sub-network. In order to explore the classification procedure for different networks, we use both the chicago and the dendrite networks (Baddeley et al., 2021b) in the simulation study, which represent linear networks with and without cycles. We use the chicago network in the third scenario, taking two nested sub-networks, generating clutter with two different clutter intensities. The fourth network is the Antonio Narino network (Moncada, 2018) which has a homogeneous structure, and we simulate features in two disjoint sub-networks, corresponding to two separate roads.

We show the results of the classification method in terms of true-positive rate (TPR), false-positive rate (FPR), and accuracy (ACC), averaging over 100 simulated point patterns generated with $E[n_c]$ and $E[n_f]$ expected number of points for clutter and feature, respectively. In addition, the $\lambda_c$ and $\lambda_f$ intensities are reported for clutter and feature. The accuracy is the proportion of correct predictions (both true-positives and true-negatives) among the total number of cases examined. We further compare the detection with $K$ equal to $\{5, 10\}$, and also the automatically selected by the segmented regression model considered in Section 3.3. The rates are computed accordingly.

4.1 Estimating feature in presence of clutter and cycles

The linear network chicago described in Section 3 is a network with loops given with edges connecting vertexes to themselves. As mentioned in that section, we generate feature and clutter point patterns as displayed in Figure 1 and report the outcomes in Table 1.

From Table 1 for stages 1 and 2, where $\lambda_c < \lambda_f$, it is possible to notice that the classification method performs remarkably well in terms of true- and false-positive rates for $K = 5$, $K = 10$ and $\hat{K}$. For stages 3 and 4, where $\lambda_c \approx \lambda_f$, the classification method performs well in true- and false-positive rates for $K = 5$, $K = 10$ and $\hat{K}$. However, stage 4 with $K = 5$ shows remarkably lower performance than $\hat{K}$. In stages 5 and 6, where $\lambda_c > \lambda_f$, the classification method again works well for $\hat{K}$, nevertheless in $K = 5$ and $K = 10$, both have lower performance than $\hat{K}$. In general, the best performance is achieved when the expected number of feature points approaches the expected number of clutter points when their ratio comes close to 1. The worst performance is when this ratio is close to 0. In the first case, there is little clutter compared to the number of features, and the classification method quickly detects the features as in stage 1; whereas in the second case, the amount of clutter is too large, and the features are more hidden to be detected, as in stage 6. The estimated $\hat{K}$ selected by minimizing the entropy level defined in equation (3) improves the classification rates. With $K = 5$ and $K = 10$, the classification method performs worse, i.e., when the expected number of feature points moves away from the expected number of clutter points, or their ratio is close to 0.
Table 1: Classification rates averaged over 100 simulated point patterns generated on the Chicago linear network with $\lambda_c$ and $\lambda_f$ intensities and $E[n_c]$ and $E[n_f]$ expected number of points for clutter and feature.

| Stage | $\lambda_c$ | $\lambda_f$ | $E[n_c]$ | $E[n_f]$ | Rate | $K$ | $5$ | $10$ | $\hat{K}$ |
|-------|-------------|-------------|----------|----------|------|-----|-----|-----|-------|
| 1     | 0.032       | 0.100       | 1000     | 300      | TPR  | 0.996 | 0.998 | 0.997 |
|       |             |             |          |          | FPR  | 0.001 | 0.001 | 0.001 |
|       |             |             |          |          | ACC  | 0.642 | 0.638 | 0.651 |
| 2     | 0.032       | 0.067       | 1000     | 200      | TPR  | 0.986 | 0.994 | 0.991 |
|       |             |             |          |          | FPR  | 0.003 | 0.001 | 0.002 |
|       |             |             |          |          | ACC  | 0.590 | 0.546 | 0.568 |
| 3     | 0.032       | 0.033       | 1000     | 100      | TPR  | 0.927 | 0.979 | 0.992 |
|       |             |             |          |          | FPR  | 0.007 | 0.002 | 0.001 |
|       |             |             |          |          | ACC  | 0.516 | 0.409 | 0.337 |
| 4     | 0.016       | 0.017       | 500      | 50       | TPR  | 0.898 | 0.979 | 0.983 |
|       |             |             |          |          | FPR  | 0.010 | 0.002 | 0.002 |
|       |             |             |          |          | ACC  | 0.538 | 0.405 | 0.325 |
| 5     | 0.064       | 0.017       | 2000     | 50       | TPR  | 0.749 | 0.865 | 0.959 |
|       |             |             |          |          | FPR  | 0.006 | 0.003 | 0.001 |
|       |             |             |          |          | ACC  | 0.445 | 0.377 | 0.239 |
| 6     | 0.128       | 0.017       | 4000     | 50       | TPR  | 0.684 | 0.765 | 0.846 |
|       |             |             |          |          | FPR  | 0.004 | 0.003 | 0.002 |
|       |             |             |          |          | ACC  | 0.425 | 0.379 | 0.312 |

4.2 Estimating feature in presence of clutter and without cycles

The linear network dendrite is the dendrite network of a neuron (nerve cell) recorded by the Kosik Lab (UCSB) [Jammalamadaka et al., 2013] and available in the package spatstat.data [Baddeley et al., 2021b]. The dendrite network has 640 intersections, 639 uninterrupted segments and a total length of 1934 $\mu$m (1.93 mm) [Baddeley et al., 2014], while the sub-network has 245 intersections, 243 uninterrupted segments and a total length of 778 $\mu$m (0.78 mm). We generate feature and clutter points on the network branches, as it is displayed in Figure 4. We report the results of the classification procedure in Table 2; the outcomes line up with the previous experiments for networks with loops. For stages 1 and 2, where $\lambda_c < \lambda_f$, and for stages 3 and 4, where $\lambda_c \approx \lambda_f$, it is possible to notice that the classification method performs well for $K = 5$, $K = 10$ and $\hat{K}$. In stages 5 and 6, where $\lambda_c > \lambda_f$, the classification method performs acceptably for $\hat{K}$; however, $K = 5$ and $K = 10$ have lower performances than $\hat{K}$. Stage 6, with $K = 5$ shows lower performance than $\hat{K}$. The estimated $\hat{K}$ performs better when the ratio between the expected number of features and the expected number of clutter approaches zero.
Figure 4: Simulated feature and clutter point patterns from a homogeneous Poisson point processes with $\lambda_c = 0.259$ and $\mathbb{E}[n_c] = 500$. The feature point pattern is simulated from a homogeneous Poisson point process with $\lambda_f = 0.193$ and $\mathbb{E}[n_f] = 150$.

| Stage | $\lambda_c$ | $\lambda_f$ | $\mathbb{E}[n_c]$ | $\mathbb{E}[n_f]$ | Rate   | $K$       | $5$ | $10$ | $\hat{K}$ |
|-------|-------------|-------------|------------------|------------------|--------|----------|-----|-----|----------|
| 1     | 0.207       | 0.386       | 400              | 300              | TPR    | 0.967    | 0.970| 0.971|          |
|       |             |             |                  |                  | FPR    | 0.025    | 0.023| 0.022|          |
|       |             |             |                  |                  | ACC    | 0.641    | 0.666| 0.660|          |
| 2     | 0.207       | 0.257       | 400              | 200              | TPR    | 0.934    | 0.942| 0.944|          |
|       |             |             |                  |                  | FPR    | 0.033    | 0.029| 0.028|          |
|       |             |             |                  |                  | ACC    | 0.549    | 0.570| 0.566|          |
| 3     | 0.388       | 0.386       | 750              | 300              | TPR    | 0.922    | 0.947| 0.945|          |
|       |             |             |                  |                  | FPR    | 0.031    | 0.021| 0.022|          |
|       |             |             |                  |                  | ACC    | 0.513    | 0.524| 0.529|          |
| 4     | 0.259       | 0.257       | 500              | 200              | TPR    | 0.906    | 0.928| 0.929|          |
|       |             |             |                  |                  | FPR    | 0.037    | 0.029| 0.028|          |
|       |             |             |                  |                  | ACC    | 0.511    | 0.523| 0.521|          |
| 5     | 0.259       | 0.193       | 500              | 150              | TPR    | 0.867    | 0.902| 0.911|          |
|       |             |             |                  |                  | FPR    | 0.040    | 0.029| 0.027|          |
|       |             |             |                  |                  | ACC    | 0.475    | 0.473| 0.468|          |
| 6     | 0.388       | 0.032       | 750              | 25               | TPR    | 0.663    | 0.715| 0.762|          |
|       |             |             |                  |                  | FPR    | 0.012    | 0.010| 0.008|          |
|       |             |             |                  |                  | ACC    | 0.377    | 0.331| 0.302|          |

Table 2: Classification rates averaged over 100 simulated point patterns generated on the dendrite linear network with $\lambda_c$ and $\lambda_f$ intensities and $\mathbb{E}[n_c]$ and $\mathbb{E}[n_f]$ expected number of points for clutter and feature.
4.3 Estimating feature in presence of two nested clutters

We consider an even more complex scenario where the features could be somehow hidden by different intensities found in the network. Therefore, to show the procedure’s usefulness, we simulate from a more complex setting dealing with the mixture of multiple intensities of a homogeneous Poisson process on the observed point pattern. We use again the chicago linear network and the feature sub-network defined in section 4.1. In this case, we consider a third intensity value in a larger nested sub-network than the feature sub-networks as shown in Figure 5; the largest nested sub-network of the chicago has 130 intersections, 199 uninterrupted segments and a total length of 11731 feet. Hence, we simulate two distinct Poisson point patterns with constant intensities in the two nested sub-networks and superimpose them to a point pattern generated on the whole network. In other words, we create a second clutter in the larger sub-network and the features points in the smaller one.

![Figure 5: Simulated feature point pattern between two nested clutter point patterns.](image)

Table 3 shows the results of the feature classification procedure for the chicago network with two nested clutter patterns. As in the previous settings, the results are computed for different combinations of the intensities \( \lambda_{c_1}, \lambda_{c_2} \) and \( \lambda_f \). In stages 1 and 2, we take \( \lambda_{c_1} \leq \lambda_{c_2} < \lambda_f \). In stages 3 and 4, we take \( \lambda_{c_1} < \lambda_{c_2} > \lambda_f \) for \( \lambda_{c_1} < \lambda_{c_2} \) in stage 3 and \( \lambda_{c_1} > \lambda_f \) for stage 4. In stages 5 and 6, we take \( \lambda_{c_1} \approx \lambda_{c_2} \approx \lambda_f \). In stages 7 and 8, we take \( \lambda_{c_1} > \lambda_{c_2} < \lambda_f \) for \( \lambda_{c_1} < \lambda_f \) in stage 7 and with \( \lambda_{c_1} > \lambda_f \) in stage 8. In stages 9 and 10, we
take $\lambda_{c_1} > \lambda_{c_2} > \lambda_f$. Looking at the classification rate in Table 3 we can see a good method performance for stages 1 to 7 in terms of true- and false-positive rates for $K = 5$, $K = 10$ and $\hat{K}$. In stages 8, 9 and 10, the classification method performs well in terms of true- and false-positive rates for $\hat{K}$, but if $K = 5$ and $K = 10$, the method performs worse. Note that in those scenarios with lower TPR rates, the algorithm performs better with $\hat{K}$ than $K = 5$ and $K = 10$.

| Stage | $\lambda_c$ | $\lambda_{c_2}$ | $\lambda_f$ | $E[n_{c_1}]$ | $E[n_{c_2}]$ | $E[n_f]$ | Rate | $K$ |
|-------|-------------|-----------------|-------------|-------------|-------------|--------|------|-----|
| 1     | 0.032       | 0.051           | 0.067       | 1000        | 600         | 200    | TPR  | 0.997 | 0.999 | 0.999 |
|       |             |                 |             |             |             |        | FPR  | 0.000 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.396 | 0.393 | 0.394 |
| 2     | 0.013       | 0.026           | 0.067       | 400         | 300         | 200    | TPR  | 0.999 | 1.000 | 0.999 |
|       |             |                 |             |             |             |        | FPR  | 0.000 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.506 | 0.493 | 0.501 |
| 3     | 0.032       | 0.085           | 0.067       | 1000        | 1000        | 200    | TPR  | 0.999 | 0.999 | 0.999 |
|       |             |                 |             |             |             |        | FPR  | 0.000 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.506 | 0.493 | 0.501 |
| 4     | 0.032       | 0.085           | 0.017       | 1000        | 1000        | 50     | TPR  | 0.992 | 0.998 | 0.995 |
|       |             |                 |             |             |             |        | FPR  | 0.000 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.268 | 0.273 | 0.268 |
| 5     | 0.032       | 0.034           | 0.033       | 1000        | 400         | 100    | TPR  | 0.983 | 0.995 | 0.993 |
|       |             |                 |             |             |             |        | FPR  | 0.001 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.361 | 0.348 | 0.357 |
| 6     | 0.064       | 0.068           | 0.067       | 2000        | 800         | 200    | TPR  | 0.987 | 0.997 | 0.995 |
|       |             |                 |             |             |             |        | FPR  | 0.001 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.361 | 0.359 | 0.361 |
| 7     | 0.064       | 0.034           | 0.134       | 2000        | 400         | 400    | TPR  | 0.996 | 0.999 | 0.998 |
|       |             |                 |             |             |             |        | FPR  | 0.001 | 0.000 | 0.000 |
|       |             |                 |             |             |             |        | ACC  | 0.485 | 0.485 | 0.494 |
| 8     | 0.128       | 0.026           | 0.033       | 4000        | 300         | 100    | TPR  | 0.828 | 0.904 | 0.932 |
|       |             |                 |             |             |             |        | FPR  | 0.004 | 0.002 | 0.002 |
|       |             |                 |             |             |             |        | ACC  | 0.398 | 0.353 | 0.323 |
| 9     | 0.064       | 0.026           | 0.017       | 2000        | 300         | 50     | TPR  | 0.867 | 0.945 | 0.958 |
|       |             |                 |             |             |             |        | FPR  | 0.003 | 0.001 | 0.001 |
|       |             |                 |             |             |             |        | ACC  | 0.395 | 0.320 | 0.300 |
| 10    | 0.128       | 0.051           | 0.017       | 4000        | 600         | 50     | TPR  | 0.840 | 0.926 | 0.936 |
|       |             |                 |             |             |             |        | FPR  | 0.002 | 0.001 | 0.001 |
|       |             |                 |             |             |             |        | ACC  | 0.358 | 0.314 | 0.298 |

Table 3: Classification rates averaged over 100 simulated point patterns generated on the *chicago* linear network with two nested sub-networks with $\lambda_{c_1}$, $\lambda_{c_2}$ and $\lambda_f$ intensities and $E[n_{c_1}]$, $E[n_{c_2}]$ and $E[n_f]$ expected numbers of feature and clutter points.
4.4 Estimating two disjoint features in presence of clutter

In order to highlight the flexibility of the classification method in terms of feature shape, we simulate a down-to-earth scenario on the streets of Antonio Narino locality in Bogota, Colombia (Moncada 2018). The road network of Antonio Narino locality has 2062 intersections, 2706 uninterrupted segments and a total length of 128.69 km. The feature points are simulated on two of the longest roads, as shown in Figure 6. The first road (right) is a sub-network with 237 intersections, 254 uninterrupted segments and a total length of 8.32 km and the second road (left) is a sub-network with 106 intersections, 112 uninterrupted segments and a total length of 3.68 km.

![Figure 6: Simulation of two disjoint feature point patterns with clutter on the streets of the Antonio Nariño locality. Point patterns simulated from three Poisson processes with λc = 0.003 and E[n_c] = 400, λ_f1 = 0.018 and E[n_f1] = 150, λ_f2 = 0.014 and E[n_f2] = 50.](image)

The report of the results of the classification procedure is shown in Table 1. In the first road, we simulate features (feature 1) with expected number E[n_f1] and intensity λ_f1. Likewise, in the second road, we simulate features (feature 2) with expected number E[n_f2] and intensity λ_f2. These results are shown for different combinations of intensities λc, λ_f1 and λ_f2. In stages 1 and 2, we take λc < λ_f1 < λ_f2. In stages 3 and 4, we take λc < λ_f1 > λ_f2 with λc < λ_f2 in stage 3 and λc > λ_f2 in stage 4. In stages 5 and 6, we take λc ≈ λ_f1 ≈ λ_f2. In stages 7 and 8, we take λc > λ_f1 < λ_f2 with λc < λ_f2 in stage 7 and λc > λ_f2 in stage 8. In stages 9 and 10, we take λc > λ_f1 > λ_f2. The classification rates of the procedure in Table 1 show a good classification method performance in all stages from 1 to 10 in terms of true- and false-positive rates for ˆK values. In the stages where the classification method performs well, its performance with K = 5 and K = 10 is very similar to that with ˆK.

Overall, these are promising simulation results, empirically proving that the classification procedure can correctly identify points between clutter and features also in this case, with a larger network and complex feature shapes.
Table 4: Classification rates averaged over 100 simulated point patterns generated on the Antonio Nariño linear network with $\lambda_c$, $\lambda_{f_1}$ and $\lambda_{f_2}$ intensities and $E[n_c]$, $E[n_{f_1}]$ and $E[n_{f_2}]$ expected number of points for clutter, feature one and feature two, respectively.

| Stage | $\lambda_c$ | $\lambda_{f_1}$ | $\lambda_{f_2}$ | $E[n_c]$ | $E[n_{f_1}]$ | $E[n_{f_2}]$ | Rate | $K$ | $5$ | $10$ | $\bar{K}$ |
|-------|-------------|-----------------|-----------------|---------|-------------|-------------|-------|-----|-----|-----|-------|
| 1     | 0.047       | 0.072           | 0.109           | 6000    | 600         | 400         | TPR   | FPR | 0.974 | 0.987 | 0.987 |
|       |             |                 |                 |         |             |             |       |     | 0.004 | 0.002 | 0.002 |
|       |             |                 |                 |         |             |             |       |     | 0.514 | 0.430 | 0.447 |
| 2     | 0.023       | 0.036           | 0.054           | 3000    | 300         | 200         | TPR   | FPR | 0.950 | 0.969 | 0.971 |
|       |             |                 |                 |         |             |             |       |     | 0.008 | 0.005 | 0.005 |
|       |             |                 |                 |         |             |             |       |     | 0.496 | 0.354 | 0.361 |
| 3     | 0.047       | 0.072           | 0.054           | 6000    | 600         | 200         | TPR   | FPR | 0.953 | 0.981 | 0.985 |
|       |             |                 |                 |         |             |             |       |     | 0.006 | 0.002 | 0.002 |
|       |             |                 |                 |         |             |             |       |     | 0.493 | 0.382 | 0.366 |
| 4     | 0.047       | 0.072           | 0.014           | 6000    | 600         | 50          | TPR   | FPR | 0.936 | 0.972 | 0.974 |
|       |             |                 |                 |         |             |             |       |     | 0.007 | 0.003 | 0.003 |
|       |             |                 |                 |         |             |             |       |     | 0.488 | 0.359 | 0.337 |
| 5     | 0.093       | 0.093           | 0.094           | 12000   | 775         | 345         | TPR   | FPR | 0.936 | 0.969 | 0.974 |
|       |             |                 |                 |         |             |             |       |     | 0.006 | 0.003 | 0.002 |
|       |             |                 |                 |         |             |             |       |     | 0.468 | 0.402 | 0.380 |
| 6     | 0.035       | 0.036           | 0.034           | 4500    | 300         | 125         | TPR   | FPR | 0.884 | 0.947 | 0.957 |
|       |             |                 |                 |         |             |             |       |     | 0.011 | 0.005 | 0.004 |
|       |             |                 |                 |         |             |             |       |     | 0.463 | 0.279 | 0.252 |
| 7     | 0.047       | 0.024           | 0.094           | 6000    | 200         | 345         | TPR   | FPR | 0.900 | 0.929 | 0.936 |
|       |             |                 |                 |         |             |             |       |     | 0.009 | 0.006 | 0.006 |
|       |             |                 |                 |         |             |             |       |     | 0.471 | 0.335 | 0.303 |
| 8     | 0.047       | 0.024           | 0.034           | 6000    | 200         | 125         | TPR   | FPR | 0.803 | 0.892 | 0.921 |
|       |             |                 |                 |         |             |             |       |     | 0.011 | 0.006 | 0.004 |
|       |             |                 |                 |         |             |             |       |     | 0.442 | 0.266 | 0.207 |
| 9     | 0.047       | 0.024           | 0.014           | 6000    | 200         | 50          | TPR   | FPR | 0.754 | 0.860 | 0.895 |
|       |             |                 |                 |         |             |             |       |     | 0.010 | 0.006 | 0.004 |
|       |             |                 |                 |         |             |             |       |     | 0.435 | 0.247 | 0.189 |
| 10    | 0.093       | 0.024           | 0.014           | 12000   | 200         | 50          | TPR   | FPR | 0.701 | 0.748 | 0.810 |
|       |             |                 |                 |         |             |             |       |     | 0.006 | 0.005 | 0.004 |
|       |             |                 |                 |         |             |             |       |     | 0.421 | 0.337 | 0.235 |

5 Estimating road traffic accident feature in two cities of Colombia

A traffic accident is an involuntary event generated by at least one vehicle in motion on roads in which there is damage to vehicles and objects and injuries or death to the people involved (WHO, 2018). According to WHO (2018), traffic accidents are the eighth cause of death worldwide and the leading cause of death for people between 5 and 29 years of age. Particularly in Colombia, ITF (2019) reports that between 2010 and 2018, the trend for the number of dead people in road accidents has been upward, and the annual number
of these deaths increased by nearly 30%. Based on the report of the National Road Safety Agency of Colombia ANSV (2021), the period 2016 to 2019, the number of dead people in road accidents annually remained between 3.4% and 3.7% of the total number of traffic accidents. When a traffic accident occurs, a chain of actions of different private and public entities begins to clarify what happened and determine the cause of this traffic accident. This means that the investigation of the reasons for which the accident is generated begins after its occurrence. Some factors that generally produce a traffic accident are human, environmental or mechanical. These factors change according to the country or city where the accidents occur; therefore, traffic accidents can be seen as events randomly distributed on a road network, and since their study is carried out once the event has occurred, we can consider that an underlying point process governs them on the linear network.

5.1 Estimation of the road traffic injuries and deaths feature in the city of Bogota

Bogota is Colombia’s capital and the country’s largest and most populous city. In 2015, it was estimated that this city had 2148541 vehicles (Secretaría Distrital de Movilidad 2015). The simplified city road network used in this work consists of 199135 intersections, 243157 uninterrupted segments and a total length of 8218.90 km. We analyzed a database of 10979 traffic accidents that resulted in injuries or deaths in 2015 in the urban area of the city, published in the OpenData portal of Bogota Town Hall, together with the road network shapefile (Moncada 2018; Secretaría Distrital de Movilidad 2022). Identifying features of road traffic injuries and deaths provides valuable information for decision-making in public policies that help prevent more of these unfortunate events.

We first estimate the non-parametric kernel-based intensity using a two-dimensional Gaussian kernel with a bandwidth of $\sigma = 600$ meters (Rakshit et al. 2019b) shown in Figure 7a. The kernel estimate of accident intensity in Bogota shows that the majority of traffic accidents occur in different sectors around the city centre. However, a large part of the road network has intermediate intensity values, which makes it difficult to decide whether it is a road with a high accident rate or not. In this way, it would be possible to delimit some sub-networks where the probability of occurrence of new traffic accidents is high, but the roads with intermediate values make it difficult to decide on the delimitation of the sub-networks that must be intervened in order to avoid new traffic accidents.

The results of the classification procedure for the traffic accidents in the city of Bogota, during 2015 for $\hat{K} = 3$, are shown in Figure 7b. The classification of the traffic accidents is depicted as a multitype point pattern for a dichotomous mark (feature or clutter), in which it is possible to identify the road segments in the feature with a high number of traffic injuries and deaths in the city of Bogota. Therefore, the classification procedure can help to decide whether the road segments with intermediate values of intensity are significantly relevant to delimit the sub-network of the spatial feature and use this as a companion tool of the intensity to study the behaviour of the traffic accidents in Bogota.
Figure 7: Traffic accidents on the road network of the city of Bogota in which at least one person was injured or death recorded in the year 2015. *Left:* Estimated non-parametric kernel-based intensity with bandwidth $\sigma = 600$ meters. *Right:* Estimated feature and clutter using the classification procedure for $K = 3$.

In detail, the estimated intensity of the traffic injuries and death accidents in Bogota has a couple of large hot spots in the central-eastern zone of the city. This zone includes the downtown, the palace of justice and the national capitol, among other important places, surrounded by government offices, museums, churches and some historic buildings where many vehicles generally move. Additionally, an important avenue crosses the entire city from south to north and has access to many other avenues that also pass through this zone. Other relevant hot spots in the city are located in the north zone, where all of the city’s financial, cultural, and recreational activity places are concentrated. The centre-west is another zone with a high probability of occurrence of traffic accidents where large factories, parks, sports centres, administrative offices, and the international airport are located, which means that the roads that cross this zone are normally crowded with vehicles. Based on the estimated intensity and considering the classification results is possible to appreciate more explicitness the roads segments into the spatial feature that has the highest number of accidents in the zone of the hot spot.
In summary, while Figure 7a shows the zones with the highest intensity values, the output in Figure 7b helps to the delimitation of the sub-networks of the feature and, in this way, to define which roads must be intervened with the highest priority to prevent traffic accidents. Therefore, it is no need to intervene on all roads in a zone with high or medium accident density.

5.2 Estimation of road traffic injuries and deaths features in Medellin

Medellin is the second most important city in Colombia after Bogota. In 2019 it was estimated that this city had 1756893 vehicles (Medellín cómo vamos, 2021). The city’s road network consider in this study consists of 110533 intersections, 115939 uninterrupted segments and a total length of 2306.48 km. We analysed a database of 22743 traffic accidents that resulted in injuries or deaths in 2019 in the city, also published in the OpenData portal of Medellin Town Hall, together with the road network shapefile (Secretaría de Movilidad, 2022).

![Figure 8](image-url)  
Figure 8: Traffic accidents on the road network of Medellin in which at least one person was injured or death in 2019. *Left:* Estimated non-parametric kernel-based intensity with bandwidth $\sigma = 700$ meters. *Right:* Estimated feature and clutter using the classification procedure for $\hat{K} = 2$.

Figure 8a shows the non-parametric kernel-based intensity estimated with bandwidth $\sigma = 700$ meters. The intensity of the traffic injuries and death accidents in Medellin has
a single large hot spot in the central-eastern zone which is the commercial and financial centre of the city. This zone includes the downtown, a large part of the city’s commerce, residences, factories, administrative offices, museums, churches, and historic buildings, among others. Additionally, four important avenues cross the hot spot, and a large proportion of the vehicles that circulate in the city every day use these roads. Figure 8b shows the result of the classification procedure of the traffic injuries and death accidents in Medellin in 2019 for $\hat{K} = 2$. Based on the intensity and the classification procedure, it is possible to differentiate between the road segments into the spatial feature with the highest number of accidents in the hot spots. Therefore, the combination of the information provided by these two exploratory tools makes it more feasible to interpret the results in geometric spaces that are not intuitive.

6 Conclusions

In this paper, we consider the problem of estimating features in a point pattern on a linear network in the presence of clutter. The proposed classification method allows for the distinction between clutter and feature in complex geometries and the parameters estimated by an EM algorithm. Furthermore, it is possible to automatically select the number of nearest neighbours to consider in the statistical analysis through the segmented regression model. The methodology of classification can be applied without assumptions about the feature shapes, and it relies on the assumption that the feature and clutter are superimposed Poisson processes. The simulation study highlights the performance and accuracy of the proposed procedure in terms of classification rates. The classification procedure applied to the point pattern of traffic accidents occurring on the road network of a city, together with the estimation of intensity, can be a helpful tool in decision making when delimiting the hot-spot roads that have priority to be intervened to prevent these accidents.

Given the growing interest in network data, we believe that the proposed approach could be used in many different contexts of application. Future work may consider modifying the EM-algorithm to estimate the intensities of more than two groups of features or clutter simulated over the same linear network, finding features with inhomogeneous intensity in the presence of clutter or extending this methodology to the spatio-temporal case on linear networks.

Acknowledgments

The research work of Juan F. Díaz-Sepúlveda and Francisco J. Rodríguez-Cortés has been partially supported by Universidad Nacional de Colombia, HERMES projects, Grant/Award Number: 56470. The research work of Giada Adelfio and Nicoletta D'Angelo has been supported by the University of Palermo, ‘FFR 2021 GIADA ADELFIO’.
References

Allard, D. and Fraley, C. (1997). Nonparametric maximum likelihood estimation of features in spatial point processes using voronoi tessellation. *Journal of the American Statistical Association*, 92(440):1485–1493.

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(4):591–617.

ANSV (2021). *Informe Sobre Seguridad Vial Para el Congreso de la República de Colombia*. Agencia Nacional de Seguridad Vial, Bogota.

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 (Applied Statistics)*, 63(5):673–694.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., and Davies, T. M. (2021a). Analysing point patterns on networks — a review. *Spatial Statistics*, 42:100435. Towards Spatial Data Science.

Baddeley, A., Turner, R., and Rubak, E. (2021b). *spatstat.data: Datasets for ’spatstat’ Family*. R package version 2.1-2.

Banfield, J. D. and Raftery, A. E. (1993). Model-based gaussian and non-gaussian clustering. *Biometrics*, pages 803–821.

Byers, S. and Raftery, A. E. (1998). Nearest-neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association*, 93(442):577–584.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, 13(2):195–212.

Cronie, O., Moradi, M., and Mateu, J. (2020). Inhomogeneous higher-order summary statistics for point processes on linear networks. *Statistics and computing*, 30(5):1221–1239.

D’Angelo, N., Adelfio, G., and Mateu, J. (2021). Assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. *Spatial Statistics*, 45:100–534.

D’Angelo, N., Adelfio, G., and Mateu, J. (2022). Local inhomogeneous second-order characteristics for spatio-temporal point processes occurring on linear networks. *Statistical Papers*, pages 1–27.
Dasgupta, A. and Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American statistical Association*, 93(441):294–302.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Federer, H. (1969). *Geometric measure theory*. Springer, Heidelberg.

ITF (2019). *Road Safety Annual Report 2019*. OECD Publishing, Paris.

Jammalamadaka, A., Banerjee, S., Manjunath, B. S., and Kosik, K. S. (2013). Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC bioinformatics*, 14(1):1–19.

McSwiggan, G., Baddeley, A., and Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44(2):324–345.

McSwiggan, G., Baddeley, A., and Nair, G. (2020). Estimation of relative risk for events on a linear network. *Statistics and Computing*, 30.

Medellín cómo vamos (2021). *Informe de Calidad de Vida de Medellín, 2020*. Medellín cómo vamos, Medellín.

Moncada, I. (2018). Análisis espacio-temporal de los accidentes de tránsito en bogotá utilizando patrones puntuales. *Tesis de Maestría. Universidad Nacional de Colombia*.

Moradi, M. M., Rodríguez-Cortés, F. J., and Mateu, J. (2018). On kernel-based intensity estimation of spatial point patterns on linear networks. *Journal of Computational and Graphical Statistics*, 27(2):302–311.

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19):3055–3071.

Muggeo, V. M. R. (2008). segmented: An R package to fit regression models with broken-line relationships. *R news*, 8(1):20–25.

Rakshit, S., Baddeley, A., and Nair, G. (2019a). Efficient code for second order analysis of events on a linear network. *Journal of Statistical Software*, 90(1):1–37.

Rakshit, S., Davies, T., Moradi, M. M., McSwiggan, G., Nair, G., Mateu, J., and Baddeley, A. (2019b). Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review*, 87(3):531–556.
Rakshit, S., McSwiggan, G., Nair, G., and Baddeley, A. (2021). Variable selection using penalised likelihoods for point patterns on a linear network. *Australian & New Zealand Journal of Statistics*, 64(1):1 – 38.

Secretaría de Movilidad (2022). *GeoMedellín*. Alcaldía de Medellín.

Secretaría Distrital de Movilidad (2015). *Movilidad en cifras 2015*. Alcaldía Mayor de Bogotá, Bogotá.

Secretaría Distrital de Movilidad (2022). *Sistema Integrado de Información sobre Movilidad Urbana Regional*. Alcaldía de Bogotá.

WHO (2018). *Global status report on road safety 2018*. World Health Organization, Geneva.