



Universitat de Lleida

Document downloaded from

<http://hdl.handle.net/10459.1/70327>

The final publication is available at:

<https://doi.org/10.1002/sta4.323>

Copyright

(c) Wiley, 2020



Directional analysis for point patterns on linear networks

Mehdi Moradi*^{1,2} | Jorge Mateu³ | Carles Comas⁴

¹Department of Statistics, Computer Science, and Mathematics, Public University of Navarre, Pamplona, Spain

²Institute of Advanced Materials and Mathematics (InaMat²), Public University of Navarre, Pamplona, Spain

³Department of Mathematics, University Jaume I, Castellón, Spain

⁴Department of Mathematics, University of Lleida, Lleida, Spain

Correspondence

* Mehdi Moradi

Email: mehdi.moradi@unavarra.es

Present Address

Department of Statistics, Computer Science and Mathematics, Public University of Navarre, 31006, Pamplona, Spain.

Summary

Statistical analysis of point processes often assumes that the underlying process is isotropic in the sense that its distribution is invariant under rotation. For point processes on \mathbb{R}^2 , some tests based on the K- and nearest neighbour orientation functions have been proposed to check such an assumption. However, anisotropy and directional analysis need proper caution when dealing with point processes on linear networks, as the implicit geometry of the network forces particular directions that the points of the pattern have to necessarily meet. In this paper, we adapt such tests to the case of linear networks, and discuss how to use them to detect particular directional preferences, even at some angles that are different from the main angles imposed by the network. Through a simulation study, we check the performance of our proposals under different settings, over a linear network and a dendrite tree, showing that they are able to precisely detect the directional preferences of the points in the pattern, regardless the type of spatial interaction and the geometry of the network. We use our tests to highlight the directional preferences in the spatial distribution of traffic accidents in Barcelona (Spain) during 2019, and in Medellin (Colombia), during 2016.

KEYWORDS:

Circular density, Dendrite, Immigration-death, K-function, Nearest neighbour orientation function, Relative risk, Traffic accidents

1 | INTRODUCTION

The analysis of point patterns is often tied with the assumptions that the underlying point process is stationary and isotropic in the sense that its distribution is invariant under both translations and rotations, respectively. Although such assumptions facilitate developing various methodologies to analyse point pattern data, they seem unlikely to be held in practice since the spread of the events may be affected by e.g. some latent attributes. Literature has focused much more on building tools to account for non-stationarity (Baddeley, Møller, & Waagepetersen 2000; van Lieshout 2011), than to take care of anisotropy, which has largely received quite less attention (Comas, Conde, & Mateu 2018; Comas, Rodriguez-Cortes, & Mateu 2015; Rajala, Redenbach, Särkkä, & Sormani 2018).

When looking at point patterns on linear networks, it is challenging to consider such assumptions due to the fact that it is not clear how such transformations or rotations should be done, since there is no non-trivial geometrical transformations that preserve the network (Baddeley, Nair, Rakshit, & McSwiggan 2017; Cronie, Moradi, & Mateu 2020), and besides the structure of the underlying network itself is rarely homogeneous and/or isotropic. Thinking about street networks, we may see a concentration of streets in downtown, with other long highways in the countryside, according to which these streets/highways may further locate in particular directions with respect to an origin. In this regard, Baddeley et al. (2017) showed that popular procedures to construct point processes on linear networks give rise to models which are no longer stationary, if the considered metric is the shortest-path distance. Notwithstanding, under different assumptions, geometrically corrected second-order and higher-order summary statistics are already built for inhomogeneous point processes on linear networks (Ang, Baddeley, & Nair 2012; Cronie et al. 2020).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/sta4.323

Clearly the geometry of a network might force the corresponding point pattern towards particular angles since different line segments are placed in particular fixed directions. In such cases, detecting (an)isotropy is even more challenging since it is not clear what is the meaning of anisotropy in such a situation. In addition, the existence of more segments in particular angles does not necessarily lead to more points in these directions due to, for instance, the effect of possible latent attributes. Further, it is not clear if the tendency of the underlying distribution of points towards particular angles is only caused by the geometry of the network.

In this line of arguments, the literature currently misses appropriate statistical tools to study the existence of tendencies towards particular directions in the distribution of points when taking the underlying network into account. Here, it is crucial to be able to distinguish a directional tendency caused by the geometry of network from a tendency coming from the underlying distribution of events. In this paper, having the work of Ohser and Stoyan (1981) and Comas et al. (2015) in mind, and upon the geometrically corrected K-function (Ang et al. 2012), we propose some tools to discover directional tendencies in the distribution of point patterns over a network apart from what is caused by the geometry of the underlying network. In addition, we propose to use the counterpart version of the nearest neighbour orientation function for linear networks and its circular density estimation (Baddeley, Rubak, & Turner 2015, Chapter 8). However, these functions will likely point to the main angles of the underlying network. In order to exclude what the underlying network inherently brings into these functions, we suggest to compare the estimated functions coming from the data, with those obtained from point patterns which are angle-wise uniformly distributed over the network. Thereby, we are able to see any discrepancy between such estimators and, furthermore, we can estimate the relative risk of the circular density of the nearest neighbour orientation function of our data relative to that of an angle-wise uniformly distributed pattern. Note that such relative risk uncovers the angles for which the pattern in question shows higher/lower densities than the case where points are angle-wise uniformly distributed.

The paper is structured as follows. In Section 2 we provide the required background about point patterns on linear networks. Section 3 is devoted to introduce the K-function-based orientation function, and the nearest neighbour orientation function. In Section 4 we study the performance of our methodology under different settings. Section 5 provides our directional analysis for traffic accidents in Barcelona (Spain) during 2019, and in Medellin (Colombia) during 2016. The paper ends with a discussion in Section 6.

2 | PRELIMINARIES

2.1 | Linear networks

A linear network $L \subset \mathbb{R}^2$ is defined as a union of finite number of line segments $l_i = \{tu_i + (1-t)v_i : 0 \leq t \leq 1\}$ such that $u_i, v_i \in \mathbb{R}^2$ are the endpoints of l_i . The length of each line segment l_i is denoted by $|l_i|$, and it gives rise to $|L|$ standing for the total length of the linear network L , being the sum of the length of all individual segments. The endpoints are usually called vertices, and the number of segments which exit from each vertex is referred to as the degree of that vertex. The distance between any two arbitrary points $u, v \in L$ is denoted by $d_L(u, v)$, which is commonly calculated by a regular metric distance (Rakshit, Nair, & Baddeley 2017). Examples of regular metrics include Euclidean and shortest-path (travel) distance. We add that the intersection between any two arbitrary segments l_i and l_j is either an empty set or a mutual vertex of both. In other words, we assume there is no under/over-pass in L . A linear network which has no closed loop is often called a "tree", that might be often seen in applications related to rivers or dendrite trees of neurons. See Baddeley et al. (2015, Chapter 17) for more details.

2.2 | Point processes on linear networks

Throughout the paper, a simple point process on the linear network L is denoted by X , being a random countable subset of \mathbb{R}^2 , and $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ is used to show a realisation of such process with $0 < n < \infty$ points. For any measurable function $f \geq 0$, the Campbell formula (Daley & Vere-Jones 2008) states that

$$\mathbb{E} \sum_{\mathbf{x} \in X} f(\mathbf{x}) = \int_L f(u) \lambda(u) d_1 u, \quad (1)$$

where $\lambda(\cdot)$ is called the intensity function of X , and d_1 stands for integration with respect to arc length. The intensity function $\lambda(u)$ gives rise to the expected number of points per unit length of L in the vicinity of $u \in L$. If $\lambda(\cdot) \equiv \lambda$, the process is called homogeneous, otherwise it is referred as inhomogeneous. Common statistical methodologies for point processes often assume that X is stationary, i.e. the distribution of X is invariant with respect to a family of transformations/shifts, and isotropic, i.e. the distribution of X is invariant with respect to a family of rotations. These assumptions, however, are quite challenging when taking a linear network as the state space of an underlying point process X since there is no guarantee that points will still live on the underlying network after applying a transformation/rotation (Baddeley et al. 2017; Cronie et al. 2020).

Moreover, we denote the number of points falling in any arbitrary measurable subset $A \subseteq L$ by $N(X \cap A)$, with the expectation

$$\mathbb{E}[N(X \cap A)] = \mathbb{E} \sum_{x \in X} \mathbf{1}\{x \in A\} = \int_A \lambda(u) d_1 u, \quad A \subseteq L, \quad (2)$$

and, further

$$\mathbb{E}[N(X \cap A)N(X \cap B)] = \mathbb{E} \sum_{x, y \in X} \mathbf{1}\{x \in A, y \in B\} = \int_A \int_B \lambda_2(u, v) d_1 u d_1 v, \quad A, B \subseteq L, \quad (3)$$

where $\lambda_2(\cdot, \cdot)$ is called the second-order product intensity function of X , and gives the expected number of points per unit length of L jointly happening in the vicinities of $u, v \in L$.

2.2.1 | Second-order summary statistics

An important while basic step when dealing with point pattern analysis, is checking for the existence of any type of interaction between points, which is essential prior to model fitting. Common types of interaction include clustering or repulsion as opposite to randomness. Popular tools to detect interaction between points have been the K - and pair correlation functions. The pair correlation function is defined as

$$g(u, v) = \frac{\lambda_2(u, v)}{\lambda(u)\lambda(v)}, \quad u, v \in L. \quad (4)$$

If $\lambda_2(u, v) > \lambda(u)\lambda(v)$ then points tend to happen jointly, whereas $\lambda_2(u, v) < \lambda(u)\lambda(v)$ means that points rather tend to maintain an inter-distance. Further, $g(v, u) = 1$ claims the non-existence of clustering or repulsion between points which is a well-known property of Poisson processes. Point processes on L for which $g(u, v) = g(d_L(u, v))$ are called second-order reweighted pseudostationary (SORS) by Ang et al. (2012), and for such processes, the geometrically corrected K -function is defined as

$$K(u, r) = \mathbb{E} \left[\sum_{x \in X} \frac{\mathbf{1}\{d_L(u, x) \leq r\}}{\lambda(x)m(u, d_L(u, x))} \middle| u \in X \right], \quad (5)$$

and

$$K(u, r) = K(r) = \int_0^r g(t) dt, \quad u \in L. \quad (6)$$

The above functions are called geometrically corrected since they account for the geometry of the network through the correction factor $m(u, r)$ which gives the number of points on L with exact distance r units from u . For Poisson processes $K(r) = r$, whereas $K(r) > r$ ($K(r) < r$) indicates clustering (repulsion) between points. See Ang et al. (2012) for more details.

3 | METHODOLOGY

We now present the K -function-based orientation function and the counterpart of the nearest neighbour orientation function for linear networks.

3.1 | Second-order orientation analysis

The use of K - and pair correlation functions, in Section 2, is usually carried out to study the existence of any pairwise interaction among points without taking possible directional effects into account. We now propose a modified version of (5) to further account for the pairwise directional interaction between points. The new K -function is of the form

$$K^{\text{anis}}(u, r, \phi) = \mathbb{E} \left[\sum_{x \in X} \frac{\mathbf{1}\{d_L(u, x) \leq r\} \mathbf{1}\{\Phi(u, x) \leq \phi\}}{\lambda(x)} \frac{m(u, d_L(u, x)|\phi)}{m(u, d_L(u, x))^2} \middle| u \in X \right], \quad (7)$$

where $\Phi(u, x)$ stands for the angle between $u \in L$ and $x \in X$, and $m(u, r|\phi)$ is the number of points laying exactly r units away from u within an angle ϕ . Note that $\Phi(u, x) = \Phi(x, u) \pm \pi$, and $K^{\text{anis}}(u, r, \phi) = K^{\text{anis}}(r, \phi)$. We highlight that $K^{\text{anis}}(r, 2\pi) = K(r)$, since $\mathbf{1}\{\Phi(u, x) \leq 2\pi\} = 1$ and $m(u, d_L(u, x)|2\pi) = m(u, d_L(u, x))$. Following Ang et al. (2012), a non-parametric estimator of K^{anis} is given by

$$\hat{K}^{\text{anis}}(r, \phi) = \frac{1}{|L|} \sum_{i=1}^n \sum_{j \neq i} \frac{\mathbf{1}\{d_L(x_i, x_j) \leq r\} \mathbf{1}\{\Phi(x_i, x_j) \leq \phi\}}{\lambda(x_i)\lambda(x_j)} \frac{m(x_i, d_L(x_i, x_j)|\phi)}{m(x_i, d_L(x_i, x_j))^2}, \quad (8)$$

which reduces to the K -function estimator proposed by Ang et al. (2012), when $\phi = 2\pi$. The advantage of K^{anis} over K is that former further analyses the pairwise interaction within angles. However, rather than focusing on this aspect, we are here more interested in making use of K^{anis}

for orientation analysis. Therefore, we define the following cumulative distribution function, in the line of the original contribution of (Ohser & Stoyan 1981),

$$O_{(r_1, r_2)}(\phi) = \int_0^\phi \int_{r_1}^{r_2} dK^{\text{anis}}(r, \xi) / \int_0^{2\pi} \int_{r_1}^{r_2} dK^{\text{anis}}(r, \xi), \quad (9)$$

which studies the orientations of the point process X for inter-distances between $0 \leq r_1$ and $r_1 < r_2$. Note that different combinations of r_1, r_2 may describe short-range, middle-range, or long-range orientation of X (Ohser & Stoyan 1981). Since $O_{(r_1, r_2)}(\cdot)$ is a cumulative function, any tendency of the distribution of points towards a particular angle appears as a jump in $O_{(r_1, r_2)}(\cdot)$. However, it might be of more interest to actually look into the vicinity of angles rather than the cumulative (scaled) count of points up to a fixed angle. Therewith, the use of the density function

$$O_{(r_1, r_2)}^*(\phi) = O_{(r_1, r_2)}(\phi + \xi) - O_{(r_1, r_2)}(\phi - \xi), \quad 0 < \xi < \phi < 2\pi, \quad (10)$$

is recommended (Comas et al. 2015). In this case, a clear tendency towards particular angles appears as sharp peaks in $O_{(r_1, r_2)}^*(\cdot)$. In practice, calculating the above functions can be quite costly, especially when dealing with big data/network (Rakshit, Baddeley, & Nair 2019).

3.2 | Nearest neighbour orientation function

Although both $O_{(r_1, r_2)}(\cdot)$ and $O_{(r_1, r_2)}^*(\cdot)$ are useful tools to check the assumption of anisotropy, their practicality, however, might be affected by the choices of r_1 and r_2 . Hence, we next present an alternative approach using the nearest neighbour orientation distribution. To do so, let $\Psi = \{\psi_1, \dots, \psi_n\}$ be the set of angles of the vectors joining each data point in $\mathbf{x} = \{x_1, \dots, x_n\}$ to its nearest neighbour. A circular probability density function, say $\rho(\cdot)$, of such angles can then highlight particular tendencies towards some specific angles. More explicitly,

$$\mathbb{P}(\Psi \leq \psi) = \int_0^\psi \rho(u) du, \quad 0 \leq \psi \leq 2\pi. \quad (11)$$

Note that in the presence of isotropy $\rho(u) \equiv 1/2\pi$ for $0 \leq u \leq 2\pi$. We note that, in practice, such circular probability density function needs to be estimated (Illian, Penttinen, Stoyan, and Stoyan (2008, Chapter 4) and Baddeley et al. (2015, Chapter 8)). In our simulation study and data analysis, we make use of a kernel-smoothed estimate (Baddeley et al. 2015, Chapter 8, pp. 278). Computationally, the nearest neighbour orientation function is more efficient than the second-order tools in Section 3.1. We add that this technique can be extended to the circular probability density function of k -th ($k > 1$) nearest neighbour orientation function.

3.3 | Relative influence of the underlying network

Generally, one possible source to make a point process in \mathbb{R}^d anisotropic might be a latent geometry which forces events to occur in some specific directions (Rajala et al. 2018). Under these circumstances, although such geometry may not be visible, however, the main detected directions by either second-order orientation analysis or the nearest neighbour orientation function might be (miss)understood as evidence of anisotropy. Clearly, when dealing with point processes on linear networks or dendrite trees, there already exists a geometry which forces events to occur in some fixed directions. We then need to explore the differences between the distribution of the angles of the observed point pattern and that of the network itself. It is to be emphasised that even if the points are randomly and angle-wise uniformly distributed over a linear network, the tools for the orientation analysis will still show the main angles of the underlying network; this, however, may not be taken into account as an evidence for anisotropy since points can not happen on other angles. We next propose some solutions for such a concern in order to exclude the directional effects of the underlying network.

With the aim of excluding the effect of the network, one first needs to get an insight in what the geometry of the network brings into such orientation analysis. Hereafter, we call the inherent orientations of the network by the **fingerprint**. Such fingerprint can be estimated through both the orientation function $O_{(\cdot, \cdot)}^*(\cdot)$ and the circular probability density function $\rho(\cdot)$. We denote the corresponding functions for the network by $O_{(\cdot, \cdot)}^{*,L}(\cdot)$ and $\rho^L(\cdot)$, so that they can be approximately estimated from point patterns over the network with no preferential direction. However, and to avoid possible random effects from simulating point patterns, we suggest to simulate some $m > 1$ times from the model in consideration (with no preferential direction), estimate the $O_{(\cdot, \cdot)}^{*,L}(\cdot)$ and $\rho^L(\cdot)$ for each, and then use such estimates to get an insight into the average behaviour of $\widehat{O}_{(\cdot, \cdot)}^{*,L}(\cdot)$ and $\widehat{\rho}^L(\cdot)$ for angle-wise uniformly distributed patterns. We here note that the type of interaction between points might not be of great importance. Therefore, we estimate the fingerprint of the network based on different models with no directional preference including Poisson, Matérn cluster, and Simple Sequential inhibition models.

Any deviation from the orientations of the network can be detected graphically. One such solution is to compare the distributions of $\widehat{O}_{(r_1, r_2)}^*(\cdot)$ with $\widehat{O}_{(r_1, r_2)}^{*,L}(\cdot)$ for different ranges of r_1, r_2 (Comas et al. 2018). Another solution would be to investigate the relative risk of the circular probability

density functions $\rho(\cdot)$ relative to $\rho^L(\cdot)$, i.e. $\rho(\cdot)/\rho^L(\cdot)$. Note the importance of considering a common bandwidth when estimating such relative risk (Hazelton 2008; Rakshit, Davies, et al. 2019). We highlight that an estimation of this relative risk underlines the angles for which the point pattern in consideration has more/less points than what is expected when points are distributed over the network with no preference towards any fix angle.

4 | SIMULATION STUDY

This section is devoted to evaluate the performance of the methods presented in Section 3. We consider different point process models with and without directional effects, including inhomogeneous Poisson, Matérn cluster, and an immigration-death model. The simulation study is carried out over a linear network as well as a dendrite tree (see Figure 1). The linear network has 64 vertices and 112 segments, maximum vertex degree 4, and total length 16 units. In addition, this network inherently has its segments on angles 0, 90, 180, and 270 degrees. The dendrite has 1000 vertices and 999 segments, maximum vertex degree 4, and an approximated total length of 21 units. The dendrite does not show any preference towards particular angles. For creation of such artificial networks we have used the R packages `spatstat` (Baddeley et al. 2015) and `emstreeR` (Quadros 2019). The models we consider are as follows:

1. Inhomogeneous Poisson process with intensity function $\lambda(u) = \lambda(u_1, u_2) = 12 + \sin(12(u_1 + u_2))$, $u = (u_1, u_2) \in L$. This model does not have any preference towards fixed angles. The simulated pattern over the linear network has 172 points, and the one on the dendrite 228 points. See the left column of Figure 1. The simulation is done using the R package `spatstat` (Baddeley et al. 2015).
2. Matérn cluster process with radius 2 units for the clusters. We consider uniformly distributed 40 parents, so that each is then replaced by 100 uniformly distributed off-springs. Amongst all off-springs we only keep those having an angle of 90 (270) degrees with respect to their parents with a threshold of approximately 1 degree. The final simulated pattern has 169 points. For simulation over the dendrite, we simulate 20 parents, uniformly distributed, and then we replace each by 1000 uniformly distributed off-springs. We then only keep the off-springs that have an angle of 60 (240) degrees with respect to their parents with a threshold of approximately 1 degree. The final generated pattern has 139 points. See the middle column of Figure 1.
3. Immigration-death. Assume that immigrants arrive randomly according to a Poisson process with constant intensity ω . Moreover, a simple death process with intensity μ is incorporated to ensure that the total number of points n remains bounded. Now new arrived points are accepted or rejected in terms of the following spatial interaction mechanism. Let $d_{i,\text{new}}$, $1 \leq i \leq n_s$, where $n_s < n$ is the number of previously established points, be the shortest-path distance between a newly arrived point and all previously established points, and $\alpha_{i,\text{new}}$ be the angle between the newly arrived point and all other points. Now immigrants are accepted with probability

$$P = \prod_{i=1}^{n_s} \Pi_i \quad (12)$$

where

$$\Pi_i = \begin{cases} 1 & \text{if } r_1 < d_{i,\text{new}} < r_2 \quad \text{and} \quad \beta - c_\beta < \alpha_{i,\text{new}} < \beta + c_\beta \\ a & \text{otherwise,} \end{cases} \quad (13)$$

where $r_1 < r_2$ is the spatial scale of anisotropy, β is the prescribed anisotropic directional effect ($0 \leq \beta \leq \pi$), c_β is a threshold, and $0 \leq a \leq 1$ is a constant that defines the strength of the anisotropic effects. The resulting point pattern, defined by (12), after a long enough "burn-in" period is an anisotropic point configuration on a linear network, with angular direction β . In the particular case where $a = 1$, the resulting point pattern is a realisation of a Poisson process with intensity ω/μ . In our simulation study, for both networks, we consider $a = 0$, $\omega = 1$, $c_\beta = 18$ degrees that somehow relaxes the condition of angles, $r_1 = 0$, and $r_2 = 0.15$ units. Moreover, we consider the death rates $\mu = 10^{-5}$ and $\mu = 10^{-3}$ for the realisations over the linear network and dendrite, respectively. The generated pattern over the linear network has 77 points, whereas the pattern simulated over the dendrite consists of 145 points. See the right column of Figure 1.

Figure 1 shows some point pattern realisations from the three mentioned models over both a linear network and a dendrite tree. It can be seen that the inhomogeneous Poisson patterns apparently are not in favour of any particular angle, nor the main angles of the network neither any other angle. However, for the simulated patterns of the Matérn cluster model, we can see that for the linear network most points live on the vertical segments, and for the dendrite the points indicate a diagonal orientation. In addition, the realisation of the immigration-death model over the linear network shows a slight diagonal preference, while its realisation over the dendrite maintains a vertical directional effect.

These indications of anisotropy, however, need to be checked more carefully. Thus, we next make use of the nearest neighbour orientation function to check any angle preference in the simulated patterns. In particular, we are interested to graphically test the null hypothesis of there is no angle effect in the distribution of points, versus the alternative hypothesis which claims the existence of directional effects. Note that having no directional effect here means that points are angle-wise uniformly distributed. In order to check directional effects, we suggest the following:

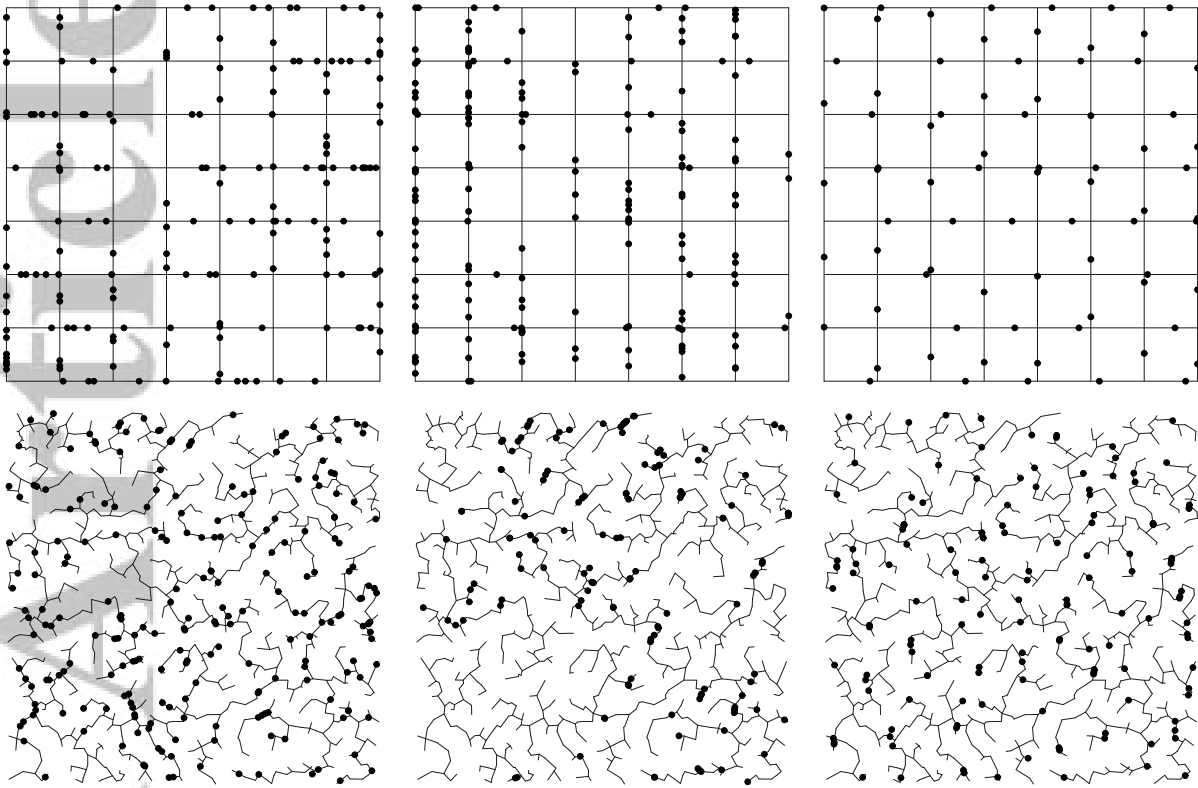


FIGURE 1 Point pattern realisations under different models. Left column: Inhomogeneous Poisson process with intensity function $\lambda(u) = \lambda(u_1, u_2) = 12 + \sin(12(u_1 + u_2))$; Middle column: Matérn cluster process with radius 2 units for the clusters, with a tendency on 90 degrees for the linear network, and a tendency on 60 degrees for the dendrite; Right column: Immigration-death with a tendency on 60 degrees for the linear network, and a tendency on 90 degrees for the dendrite

1) estimate the circular probability density function (11) for the point pattern realisation; 2) generate some $m > 1$ number of realisations from the null hypothesis, and estimate their circular probability density functions; 3) estimate the relative risk of the circular probability density function of the empirical point pattern relative to the average of m estimated probability functions in step 2. We emphasise the importance of the use of a common bandwidth in the estimation of relative risk (Rakshit, Davies, et al. 2019). We make use of the R package `spatstat` for the kernel estimation of the density of the nearest neighbour orientation function.

In order to simulate from the null hypothesis, we consider homogeneous angle-wise uniformly-distributed patterns with three different types of interaction: Poisson, Matérn cluster, and simple sequential inhibition. In each of these models, we simulate the same number of points as the empirical pattern. When estimating the relative risk, bandwidth is selected using the Silverman's rule-of-thumb (Silverman 1986) for each individual pattern, and then a common bandwidth is estimated as the geometrical mean of all individuals.

Figure 2 represents the estimated relative risks for the patterns in Figure 1 relative to their corresponding null hypothesis. These relative risks highlight the differences in the behaviour of the corresponding point pattern with respect to different angles, underlining angle preferences in the spatial distribution of the point pattern. The very first thing we can note from Figure 2 is that the relative risks obtained for the different null hypothesis almost fully coincide. This means that the type of spatial interaction for the patterns defining the null hypothesis has no or little effect, since interest here focuses on angles. In addition, looking closer to the relative risks in Figure 2, we can see that for the inhomogeneous Poisson patterns, the estimated relative risks are basically the same with respect to different angles. In other words, the estimated circular densities of the nearest neighbour orientation function for the simulated patterns in the left column of Figure 1 somehow mimics those of the null hypothesis, confirming no angle preference in the simulated patterns. Regarding the simulated patterns of the Matérn cluster patterns, see middle column in Figure 1, we can see the corresponding estimated relative risks, middle column in Figure 2, clearly highlight the preference towards angles 90 (270) degrees for the pattern on the linear network, and angles 60 (240) degrees for the pattern on the dendrite. In other words, these patterns have quite more points in the detected angles compared to the case where points are angle-wise uniformly distributed. Turning finally to the simulated patterns from the immigration-death model, the interesting point in this case is to have an angle preference towards an angle which is not the

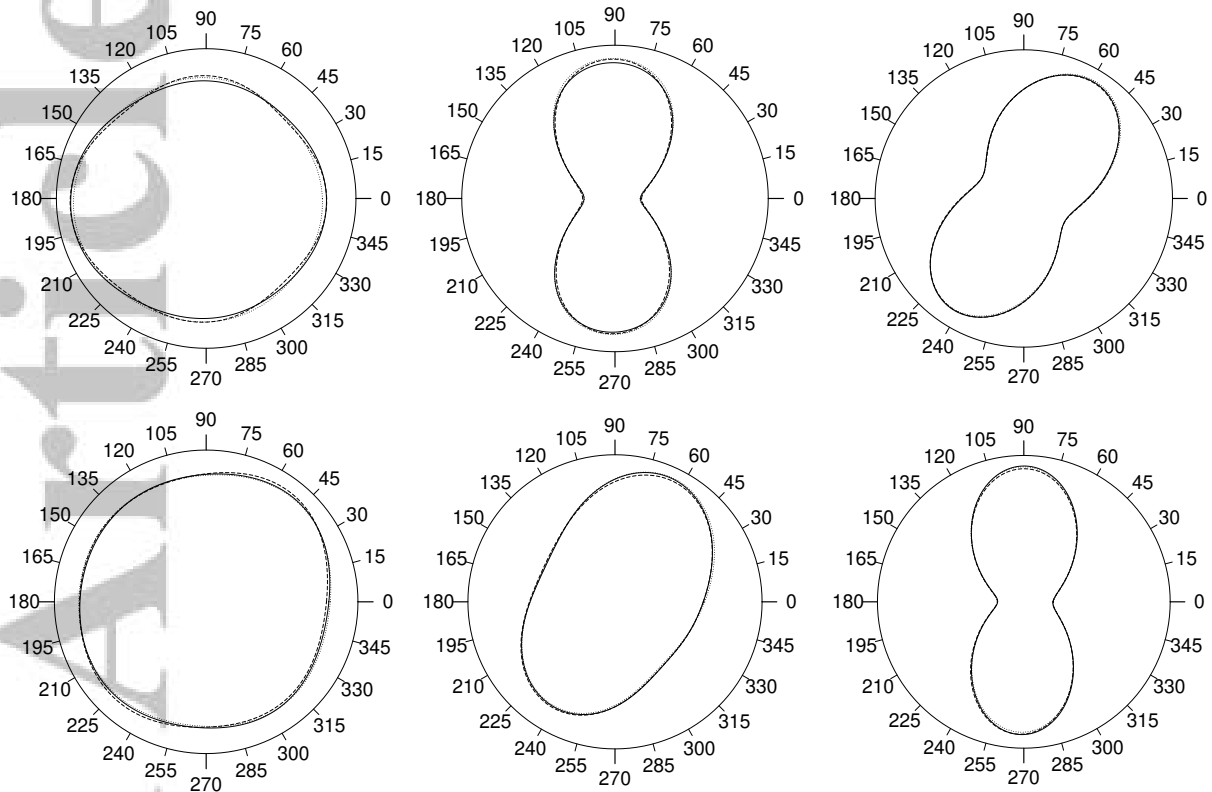


FIGURE 2 Relative risks of the estimated circular density function of the patterns in Figure 1, with the same order, relative to the estimated averaged circular density function under the null hypothesis, which is angle-wise uniformly-distributed Poisson (solid), Matérn cluster (dotted), and simple sequential inhibition (dashed).

main angle of the network, see top-right plot in Figure 1. However, the estimated relative risk for this case clearly reveals the tendency towards the angle 60 (240) degrees for the linear network, and angles 90 (270) for the dendrite. Having checked the presented relative risks in Figure 2, we have generally seen that such suggested relative risk can detect any angle preference in the spatial distribution of points over the network, even if it is not the main angle of the network.

We next turn to comment on the practical use of the K-function-based orientation function. Figure 3 shows the estimated O^* for different ranges of r_1 and r_2 , and for the simulated patterns of the Matérn model over the dendrite tree (middle plot in the bottom row of Figure 1), and the immigration-death model over the squared linear network (right plot in top row of Figure 1). In each case, for a fixed $r_1 > 0$, we estimate O^* for several choices of r_2 . Looking closer to e.g. the two left-side plots in Figure 3, we can see that although the two peaks of \hat{O}^* , in the very left panel, point to the angle preference of the simulated pattern, the behaviour of O^* strongly depends on the choices of r_1 and r_2 , which makes it difficult to come to a concrete conclusion. For the two right-side plots in Figure 3, we can see that for the selected choices of r_1 and r_2 , we can not see the angle preferences of the simulated pattern, i.e. 60 (240) degrees. Note that there might be other choices of r_1 and r_2 which can highlight the angle preferences. We add that it even becomes more complicated when thinking to build e.g. an envelope based on the simulated patterns from the null hypothesis for more solid conclusions. We further note that, in this case, the type of interaction amongst points may make a difference.

Although both presented tools, the nearest neighbour orientation function and the K-function-based orientation function, are prepared to respond to detection of particular main directions making the point pattern anisotropic, we have shown that the former tool is more reliable, computationally faster, and much easier to interpret.

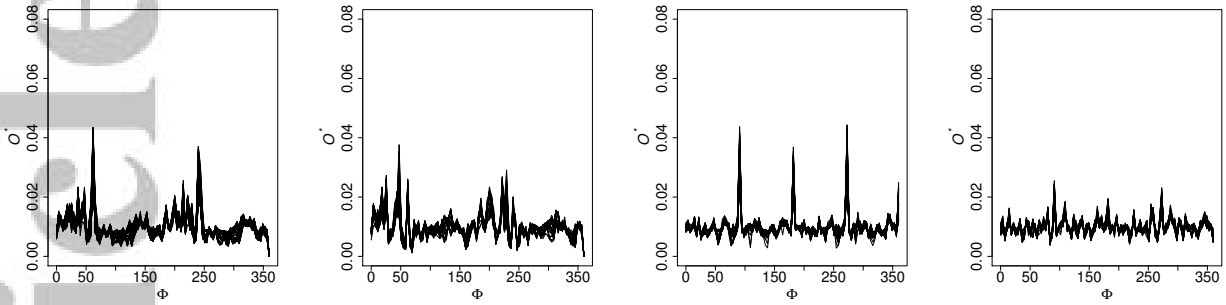


FIGURE 3 Results of the K-function-based orientation function for different ranges of r_1 and r_2 . From Left to Right: Matérn model over the dendrite tree with $r_1 = 0.01$, $0.75 \leq r_2 \leq 1.85$; Matérn model over the dendrite tree with $r_1 = 0.75$, $1.11 \leq r_2 \leq 1.85$; Immigration-death model over the linear network with $r_1 = 0.01$, $0.55 \leq r_2 \leq 1.38$; Immigration-death model over the linear network with $r_1 = 0.55$, $0.83 \leq r_2 \leq 1.38$;

5 | REAL DATA ANALYSIS

This section is devoted to detect main directions and angle preferences in two real cases, traffic accidents in Barcelona (Spain), during 2019, and traffic accidents in Medellín (Colombia), during 2016. The data in Barcelona are provided by the local Town Hall (Open Data BCN¹), and are usable under license CC BY 4.0. The traffic accidents data in Medellín are published in the Open Data Portal of Medellín Town Hall², and were previously studied for the purpose of intensity estimation and spatially-varying relative risk by Moradi (2018) and Rakshit, Davies, et al. (2019).

Figure 4 shows, on the left panel, the locations of traffic accident occurred in Barcelona during 2019. The initial pattern had 10536 records, however, some were duplicated due to having more than a single collision type. Moreover, there have been several traffic accidents with the exact same location but e.g. different time occurrence. Since we are interested in detecting angle preferences in the spatial distribution of points, we do not consider these records, and it leads to a final pattern with 5700 points. The corresponding linear network is constructed by 25888 vertices and 31235 line segments. The total length of the network is 1377.444 km, and the maximum degree of the vertices is 7. The right panel in Figure 4 shows its estimated intensity, based on the quick kernel-based intensity estimator using 2D convolution (Rakshit, Davies, et al. 2019), after considering the Scott's rule-of-thumb to select the bandwidth which yielded 0.54 km (Scott 2015). It can be seen that the highest intensity is estimated at the "Eixample" district which is characterised by long straight streets, a strict pattern of square blocks crossed by wide avenues. Moreover, several touristic places are located in this district. In addition, apart from some areas in the north-west, north-east, and south-west which show proportionally lower intensities than other areas, the rest of the network apparently shows a similar intensity. Figure 5 shows the traffic accidents in Medellín during 2016. In this case, traffic accidents are further labelled based on accidents' severity by "fatal" (133 points), "personal injury" (6004 points), and "property damage" (4627 points). The road network has 54164 segments, 48194 vertices, maximum vertex degree 6, and total length 1244.202 km. The corresponding estimated intensities and relative risks can be found in Moradi (2018) and Rakshit, Davies, et al. (2019).

We now analyse if there is any tendency towards particular angles in the two datasets. We use kernel estimate of the circular density function of the nearest neighbour orientation function. For each dataset, we generate 99 angle-wise and spatially uniformly distributed patterns with the same number of points as the traffic data. Then we estimate the corresponding circular density function with a common bandwidth, and obtain the average of these 99 estimations. Note that in Section 4, we showed that the type of spatial interaction has a neglected effect on the null hypothesis. Left panel of Figure 6 shows the relative risk for traffic accidents in Barcelona using the common bandwidth 16.48 degrees; we can see there are slight tendencies towards 45, 135, 225, and 315 degrees. Note that angle 45 (225) belongs to the streets going from south-west to north-east, whereas angle 135 (315) corresponds to the streets going from south-east to north-west of Barcelona.

The corresponding relative risk for traffic accidents in Medellín, using as a common bandwidth 15.65 degrees, shows tendencies towards six main angles 15, 60, 120, 195, 240, and 300 degrees. Traffic accidents in Medellín show such angles more than if they were angle-wise uniformly distributed. Since traffic accidents in Medellín are labelled by their type of severity, we further show the relative risk of those causing fatality and personal injury relative to those causing property damage. From Figure 6, we can see that the relative risk of accidents involving fatality relative to those involving property damage (using common bandwidth 24.13 degrees) is higher for angles between 90 to 180 degrees than other angles, with some other peaks towards 20, 220, and 280 degrees. On the contrary, the relative risk of traffic accidents causing personal injuries relative to

¹<https://opendata-ajuntament.barcelona.cat/en/>

²<https://www.medellin.gov.co/geomedellin/index.hyg>

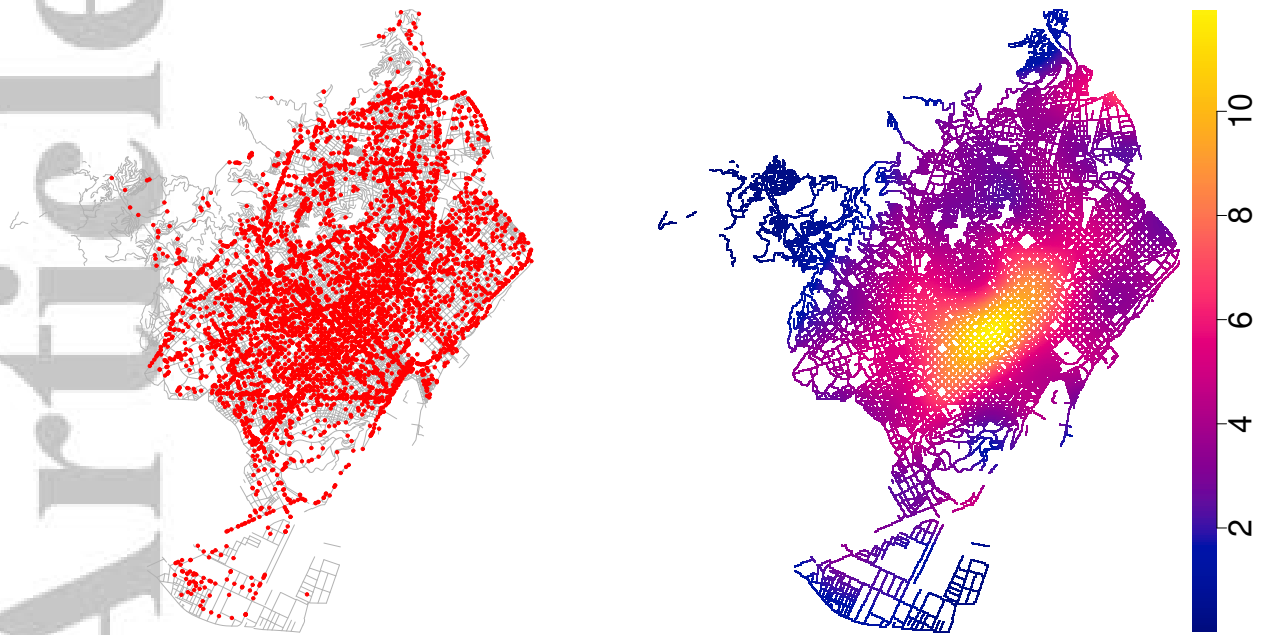


FIGURE 4 Traffic accidents in the city of Barcelona (Spain) during 2019. Left: Traffic accidents' locations. Right: Kernel-based estimated intensity of traffic accidents by the method of Rakshit, Davies, et al. (2019) and selected bandwidth using the Scott's rule of thumb (Scott 2015). Intensity values are accidents per kilometer.



FIGURE 5 Traffic accidents in Medellin (Colombia) during 2016, based on their severity type. Left: fatal, Middle: personal injury, Right: property damage.

those involving property damage (using common bandwidth 16.80 degrees) does not show any tendency towards any fixed angle, being uniformly distributed with respect to angles from 0 to 360 degrees. In other words, angle-wise speaking, they show a similar distribution.

6 | DISCUSSION

We have presented different tools to check directional preferences in the spatial distribution of points over linear networks. The main challenge is that the structure of the underlying network drives events to occur at particular fixed angles, i.e. the angles of the network. This, however, can not be taken into account as directional preference since events are forced to maintain such directions. Nevertheless, it is of great interest to check if

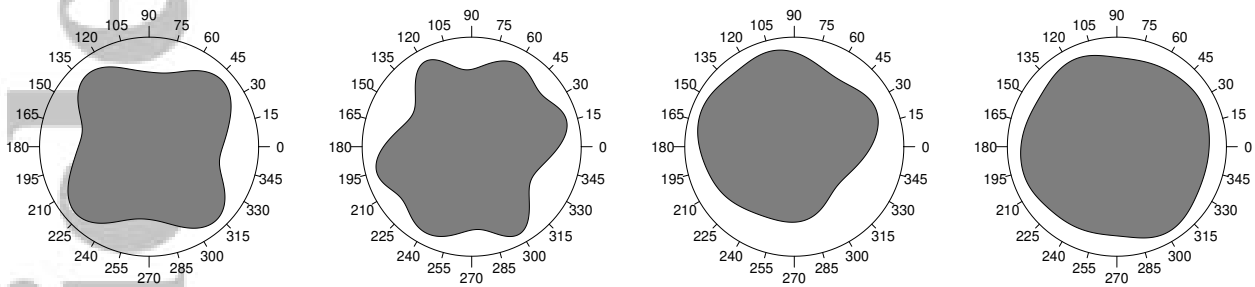


FIGURE 6 From Left to Right: estimated relative risk of traffic accidents in Barcelona relative to the case if they were angle-wise uniformly distributed; Estimated relative risk of traffic accidents in Medellin relative to the case if they were angle-wise uniformly distributed; Estimated relative risk of traffic accidents in Medellin causing fatality relative to those causing property damage; Estimated relative risk of traffic accidents in Medellin causing personal injury relative to those causing property damage.

there is any preference among the possible directions caused by the network, or any preference towards angles different than those derived by the geometry of the network.

We have presented the counterparts of the K-function-based orientation function and the nearest neighbour orientation function for linear networks. Since these functions are unavoidably affected by the geometry of the network, we have suggested to compare their estimators with those obtained based on angle-wise uniformly distributed patterns on the same network, and with the same number of points. Thereby, we can have an insight into the angle preferences caused by the inherent geometry of the network. We have then recommended to estimate the relative risk of the circular density function of the nearest neighbour orientation function of the data in question relative to that, on average, of the angle-wise uniformly distributed patterns. Therefore, we obtain a measure according to which we can uncover angle preferences amongst the main angles of the network or any other angle. We have seen that such relative risk is somehow not affected by the type of spatial interaction between points of the angle-wise uniformly distributed patterns. In contrast, the practicality of the K-function-based orientation function is, however, quite limited due to being influenced by the distance ranges (Baddeley et al. 2015, Chapter 7, pp. 239), and computationally expensive (Rakshit, Baddeley, & Nair 2019). Moreover, the type of interaction among points may also have secondary effect.

The problem of anisotropy on networks is still in its infancy. There are many points that still need to be solved. One is formally testing using circular statistics over the estimated circular densities. We have approached the problem from a more graphical point of view. Clearly statistical formality should reinforce our arguments. Another aspect is considering the evaluation of the strength of anisotropy. This could be a subsequent stage in the analysis. This comes together with considering anisotropic point process models for networks, an area completely open for research.

ACKNOWLEDGMENTS

J. Mateu has been partially funded by grants PID2019-107392RB-I00 from the Spanish Ministry of Science, AICO/2019/198 from Generalitat Valenciana, and UJI-B2018-04 from UJI, and C. Comas by grant MTM2017-86767-R from the Spanish Ministry of Science.

Calculations were performed in the R language, and our codes are accommodated in the R package `anisotropy1pp` available at <https://github.com/Moradii/anisotropy1pp> in order to facilitate the use of the corresponding developed methodologies.

Author contributions

The three authors contributed equally.

Conflict of interest

The authors declare no potential conflict of interests.

References

- Ang, Q. W., Baddeley, A., & 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.
- Baddeley, A., Møller, J., & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3), 329–350.
- Baddeley, A., Nair, G., Rakshit, S., & McSwiggan, G. (2017). "stationary" point processes are uncommon on linear networks. *Stat*, 6(1), 68–78.
- Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. CRC Press.
- Comas, C., Conde, J., & Mateu, J. (2018). A second-order test to detect spatio-temporal anisotropic effects in point patterns. *Statistics*, 52(4), 717–733.
- Comas, C., Rodriguez-Cortes, F., & Mateu, J. (2015). Second-order analysis of anisotropic spatiotemporal point process data. *Statistica Neerlandica*, 69(1), 49–66.
- Cronie, O., Moradi, M., & Mateu, J. (2020). Inhomogeneous higher-order summary statistics for point processes on linear networks. *Statistics and Computing*, 30, 1221 – 1239.
- Daley, D. J., & Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure* (Second ed.). Springer-Verlag New York.
- Hazelton, M. L. (2008). Kernel estimation of risk surfaces without the need for edge correction. *Statistics in Medicine*, 27(12), 2269–2272.
- Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.
- Moradi, M. (2018). *Spatial and Spatio-Temporal Point Patterns on Linear Networks* (PhD Dissertation). University Jaume I.
- Ohser, J., & Stoyan, D. (1981). On the second-order and orientation analysis of planar stationary point processes. *Biometrical Journal*, 23(6), 523–533.
- Quadros, A. (2019). emstreeR: Tools for fast computing and plotting euclidean minimum spanning trees [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=emstreeR> R package version 2.2.0.
- Rajala, T., Redenbach, C., Särkkä, A., & Sormani, M. (2018). A review on anisotropy analysis of spatial point patterns. *Spatial Statistics*, 28, 141–168.
- Rakshit, S., Baddeley, A., & Nair, G. (2019). Efficient code for second order analysis of events on a linear network. *Journal of Statistical Software*, 90(1), 1–37.
- Rakshit, S., Davies, T. M., Moradi, M., McSwiggan, G., Nair, G., Mateu, J., & Baddeley, A. (2019). Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review*, 87(3), 531–556.
- Rakshit, S., Nair, G., & Baddeley, A. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, 22, 129–154.
- Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis* (Vol. 26). CRC press.
- van Lieshout, M. N. M. (2011). A J-function for inhomogeneous point processes. *Statistica Neerlandica*, 65(2), 183–201.

How to cite this article: Moradi M., J. Mateu, and C. Comas (2020), Directional analysis for point patterns on linear networks, *Stat*, 2020;00:1–11.