Citation
Baddeley, A. and Nair, G. and Rakshit, S. and McSwiggan, G. and Davies, T. 2021. Analysing point patterns on networks — A review. Spatial Statistics. 42: pp. 1-35.
http://doi.org/10.1016/j.spasta.2020.100435

Analysing point patterns on networks — a review

Adrian Baddeley[a,b,*], Gopalan Nair[c], Suman Rakshit[d], Greg McSwiggan[c], Tilman M. Davies[e]

[a] *School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Perth, Australia*
[b] *Data61, CSIRO, Perth, Australia*
[c] *School of Mathematics and Statistics, University of Western Australia, Perth, Australia*
[d] *SAGI-West, School of Molecular and Life Sciences, Curtin University, Perth, Australia*
[e] *Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand*

Abstract

We review recent research on statistical methods for analysing spatial patterns of points on a network of lines, such as road accident locations along a road network. Due to geometrical complexities, the analysis of such data is extremely challenging, and we describe several common methodological errors. The intrinsic lack of homogeneity in a network militates against the traditional methods of spatial statistics based on stationary processes. Topics include kernel density estimation, relative risk estimation, parametric and non-parametric modelling of intensity, second-order analysis using the K-function and pair correlation function, and point process model construction. An important message is that the choice of distance metric on the network is pivotal in the theoretical development and in the analysis of real data. Challenges for statistical computation are discussed and open-source software is provided.

Keywords: distance metric, kernel density estimation, K–function, nonparametric estimation, pair correlation function, point process model, stationary process.
2010 MSC: 62H11, 62M30

*Corresponding author
Email address: adrian.baddeley@curtin.edu.au (Adrian Baddeley)

1. Introduction

This paper reviews recent research on the spatial analysis of events that occur along a network of lines. Figure 1 displays a motivating example, in which the locations of traffic accidents in a city are mapped together with the road network. Such data require the development of novel statistical methodology and computational techniques (Okabe and Sugihara, 2012; Ver Hoef et al., 2006; Baddeley et al., 2015, Chapter 17).

Figure 1: Traffic accidents recorded in 2011 in the central business district (CBD) of the city of Perth, Western Australia. Black dots are accident locations; grey lines are state roads.

Spatial patterns of points along a network of lines are found in many other applications. The network might reflect a map of railways, rivers, electrical wires, nerve fibres, airline routes, irrigation canals, geological faults or soil cracks. Observations of interest could be the locations of traffic accidents, bicycle incidents, vehicle thefts or street crimes (Yamada and Thill, 2004; Lu and Chen, 2007; Xie and Yan, 2008; Ang et al., 2012; Vandenbulcke-Plasschaert, 2011); roadside trees or invasive species (Spooner et al., 2004; Deckers et al., 2005); retail stores, roadside kiosks or urban parks (Okabe and Kitamura, 1996; Okabe and Okunuki, 2001; Okunuki and Okabe, 2003; Comber et al., 2008); insect nests (Voss, 1999; Voss et al., 2007; Ang et al., 2012); neuroanatomical features (Yadav et al., 2012; Jammalamadaka et al., 2013; Baddeley et al., 2014) or sample points along a stream (Ver Hoef et al., 2006; Ver Hoef and Peterson, 2010; Som et al., 2014). John Snow's pioneering observations of cholera cases in London (Snow, 1855) could also be described as a point pattern on a linear network. Spatial analysis of such data can have immediate practical value, as it did when Snow's analysis suggested the cause of cholera transmission.

Statistical analysis of network data presents severe challenges. A network is not spatially homogeneous, which creates geometrical and computational com-

plexities and leads to new methodological problems, with a high risk of methodological error. Real network data can also exhibit an extremely wide range of spatial scales. These problems pose a significant and far-reaching challenge to the classical methodology of spatial statistics based on stationary processes, which is largely inapplicable to data on a network.

Analysis of point patterns on linear networks has been a focus of recent research in Geographical Information Systems (GIS) (Borruso, 2005, 2008; Downs and Horner, 2007a; Okabe and Yamada, 2001; Okabe et al., 2009, 2008, 2000; Shiode, 2008; Okabe and Satoh, 2006; Okabe et al., 1995, 2006b,a; Okabe and Satoh, 2009; Shiode and Shiode, 2009; Warden, 2008; Yamada and Thill, 2007). For surveys, see Okabe and Satoh (2009); Okabe and Sugihara (2012). In recent years there has been increased interest from the spatial statistics community (Ang et al., 2012; Baddeley et al., 2014; McSwiggan et al., 2016; Baddeley et al., 2017; Anderes et al., 2017; Rakshit et al., 2017; Moradi et al., 2018; Briz-Redón et al., 2019; Rakshit et al., 2019a,b; McSwiggan et al., 2019; Moradi et al., 2019; McSwiggan, 2019, Baddeley et al., 2015, Chap. 17). Spatial analysis on a network of rivers or streams is also a highly active research topic in spatial ecology and spatial statistics, although this is more focused on geostatistical techniques for spatial variables (Cressie and Majure, 1997; Cressie et al., 2006; Isaak et al., 2014; O'Donnell et al., 2014; Ver Hoef et al., 2006; Ver Hoef and Peterson, 2010).

Section 2 presents a range of example datasets and applications. Section 3 surveys the main statistical challenges for analysis of point patterns on a network. Section 4 gives some basic technical definitions. Section 5 discusses kernel smoothing on a network. Section 6 discusses nonparametric estimation of intensity as a function of covariates. Section 7 discusses parametric modelling of point processes, model-fitting and variable selection. Section 8 discusses the analysis of clustering using the K–function and related tools. Section 9 describes the fundamental problem of existence of models with specified properties. Section 10 presents alternative ways of measuring the distance between locations on a network, and the implications for statistical analysis and modelling.

2. Data examples

In this section we present datasets from several different applications, illustrating different features and challenges. Various techniques will be demonstrated on these datasets throughout the paper. Where possible, the datasets are publicly available in the R packages `spatstat` (Baddeley and Turner, 2005; Baddeley et al., 2015) or `spatstat.Knet` (Rakshit et al., 2019a).

2.1. Western Australia traffic accidents

The traffic accident pattern shown in Figure 1 is part of a much larger dataset giving the locations of all traffic accidents recorded in the State of Western Australia over several years. The data, curated by the state government agency Main Roads WA, include spatial coordinates of each road segment; the spatial location of each accident; properties of the road, such as speed limit and curvature; and attributes of each accident, such as severity and time of day.

Figure 2 shows the accidents recorded in the Perth metropolitan area for the year 2011, delimited by a 59×67 km rectangle, comprising 12,408 accident locations on a road network of total length 10,319 km.

Figure 2: Accidents recorded in the Perth metropolitan area.

The full dataset for the year 2011 is shown in Figure 3; it gives the locations of 14,562 traffic accidents for the entire state of Western Australia on a road network of 115,169 segments with total length 97,166 km. This is available in the `spatstat.Knet` package as the dataset `WAcrashes`. Both the network and the accident observations are extremely dense in the urban area along the west coast; the raw data also reveal major arterial highways.

Questions of interest include the spatial variation of accident risk, the relationship between accident risk and road characteristics, and the effect of safety management interventions such as the installation of traffic lights.

2.2. Chicago crimes

Figure 4 is a record of 116 crimes (classified into 7 types) reported in a neighbourhood of the University of Chicago. The locations are the street addresses associated with each crime report.

Questions of interest include whether there are spatial concentrations of crime, and whether the spatial pattern is different for different types of crime (Johnson, 2010).

Figure 3: Traffic accidents (black dots) recorded in the year 2011 on the state road network (grey lines) of the southern half of Western Australia.

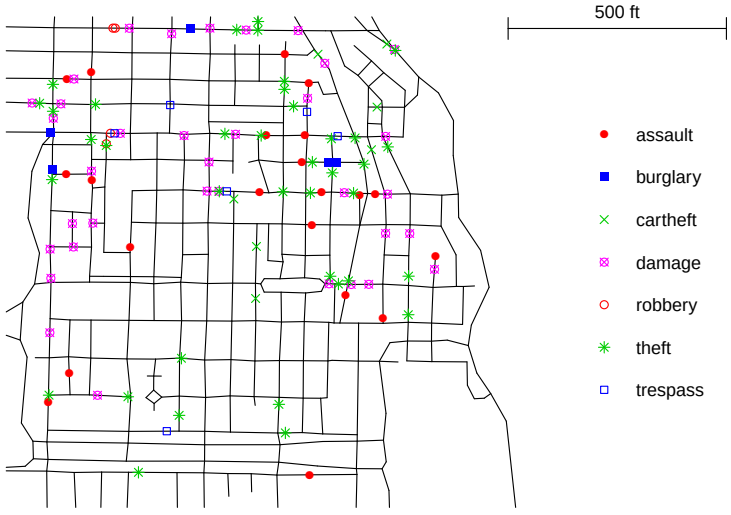

Figure 4: Chicago crimes data, classified by type of crime (Chicago Weekly News).

2.3. Spider shelter webs — replicate observations

Figure 5 shows the positions of spider webs of the urban wall spider *Oecobius navus* on the mortar lines of a brick wall, observed on six successive Thursdays, recorded by Voss (1999); Voss et al. (2007) and digitised by Mark Handcock.

Observations were confined to a square of side 1.125 metres; the linear network has 156 vertices and a total length of 20.22 metres. Questions concern the spiders' habitat preferences (Voss et al., 2007) and interaction between nearby individuals. The spiders are free to roam over the wall surface, but the mortar spaces provide the only opportunity for constructing webs (Voss, 1999).

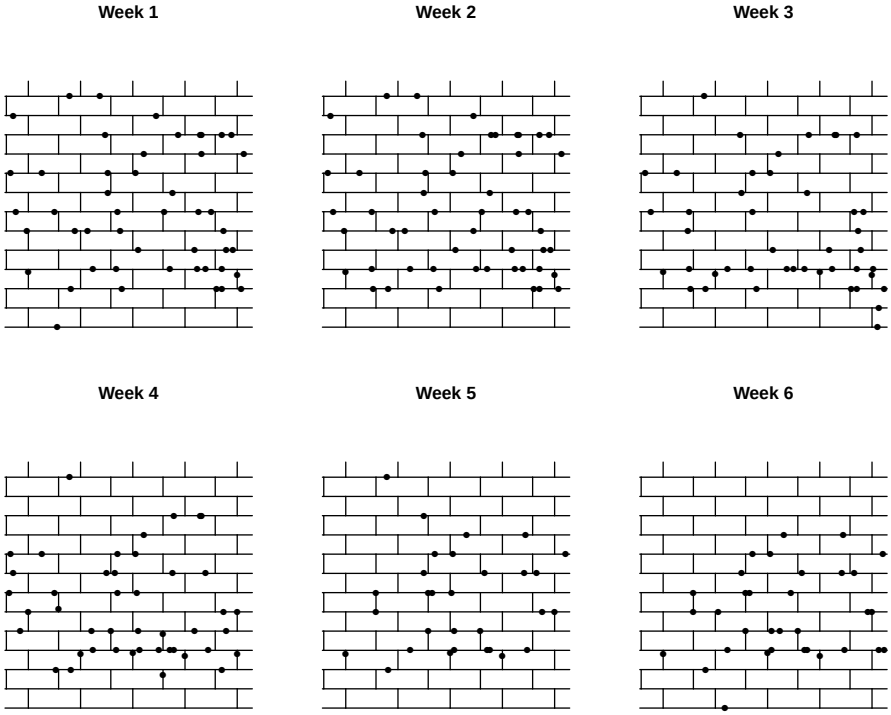

Figure 5: Spider shelter webs on a brick wall, observed once a week for six weeks.

The data in the top left panel of Figure 5 were analysed in Ang (2010); Ang et al. (2012); Rakshit et al. (2017), Baddeley et al. (2015, Chap. 17) and are available in the `spatstat` package as the dataset `spiders`. The full sequence of observations shown in Figure 5 was analysed in Voss (1999); Ang (2010); Voss et al. (2007). The availability of replicated observations of a point pattern allows a more searching analysis; methods for this purpose are well-developed for two-dimensional point patterns (Baddeley et al., 1993; Bell and Grunwald, 2004; Diggle et al., 1991, 2000; Mateu, 2001; Myllymäki et al., 2013; Webster et al., 2005; Baddeley et al., 2015, Chap. 16) but for linear networks we demonstrate them for the first time in this paper.

2.4. Dendritic spines

Figure 6 shows the locations of small protrusions called *spines* on part of the dendrite network of a neuron (nerve cell) recorded by the Kosik Lab (UCSB) and Aruna Jammalamadaka. The spines are also classified into three types, based on their shape. This neuron was grown *in vitro*, lying almost flat, and Figure 6 shows a two-dimensional projection. Questions of interest include whether the spines are uniformly distributed; whether different spine types have different

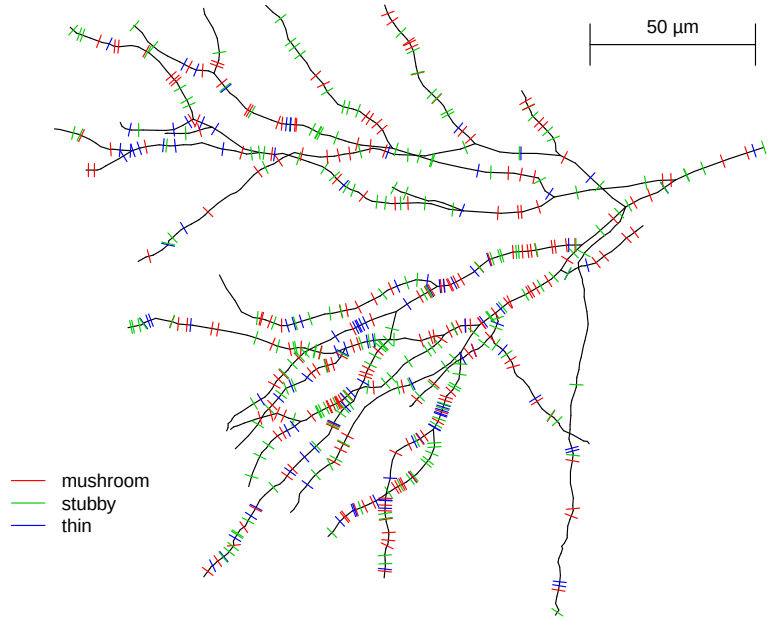

Figure 6: Dendritic spine locations on the dendrite network of a neuron. Grey lines show dendrite network. Coloured tickmarks show spine locations; tickmarks are not to scale. The root of the dendritic tree is at the top right.

distributions; and whether spines are clustered or randomly located. These data were analysed by Jammalamadaka et al. (2013); Baddeley et al. (2014) and are available in the `spatstat` package as the dataset `dendrite`.

2.5. Geelong road accidents

The upper panel of Figure 7 shows the locations of high-severity road traffic accidents in the Australian regional city of Geelong. The lower panel shows the traffic volume along each road segment, measured as the average daily total number of cars that triggered a counting device on the road segment. The traffic volume is an example of an explanatory variable (spatial covariate) that should be included in any realistic analysis of accident risk (McSwiggan, 2019).

2.6. Subtleties and differences between applications

To avoid methodological errors, it is important that network data should not be viewed simply as a "type of data" for which we can develop software solutions. Each application has its own characteristics and exigencies requiring attention.

Spatial point process methods are appropriate only when the spatial locations are recorded. In road accident research, the precise spatial locations of accidents are recorded only for high-severity accidents, which are investigated by authorities; low-severity accidents are often self-reported (and under-reported) at a much lower spatial precision, often suitable only for use as aggregated count data.

An important question about all network data is to understand the relevance of the space "outside" the network. This varies between applications. The

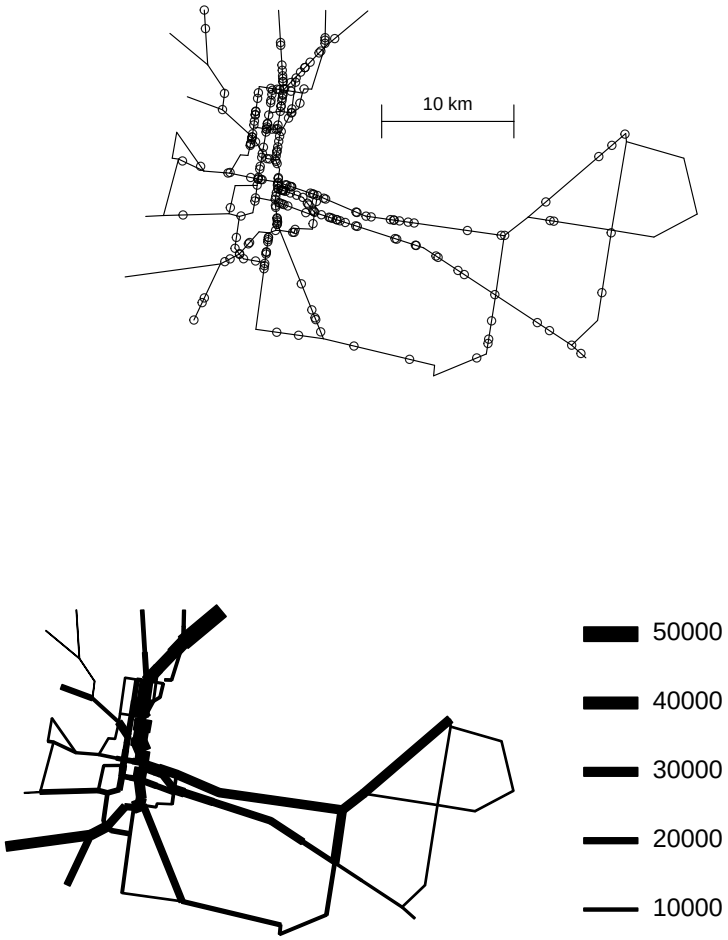

Figure 7: Geelong accident data. *Top:* High severity traffic accidents on state-declared roads in Geelong, Australia, 2009–2011. *Bottom:* traffic volume (vehicles per day) along the network; line widths are proportional to traffic volume according to the scale at right, in the style of Xie and Yan (2008). Downloaded from CrashStats interactive database at `www.vicroads.vic.gov.au`.

spider webs in Figure 5 are built in the mortar lines of the brick wall, but the spiders are otherwise free to roam across the wall, unlike road vehicles, which are constrained to stay on the road network. River ecology depends on the surrounding landscape ecology; dendritic spines are electrical connections into the surrounding tissue. This information could be used for better inference (Rakshit et al., 2017).

Point locations recorded on a network may be the true locations of events (such as severe road accidents) or they may be the projections of the true spatial locations onto the network (such as the street addresses of crime reports in the Chicago data, Figure 4).

Figure 8: Road accidents in the inner city of Fremantle, Western Australia.

Road traffic accidents frequently occur at a road intersection. Figure 8, provided by Mr Isaac Gravestock, displays traffic accident locations in the port city of Fremantle, Western Australia showing a high proportion of accidents at intersections. In many point process models, there is zero probability that a random point will occur at a predetermined fixed location. Point process models need to be modified to allow points to occur exactly at a vertex of the network (McSwiggan, 2019). Alternatively, count regression models can be used by introducing a separate accident count at each intersection (Briz-Redón et al., 2019).

Most networks are constantly changing through engineered, biological or ecological processes. Existing techniques treat the network as fixed, which is a reasonable approximation for short periods, but a spatio-temporal approach would be more powerful.

There are methodological challenges with the data sources. Institutional spatial databases have constraints such as inclusion criteria, privacy and record-keeping laws, which can introduce subtle biases of Berkson's (1946) type. Changes in a network often *depend on* the events of interest, such as road improvements in response to accident patterns. When a dangerous road intersection is modified to improve safety, some authorities delete the old accident records, because they are associated with a road segment that no longer exists. This introduces confounding (making it impossible to demonstrate the benefit of the road modification) or Berkson bias (producing spurious associations). Spatio-temporal analysis could resolve these problems, provided the old data are retained.

Linkage of data in different databases is needed. Road networks are typically partitioned into several sub-networks, administered by different levels of government. Accident databases are often separate from databases recording the road network and road condition. Traffic volume is an important explanatory variable which is often available only for one class of road.

River and stream networks, and other branching networks such as the dendrite network in Figure 6, raise a different class of questions, and are amenable to different kinds of analysis, from the general network. Typical questions include whether different branches of the network have different characteristics, and whether statistical properties change progressively downstream.

3. Overview of problems and significance

In this section we survey the main problems for analysis of point patterns on a linear network. For concreteness we shall often use terms that would be appropriate to road traffic accident analysis.

3.1. The need for truly spatial analysis

For agencies which build and manage road networks, the main question of interest in traffic accident analysis is the relationship between accident risk and road design characteristics such as speed limit, road curvature, road width, number of lanes, presence of a kerb, and presence of traffic lights. These questions require some kind of spatial analysis.

A pragmatic strategy is to aggregate the data by counting the number of accidents that occurred along each segment of road. The road network may be divided into segments in any arbitrary fashion, but ideally in such a way that the accident risk is approximately constant along a given segment. Count regression models can then be used to model the dependence of accident frequency on explanatory variables, provided these data are available for each road segment (Lord and Mannering, 2010).

Potential weaknesses of this *crash frequency* approach include substantial bias due to aggregation (the "ecological fallacy" or "modifiable unit area problem") and the loss of spatial information needed to assess evidence for clustering. It has been reported that accident count data are often over-dispersed and zero-inflated, relative to Poisson regression (Lord et al., 2005). This could be partly attributable to aggregation bias. A more subtle problem is that the crash frequency approach favours the use of covariates which are approximately constant along each road segment, to the exclusion of potentially important covariates such as sighting distance or distance to the nearest road intersection (McSwiggan, 2019). It could be argued that current crash frequency methodology is not capable of correctly identifying "black spots" of high localised accident risk.

When the available data include the exact spatial location of each road accident, it would be highly desirable to have techniques for spatial analysis of the accident locations, analogous to existing methods for analysing spatial point patterns in two dimensions (Baddeley et al., 2015; Diggle, 2014; Gelfand et al., 2010; Illian et al., 2008). This is the topic of this paper.

3.2. Taking account of the network

The core problem is that — since traffic accidents can only occur on roads — a correct spatial analysis of traffic accident locations must take account of the layout of the road network.

It would clearly be incorrect to simply ignore the road network, retaining only the spatial coordinates of the accident locations, and analysing them as a spatial point pattern in two-dimensional space using the standard methods of spatial statistics (Baddeley et al., 2015; Diggle, 2014; Gelfand et al., 2010; Illian et al., 2008). Using only the two-dimensional point locations, a kernel estimate of the spatially-varying density of accidents in two dimensions (Diggle, 1985; Bithell, 1990) would give spuriously high density values in areas where the road network is more dense. An estimate of Ripley's K-function for two-dimensional point patterns (Ripley, 1977) could give spurious evidence of short-range clustering (due to concentration of points along each road) and long-range

regularity (due to spatial separation of different roads). This has been observed in real data on traffic accidents (Yamada and Thill, 2004) and urban crime (Lu and Chen, 2007).

A linear network is not a homogeneous space: the geometry of the network is different at each location. Consequently, even elementary statistical tools can be difficult to implement, and must be "adjusted" to compensate for the spatially-varying geometry of the network. Kernel smoothing of point events, which is simple to define and very fast to compute in two dimensions (Diggle, 1985; Silverman, 1982), is mathematically complicated and time-consuming to perform on a network (Okabe et al., 2009, Okabe and Sugihara, 2012, Chap. 9, McSwiggan et al., 2016). The correlation analysis of point patterns is straightforward in two dimensions using Ripley's K–function or the pair correlation function (Baddeley et al., 2015, Sec. 7.6, Illian et al., 2008, pp. 218–244) but on a network, the definition of the K–function is more intricate (Okabe and Yamada, 2001; Ang et al., 2012; Baddeley et al., 2014) and its theoretical justification is weaker (Baddeley et al., 2017).

The problems of spatial analysis on a network have far-reaching relevance for the discipline of spatial statistics. Many spatial phenomena, although observed in two- or three-dimensional space, are confined to a subset of the space, either by preference for a supportive "substrate", or by a physical constraint. A linear network is the simplest non-trivial example of a substrate.

3.3. Methodological errors and challenges

A popular strategy for data analysis of point patterns on a linear network is to take existing *computational* procedures, originally developed in statistical science for analysing point patterns in two-dimensional space, and adapt or transfer them directly to the new setting of a linear network.

This approach implicitly assumes that the *statistical* basis for such procedures is also transferable from the two-dimensional plane to the linear network. Unfortunately, this is not always established. Consequently, the practical interpretation and statistical validity of these techniques are open to question. In the worst case, they could lead to fallacious conclusions (Yule, 1903; Berkson, 1946; Andersen, 1990).

3.3.1. Kernel density estimation

A natural first step in analyzing road accident locations is to form a kernel density estimate (Silverman, 1986) of the spatially-varying accident rate. However, this is not straightforward on a linear network. A common fallacy is to take the kernel density estimate on the one-dimensional real line

$$\widehat{f}(u) = \frac{1}{n}\sum_{i=1}^{n} k(u - x_i), \tag{1}$$

and translate it directly to the linear network as

$$\widehat{f}(u) = \frac{1}{n}\sum_{i=1}^{n} k(d_L(u, x_i)), \tag{2}$$

where u is any location on the network, $x_1, \ldots, x_n$ are the observed data locations, $d_L(u, x_i)$ is the shortest-path distance in the network from u to x_i, and k is a smoothing kernel (probability density) *on the real line*.

Okabe et al. (2009), Okabe and Sugihara (2012, p. 180) and McSwiggan et al. (2016) pointed out that (2) is a fallacious estimate, because it does not conserve mass: the induced kernel $K(\cdot, x_i) = k(d_L(\cdot, x_i))$ is not a probability density on the linear network; the density estimate (2) is not a probability density. The true probability density will be grossly overestimated in the denser parts of the network, leading to artefacts. We discuss this in Section 5.

3.3.2. K–function

Spatial clustering of points in two dimensions is often investigated using estimates of Ripley's K–function (Ripley, 1977) or the pair correlation function (see e.g. Illian et al., 2008). The K–function of a completely random pattern is easily recognisable; two datasets from different survey regions can be compared using their K–functions. For point patterns on a network, Yamada and Thill (2004); Lu and Chen (2007) drew attention to the "false alarm" (spurious evidence of clustering) which may occur if these computational tools are applied to the spatial coordinates ignoring the network. Okabe and Yamada (2001) proposed a modification of (29) in which distances between pairs of points x_i, x_j are measured using the shortest-path distance $d_L(x_i, x_j)$ along the network, instead of the Euclidean distance $\|x_i - x_j\|$. This accounts for the geometry at small scales, but does not adjust for the fact that different spatial locations on the network are surrounded by different configurations of lines. The Okabe-Yamada K–function is highly dependent on the network geometry and cannot be used to compare the spatial pattern of road accidents in two different cities.

A correction for the local configuration of the network was proposed by Ang et al. (2012) and some basic statistical properties were established. This is discussed in Section 8.

3.4. Inapplicability of classical tools

A point pattern can be modelled as the outcome of a random *point process*. Probability theory for point processes on any space is highly-developed and highly flexible (Daley and Vere-Jones, 2003, 2008). Statistical methodology for spatial point processes is also available (Diggle, 2014; Gaetan and Guyon, 2009; Gelfand et al., 2010; Møller and Waagepetersen, 2004). This would suggest that the analysis of point patterns on a network can easily be handled by adapting existing statistical tools. Unfortunately that is not the case.

The most popular statistical techniques for spatial point patterns in two dimensions rely heavily on assuming that the random point process which generated the pattern is *stationary*, meaning that statistical properties are the same at all spatial locations (Illian et al., 2008). For example, Ripley's function $K(r)$ is defined as (a multiple of) the expected number of random points falling within a distance r of a "typical point" (Ripley, 1977; Baddeley et al., 2015, Sec. 7.3.2). In order for this definition to be meaningful, the expected number must not depend on spatial location, so we are implicitly assuming that the underlying point process is stationary, at least for the first and second moments (Ripley, 1976, 1977, 1981, 1988; Illian et al., 2008). Similar comments apply to the pair correlation function (Illian et al., 2008).

In two dimensions, the assumption of a stationary point process is often a reasonable working approximation, and is not very restrictive. The class of stationary point processes in two dimensions is large; there are many simple devices

for constructing them; and they arise in many plausible scenarios. This richness justifies the use of the K–function and allows us to call it a "non-parametric" property (Diggle, 2010). The scope of application of the K–function has been extended even further, to inhomogeneous point processes with stationary correlation structure (Baddeley et al., 2000).

Unfortunately, it is not possible to define stationary processes on a linear network (in the strict sense). The network itself is not homogeneous, that is, different neighbourhoods have different geometry. This cuts off one of the main routes to developing statistical methodology. For example, there is no analogue for linear networks of the empty space function F and nearest-neighbour distance function G (Illian et al., 2008, Baddeley et al., 2015, Chap. 8).

Some progress is possible for first and second moment statistics. Okabe and Yamada (2001) and Ang et al. (2012) adapted the K–function and pair correlation function to point patterns on a linear network. Distance between points is measured by the shortest path in the network. Just as in the two-dimensional case, in order for the K–function and pair correlation function to be well-defined, the point process is required to have a stationary correlation structure, in the sense that the pair correlation is a function of shortest-path distance (Baddeley et al., 2015, Chapter 17).

However, we recently discovered that it may be unreasonable to assume that a point process on a network has stationary correlation structure, unless the network is a tree. Baddeley et al. (2017) consider some of the standard recipes for constructing point process models, and show that many of these constructions do not produce a stationary correlation structure. The findings suggest that such processes may be quite rare on a general network. Indeed Anderes et al. (2017) prove that stationary correlation structures do not exist on any network which contains certain kinds of loops. This severely weakens the rationale for using the K–function and pair correlation function on a network in real applications. Modelling and inference also become much more complicated.

This calls for a fundamental shift in statistical methodology for spatial point processes. We must abandon the cherished assumption of stationarity — which means relinquishing about half of all existing statistical techniques for spatial point patterns — and develop a fundamentally new approach that applies to nonstationary point processes on inhomogeneous spaces.

3.5. Multiple scales and local analysis

The analysis of real network data often involves a wide range of spatial scales — which is another argument against the assumption of stationarity.

The Western Australian road accidents shown in Figure 3 are highly concentrated in the city of Perth, in a few regional towns, and along major highways. A novel complication is that the road network itself has huge spatial variation, being highly concentrated around metropolitan Perth and the west coast. This makes it much more difficult to assess patterns by eye. It also causes computational difficulties, since a much finer spatial resolution is required in urban areas than in the remote desert. The data in Figures 1, 2 and 3 have average densities of 11 km, 2.6 km and 0.09 km (respectively) of road per sq km of land area. There are, respectively, 5.4, 3.1 and 0.15 accidents per km of road. Different accident processes are happening in a city block, on a busy freeway, and on a remote highway. The three regions shown in by Figures 1–3 deserve to

be analysed separately, or the entire dataset should be analysed using localised methods such as adaptive smoothing (Abramson, 1982; Hall and Marron, 1988; Davies and Hazelton, 2010; Davies and Baddeley, 2018; Davies et al., 2016), Local Indicators of Spatial Association (Anselin, 1995; Cressie and Collins, 2001; Yamada and Thill, 2007), or geographically weighted regression and local likelihood (Fotheringham et al., 2003; Baddeley, 2017). Pragmatic alternatives may include constructing spatial covariates which describe the local configuration of the network, such as the spatial density of lines (Borruso, 2008).

3.6. Space-time

Network point patterns often evolve over time, and their analysis should include the time coordinates (Cressie and Wikle, 2011; Diggle, 2014). The time of day at which a road accident occurred is an important covariate; it determines the speed limit and road rules applicable at the time, and is a surrogate for unobserved variables such as traffic conditions and weather. The calendar date is likewise a surrogate for seasonal effects. Accident risk changes over short time scales due to changes in traffic conditions. Spatio-temporal analysis of network data is a highly fertile research area.

3.7. Computational challenges

The geometrical complexity of networks causes many computational challenges. Data structures and algorithms must be designed for efficient and correct computation of geometrical quantities, shortest-path distances between data points, identification of nearest neighbours, statistical summaries, numerical integrals on the network, subdivisions or modifications of the network, and stochastic simulation. Software for these purposes includes `SANET` (Okabe et al., 2006b,a; Okabe and Sugihara, 2012) and the open-source R packages `spatstat` (Baddeley et al., 2015, Chap. 17) and `spatstat.Knet` (Rakshit et al., 2019a). The effort required to develop such code has been very substantial.

Point pattern data on a network may involve a very large number of variables, both explanatory spatial variables and attributes of individual points. There are growing opportunities to collect real-time data on traffic flow through crowdsourcing apps, along with real-time contextual information such as weather radar. Consequently the analysis of network data is increasingly viewed through the lens of "big data" and predictive analytics.

3.8. Distance metrics

The analysis of spatial clustering and correlation on a linear network also depends crucially on how we measure the distance between points. Standard practice is to measure distance by the length of the shortest path. However, this is not obligatory, and may be inappropriate in some applications (Okabe and Sugihara, 2012, p. 9).

Other choices of distance metric include Euclidean distance, weighted metrics (with a separate cost factor for each edge), directed distances (one-way roads, river systems) and flow-based metrics (river systems, traffic flows, electrical resistance). The choice of distance metric involves implicit model assumptions and should be considered carefully.

Choosing a metric other than the shortest-path distance may help to resolve many technical difficulties. Rakshit et al. (2017) developed versions of the K–function on a network with respect to any distance metric. There is a rich supply

of point processes on a network which have stationary correlation structure with respect to Euclidean distance (Baddeley et al., 2017; Anderes et al., 2017) and possibly with respect to electrical resistance distance (Bapat, 2004).

4. Technical definitions

Here we collect a few technical definitions and preliminaries.

4.1. Point patterns on a network

Define a *linear network* L as the union of N line segments l_i such that $L = \cup_{i=1}^{N} l_i$ and $N < \infty$. Each line segment takes the form $l_i = [u_i, v_i] = \{w : w = tu_i + (1-t)v_i,\ 0 \le t \le 1\}$, where $u_i, v_i \in \mathbb{R}^2$ are the endpoints of l_i. The intersection of any two segments l_i and l_j for $i \neq j$ is assumed to be either empty or an endpoint of both segments. We denote by $|B|$ the total length of a subset $B \subseteq L$.

A finite *point pattern* $\mathbf{x}$ on a linear network L is a finite unordered set $\mathbf{x} = \{x_1, \ldots, x_n\}$, where each point x_i represents a location on L, and the number of points n is not fixed in advance.

4.2. Random point processes

An observed point pattern $\mathbf{x}$ will be regarded as the outcome of a random point process $\mathbf{X}$ on L. For general definitions, see Daley and Vere-Jones (1988). We assume that the total number of points is finite with probability 1, that the total number of points has finite mean and finite variance, and that there are no multiple coincident points (Daley and Vere-Jones, 2003, Chap. 5).

The point process $\mathbf{X}$ has *intensity function* $\lambda(u)$, $u \in L$, if

$$E[N(\mathbf{X} \cap B)] = \int_B \lambda(u)\, \mathrm{d}u, \tag{3}$$

for all closed subsets $B \subset L$, where $N(\mathbf{X} \cap B)$ is the number of points of $\mathbf{X}$ falling in B, and $\mathrm{d}u$ denotes integration with respect to arc length on the network. Thus an intensity function on the network has values with dimension length^{-1}, points per unit length of network, and we may interpret $\lambda(u)$ as yielding the expected number of points per unit *length* of network in the vicinity of location u.

4.2.1. Shortest-path distance

Distances in a network will often be measured by the shortest path in the network. 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 such 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 $x_0, x_1, \ldots, x_m$ is defined to be $\sum_{i=0}^{m-1} \|x_{i+1} - x_i\|$, where $\|\cdot\|$ denotes Euclidean distance. 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.

The *disc* (in the shortest-path metric), with radius $r > 0$ and centre point u, in the network L is the set of all points v in the network that lie no more than a distance r from the location u, in the shortest path distance:

$$b_L(u, r) = \{v \in L :\ d_L(u, v) \le r\}.$$

Figure 9: A "disc" of radius $r = 2$ km in the Perth CBD road network defined using the shortest-path distance metric. Filled circle: centre point u. Thick lines: disc $d_L(u, r)$. Open circles: points counted in the circumference $m(u, r)$.

An important quantity in this paper is the *circumference* $m(u, r)$, which is the number of points v on the network satisfying $d_L(u, v) = r$. These concepts are illustrated in Figure 9.

5. Kernel density estimation

Estimation of the spatially-varying density of events is crucially important in practice. In studying road safety or transport planning, for example, such estimates are essential for accident research, the formulation of emergency response strategies, urban modelling and other purposes. Even when it is not the main focus of interest, we may need to adjust for spatially-varying density in order to study other properties. Non-uniform density can easily be confounded with clustering between points, (Baddeley et al., 2015, Chaps. 7, 8, 12) as demonstrated by the analysis of the dendritic spines data (Baddeley et al., 2014).

The goal of density estimation is to statistically estimate the spatially-varying density from an observed point pattern $\mathbf{x}$ with only minimal assumptions about the underlying point process. Kernel density estimation (Silverman, 1985; Wand and Jones, 1995) is arguably the most popular technique. For spatial point pattern data in two dimensions, kernel estimates are conceptually simple (Diggle, 1985; Bithell, 1990), and can be computed rapidly using the Fast Fourier Transform (FFT) (Silverman, 1982). However, kernel estimation on a network is mathematically and computationally intricate; many different techniques have been proposed (Borruso, 2003, 2005, 2008; Downs and Horner, 2007a,b, 2008; Xie and Yan, 2008; Okabe et al., 2009; Sugihara et al., 2010;

Okabe and Sugihara, 2012, Chap. 9; Anderson, 2009; McSwiggan et al., 2016; Moradi et al., 2018; Rakshit et al., 2019b).

5.1. Kernel sums

For real-valued observations $x_1, \ldots, x_n$ the classical (fixed-bandwidth) kernel estimator of probability density is given in (1). The corresponding estimator for the *intensity* function $\lambda(u)$ is obtained by omitting the normalising factor $1/n$:

$$\widehat{\lambda}(u) = \sum_{i=1}^{n} k(u - x_i). \tag{4}$$

For point events $x_1, \ldots, x_n$ on a linear network L, it may seem plausible that we could adapt the kernel estimation procedure to the network by replacing the vector difference $u - x_i$ by the shortest path distance $d_L(u, x_i)$, giving an intensity estimate

$$\widehat{\lambda}(u) = \sum_{i=1}^{n} k(d_L(u, x_i)), \tag{5}$$

where k is still a smoothing kernel on the real line (e.g. Xie and Yan, 2008). This implicitly assumes that the statistical basis for kernel estimation can be transferred from the real line to the linear network. Unfortunately, that is not true. The estimator (5) is severely biased. The total contribution from a point x_i to the intensity estimate is the integral $\int_L k(d_L(u, x_i))\,\mathrm{d}u$, which may be substantially different from 1. Thus (5) is highly susceptible to artefacts (Okabe et al., 2009; Sugihara et al., 2010; Okabe and Sugihara, 2012, p. 180; McSwiggan et al., 2016). These problems can be avoided on river and stream networks (Cressie and Majure, 1997; Ver Hoef et al., 2006; Cressie et al., 2006; Ver Hoef and Peterson, 2010; O'Donnell et al., 2014) where there is a unique shortest path between any two points which are connected.

5.2. Corrected kernel sums

Several modifications of the naive estimator (5) have been proposed in order to reduce its severe bias (Borruso, 2005, 2008; Okabe et al., 2009, Sec. 3).

Moradi et al. (2018) and McSwiggan (2019) pointed out that the bias can be removed by adapting classical edge corrections from spatial statistics. For a two-dimensional spatial point pattern observed inside a restricted survey region $W \subset \mathbb{R}^2$, Diggle (1985) proposed the edge-corrected estimator of intensity

$$\widehat{\lambda}^{\mathrm{U}}(u) = \frac{1}{c_W(u)} \sum_{i=1}^{n} \zeta(u - x_i), \quad u \in W, \tag{6}$$

where ζ is a *two-dimensional* kernel (a probability density on the two-dimensional plane), and

$$c_W(u) = \int_W \zeta(u - v)\,\mathrm{d}v, \quad u \in W, \tag{7}$$

is the mass of the kernel centred at u that falls inside the window (see also Bithell, 1990). This ensures that $\widehat{\lambda}^{\mathrm{U}}$ is an unbiased estimator when the true intensity is constant. Jones (1993) proposed the alternative estimator

$$\widehat{\lambda}^{\mathrm{JD}}(u) = \sum_{i=1}^{n} \frac{\zeta(u - x_i)}{c_W(x_i)}, \quad u \in W, \tag{8}$$

which conserves total mass, that is, $\int_W \widehat{\lambda}^{\mathrm{JD}}(u)\,\mathrm{d}u = n$.

These statistically-principled edge corrections can be adapted to a linear network and applied to the biased estimator (5). For points $x_1, \ldots, x_n$ observed on a network L, the analogue of Diggle's (1985) uniform correction is

$$\widehat{\lambda}^{\mathrm{U}}(u) = \frac{1}{c_L(u)} \sum_{i=1}^{n} k(d_L(u, x_i)) \quad u \in L, \tag{9}$$

and the analogue of Jones's (1993) correction, proposed by Moradi et al. (2018), is

$$\widehat{\lambda}^{\mathrm{JD}}(u) = \sum_{i=1}^{n} \frac{k(d_L(u, x_i))}{c_L(x_i)}, \quad u \in L, \tag{10}$$

where

$$c_L(u) = \int_L k(d_L(v, u))\,\mathrm{d}v. \tag{11}$$

In the special case where k is the uniform density on an interval $[0, h]$, we find $c_L(u) = |b_L(u, r)|$ is the total length of the disc of radius h in the shortest path metric, centred at u. Unfortunately, computation of the correction factor $c_L(u)$ or $c_L(x_i)$ is expensive, even in this simple case. Moradi et al. (2018) develop an algorithm for computing (10).

5.3. Path enumeration methods

Kernel density estimators on a general network were first investigated systematically by Okabe et al. (2009); Sugihara et al. (2010), summarised in Okabe and Sugihara (2012, Chap. 9). They considered computational algorithms which start with a kernel k on the real line, and progressively redistribute the mass of this kernel over the network. Desirable properties of a kernel estimator were listed. Two kernel estimators were found to satisfy many of the desired properties: the "equal-split discontinuous" and "equal-split continuous" rules (Okabe and Sugihara, 2012, Sec. 9.2.2 and 9.2.3).

The "continuous" rule has excellent properties: it is symmetric, conserves mass, and is unbiased when the true intensity is uniform. In comparison, the edge-correction estimator (9) is unbiased, but does not conserve mass; while (10) conserves mass, but is not unbiased; neither (9) nor (10) is symmetric. Unfortunately, the "continuous" rule is extremely slow to compute by the original algorithm of Sugihara et al. (2010); see McSwiggan et al. (2016, Table 1). The "discontinuous" rule is faster, but has less desirable properties (Okabe and Sugihara, 2012, Sec. 9.3.2). On a general network, both methods require a kernel on the real line with bounded support (i.e. such that $k(x) = 0$ when $|x| > h$ for some value h), which excludes the Gaussian kernel. Computational cost increases exponentially with the bandwidth, so that automatic bandwidth selection is computationally prohibitive.

The "equal-split discontinuous" algorithm is described in Okabe and Sugihara (2012, Algorithms 9.1 and 9.2, Sec. 9.3.2) and equivalently in McSwiggan et al. (2016, Appendix, Algorithm 1). Effectively, the algorithm constructs a copy of the kernel k for each data point x_i on the network L. For locations u on the same line segment as the starting point x_i, the kernel estimate is $k(d_L(u, x_i))$. At each fork in the network, the kernel's remaining tail mass is divided equally over the outgoing line segments, preserving the total mass.

Essentially the same rationale was proposed independently by Ver Hoef and Peterson (2010) for nonparametric regression on a network of rivers or streams.

On a network without loops, the equal-split discontinuous kernel at location u due to a data point at x takes the value

$$K^{\mathrm{D}}(u \mid x) = \frac{k(d_L(x,u))}{(m_1-1)(m_2-1)\dots(m_P-1)}, \tag{12}$$

where $m_1, m_2, \dots, m_P$ are the degrees of each vertex (excluding u and x) along the shortest path from x to u. On a general network,

$$K^{\mathrm{D}}(u \mid x) = \sum_{\pi}^{*} k(\ell(\pi)) a^{D}(\pi), \tag{13}$$

(McSwiggan et al., 2016, Thm. 1), where the sum is over all paths $\pi = (x, v_1, \dots, v_{P-1}, u)$, of length less than or equal to h, from x to u that are non-reflecting (i.e. $e_{i+1} \neq e_i$, with e_i as the edge containing v_{i-1} and v_i), and $\ell(\pi)$ is the length of the path, and $a^{D}(\pi) = 1/((m_1-1)(m_2-1)\dots(m_{P-1}-1))$, where m_i is the degree of v_i.

The more expensive "equal-split *continuous*" algorithm is formally described in Okabe and Sugihara (2012, Algorithm 9.3, Sec. 9.3.3) and an equivalent version is given in McSwiggan et al. (2016, Appendix, Algorithm 2). It traverses *all* paths of length less than the kernel width h, including the self-intersecting paths that reflect at a vertex. When a path reaches a vertex of degree m, there are $m-1$ outgoing branches and one incoming branch which the path has just traversed. Each outgoing branch receives a weighted copy of the tail with equal weight $2/m$, and the incoming branch receives a weighted copy of the tail with negative weight $(2/m) - 1$. The result of this algorithm has the path sum representation, analogous to (13),

$$K^{\mathrm{C}}(u \mid x) = \sum_{\pi} k(\ell(\pi)) a^{C}(\pi), \tag{14}$$

(McSwiggan et al., 2016, Thm. 2), where the sum is now over *all* paths from x to u, and $a^{C}(\pi) = c_1 \dots c_{P-1}$, where $c_j = \frac{2}{deg(v_j)} - \delta_j$, in which $\delta_1 = \delta_P = 0$ and $\delta_j = \mathbf{1}\{e_j = e_{j-1}\}$ is the indicator that equals 1 if the path reverses at step j, and 0 otherwise.

5.4. Heat kernel

McSwiggan et al. (2016) developed a statistically principled kernel estimator on a linear network by exploiting the connection between kernel smoothing and diffusion (Chaudhuri and Marron, 2000; Botev et al., 2010). On a network, the counterpart of the Gaussian kernel is the heat kernel, the function in classical physics which describes the diffusion or propagation of heat along the network. A kernel estimator based on the heat kernel can be visualised by imagining that the network is a physical structure made of steel wire; heat is applied to the network at the locations of the observed events x_i (with a unit amount of heat applied at each x_i instantaneously); after a time t has elapsed, the distribution of temperature over the network is recorded. The kernel estimate can be computed rapidly by imitating this process, that is, by numerically solving the

time-dependent Fourier heat equation over the network, with initial condition given by the observed event locations. Computation is fast, indeed fast enough to allow data-based bandwidth selection. Many statistical properties of the heat kernel estimator can be established.

The time-dependent heat equation on the network is the partial differential equation

$$\frac{\partial f}{\partial t} = \beta \frac{\partial^2 f}{\partial x^2} \tag{15}$$

which is well-defined everywhere except at a vertex of the network. The solution of (15) with initial condition $f_0(x) = g(x)$ is

$$f_t(u) = \int_L g(x)\kappa_t(u \mid x)\,\mathrm{d}x, \tag{16}$$

where $\kappa_t(u \mid x)$ is the *"heat kernel"* or heat transfer function from source x to destination u over time t. There is typically no closed-form expression for the heat kernel.

The *diffusion estimator* of intensity $\widehat{\lambda}^{\mathrm{H}}(u)$ can be defined as a sum of heat kernels (McSwiggan et al., 2019)

$$\widehat{\lambda}^{\mathrm{H}}(u) = \sum_{i=1}^{n} \kappa_t(u \mid x_i), \quad u \in L, \tag{17}$$

where elapsed time $t = \sigma^2$ is the squared bandwidth. This expression would not be used in computation. Rather, one can use the fact that the diffusion estimator is a solution of the time-dependent heat equation (15), and solve this numerically up to the desired time $t = \sigma^2$. Numerical solution of the heat equation is many orders of magnitude faster than path-enumeration algorithms (McSwiggan et al., 2016, Table 1). Computation time increases quadratically with bandwidth. Figure 10 shows a diffusion estimate of accident intensity in the Perth CBD.

Figure 10: Estimate of traffic accident intensity in the Perth CBD using the heat kernel algorithm with bandwidth 330 metres. Intensity values (accidents per km per year) proportional to line widths.

Statistical properties of the diffusion estimator can be deduced from the heat kernel representation (16). The heat kernel is symmetric; the diffusion estimator conserves mass, and is unbiased *if and only if* the true intensity is uniform (McSwiggan et al., 2019).

Surprisingly, the heat kernel estimator is mathematically *equivalent* to the "equal-split continuous" estimator extended to the Gaussian kernel. In stochastic process theory there are theorems which express the heat kernel on a network as an infinite sum, over all paths in the network, of weighted contributions involving the Gaussian probability density (Kostrykin et al., 2007, 2012). It was shown by McSwiggan et al. (2016) that the representation of the heat kernel in Kostrykin et al. (2007, Corollary 3.4) is formally equivalent to the path-sum representation (14) for the equal-split continuous algorithm if k is taken to be the Gaussian kernel and an infinite sum over all paths in the network is permitted. Accordingly, the heat kernel estimator is the unique estimator of intensity which satisfies the long list of desirable properties given by Okabe et al. (2009); Sugihara et al. (2010) as well as agreeing locally with the Gaussian kernel.

5.5. Bandwidth selection

Kernel methods require a choice of bandwidth. Classical univariate methods for bandwidth selection (Silverman, 1986; Wand and Jones, 1995; Jones et al., 1996b) have been extended to spatial data (Sain et al., 1994; Duong and Hazelton, 2003, 2005; Davies et al., 2018, Section 3) and to data on a linear network (McSwiggan et al., 2016; Rakshit et al., 2019b). They include likelihood cross-validation in which we maximise the criterion

$$\mathsf{cv}(\sigma) = \sum_{i=1}^{n} \log(\widehat{\lambda}_{\sigma}^{-i}(x_i)) - \int_L \widehat{\lambda}_{\sigma}(u)\,\mathrm{d}u, \tag{18}$$

based on the Poisson point process likelihood, where $\widehat{\lambda}_{\sigma}(u)$ is one of the kernel estimators described here, and $\widehat{\lambda}_{\sigma}^{(-i)}(x_i)$ is the corresponding "leave-one-out" kernel estimate at x_i, defined by omitting the contribution from x_i. For kernel estimators which conserve mass, such as the diffusion estimator, the integral term in (18) is constant and can be omitted. The leave-one-out estimates for the diffusion estimate (17) are hard to compute (McSwiggan et al., 2016, Section 7.1) although a workable fast approximation is simply to truncate the series expansion (14) of the heat kernel (McSwiggan et al., 2019).

We may also turn to simple rules of thumb for bandwidth selection. While crude, they are fast and easy to calculate and remain useful for initial exploration, provided of course that they are adaptable to linear network data. Analogues of Scott's rule of thumb (Scott, 1992, p. 152), for example, perform well on a network in our experience (McSwiggan et al., 2016; Rakshit et al., 2019b). However, the same cannot be said for all such selection methods: It is unclear, for example, how to apply the oversmoothing principle of Terrell (1990) to data on a network.

5.6. Euclidean kernel

Density can also be estimated using two-dimensional geometry. Borruso (2003, 2005, 2008) described several *ad hoc* techniques for kernel density esti-

mation including the "Euclidean, divide-by-length" estimator (Borruso, 2008)

$$\widehat{\lambda}^{\mathrm{B}}(u) = \frac{N(\mathbf{x} \cap b(u,r))}{|L \cap b(u,r)|}, \quad u \in \mathbb{R}^2, \tag{19}$$

where $b(u,r) = \{v \in \mathbb{R}^2 : \|v-u\| \le r\}$ is the *two-dimensional* Euclidean disc of fixed radius $r > 0$ centred at the query location u. That is, (19) is the number of data points, divided by the total network length, within a Euclidean disc of radius r.

Rakshit et al. (2019b) provided a rigorous justification and generalisation of (19). The point locations, and the network itself, are convolved with a *two-dimensional* smoothing kernel, then combined into an intensity function on the network. Let ζ denote a bivariate kernel function, that is, a probability density on $\mathbb{R}^2$. The *convolution kernel estimator* of intensity is, with the *uniform correction*,

$$\widehat{\lambda}^{\mathrm{U}}(u) = \frac{1}{a_L(u)} \sum_{i=1}^{n} \zeta(u - x_i), \quad u \in L, \tag{20}$$

and with the *Jones-Diggle correction*

$$\widehat{\lambda}^{\mathrm{JD}}(u) = \sum_{i=1}^{n} \frac{\zeta(u - x_i)}{a_L(x_i)}, \quad u \in L, \tag{21}$$

where

$$a_L(u) = \int_L \zeta(v - u)\, \mathrm{d}v, \quad u \in L. \tag{22}$$

Statistical properties are analysed by Rakshit et al. (2019b). The estimator can be computed rapidly using the Fast Fourier Transform, even on large networks and for large bandwidths. The estimator is consistent and its statistical efficiency is only slightly sub-optimal.

In the special case where ζ is the uniform density on a disc of fixed radius $r > 0$, the uniform-correction estimator (20) reduces to Borruso's estimator (19), but now restricted to query locations u lying on the network. The Jones-Diggle correction estimator (21) reduces to

$$\widehat{\lambda}^{\mathrm{JD}}(u) = \sum_{x_i \in b(u,r)} \frac{1}{|L \cap b(x_i, r)|}, \quad u \in L. \tag{23}$$

This corresponds to associating, with each data point x_i, a unit mass which is then uniformly spread over the part of the network lying within Euclidean distance r of x_i.

Unlike estimators of intensity based on path distances in the network, the convolution estimators (20)–(21) are robust against errors in the geometry of the network Rakshit et al. (2019b). Leave-one-out estimates are also easy to compute, so that bandwidth selection using likelihood cross-validation is feasible.

Figures 11 and 12 show the Euclidean kernel estimates of accident intensity for the Perth CBD and for the Perth metropolitan area, using bandwidths of 91 and 357 metres, respectively, selected by leave-one-out cross-validation.

Figure 11: Kernel estimate of accident intensity for Perth CBD, using Euclidean kernel with fixed bandwidth 91 metres selected by leave-one-out cross-validation. Uniform correction (20). Line thickness proportional to intensity values.

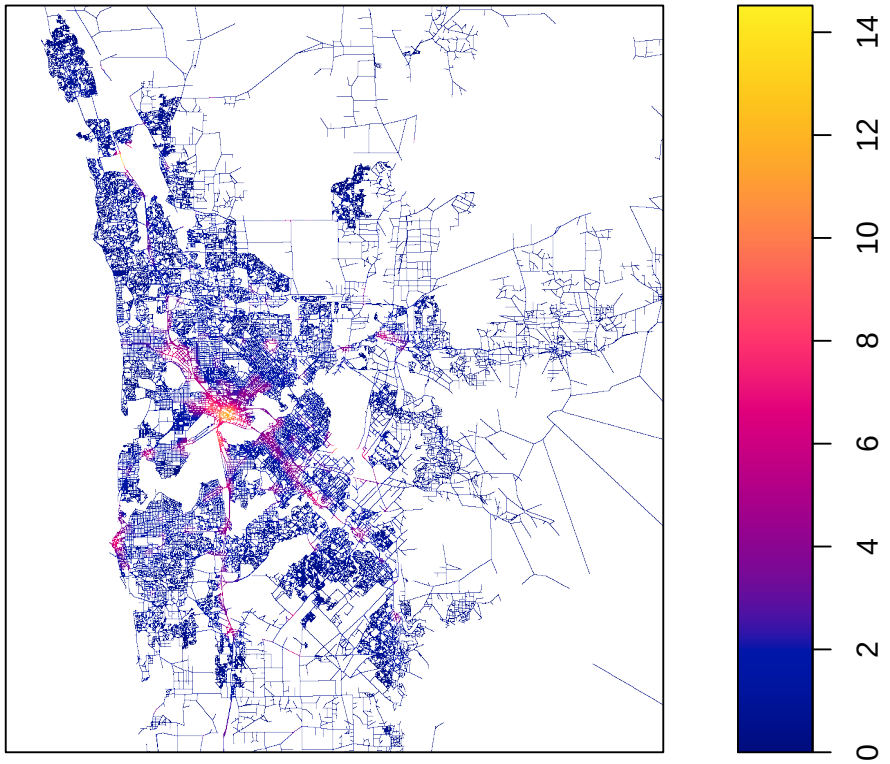

Figure 12: Kernel estimate of accident intensity for Perth metropolitan area, using Euclidean kernel with fixed bandwidth 357 metres selected by leave-one-out cross-validation. Jones-Diggle correction (21). Intensity values represented as colours according to the colour scale at right.

5.7. Adaptive estimators

The Western Australian road accident data, shown in Figure 3, exhibit huge spatial variation in the concentration of data points. In such situations it is well

known that fixed-bandwidth kernel estimation can perform poorly. Dense concentrations of accidents in the urban areas will be over-smoothed, and sparse accidents in the remote desert will be under-smoothed. Adaptive (variable-bandwidth) kernel estimation can perform substantially better in this context (Abramson, 1982; Hall and Marron, 1988). This has been investigated for networks by Rakshit et al. (2019b).

Following Marshall and Hazelton (2010) one may construct a spatially-varying bandwidth function $\sigma(u),\ u \in L$ and estimate the intensity by

$$\widehat{\lambda}^{\mathrm{U}}(u) = \frac{1}{c_L(u, \sigma(u))} \sum_{i=1}^{n} \kappa_{\sigma(x_i)}(u - x_i), \quad u \in L, \tag{24}$$

analogous to the uniform correction (20). Alternatively, each data point x_i may be assigned its own smoothing bandwidth σ_i, and we estimate the intensity by

$$\widehat{\lambda}^{\mathrm{JD}}(u) = \sum_{i=1}^{n} \frac{\kappa_{\sigma_i}(u - x_i)}{c_L(x_i, \sigma_i)}, \quad u \in L, \tag{25}$$

analogous to the Jones-Diggle correction (21). The higher computational cost can be reduced by partitioning (Davies and Baddeley, 2018, Section 4).

The adaptive bandwidths σ_i may be assigned, following the prescription of Abramson (1982), by first computing a pilot estimate of intensity $\tilde{\lambda}(\cdot)$, then computing initial bandwidths $a_i = \sqrt{(n/\tilde{\lambda}(x_i))}$, and finally assigning bandwidths $\sigma_i = (a_i/a)\sigma$, where $a = (\prod_i a_i)^{1/n}$ is the geometric mean of the initial bandwidths, and σ is the "global" bandwidth.

Figure 13 shows two panels representing the fixed-bandwidth and adaptive estimates, respectively, as three-dimensional surfaces. We refer to these as "heightened network" (HEN) plots. They are similar in style to those of Borruso (2008), where they portray a function defined on the two-dimensional plane as a surface viewed in perspective. In our case, the function is simply the extension of the estimator (20), (21), (24) or (25) to all locations $u \in \mathbb{R}^2$, which is an intermediate result in the convolution method calculations. The surface height is proportional to this function value; the surface colour also represents the function value; and the network itself is overlaid onto the surface.

Figure 13 is a screenshot of an interactive 3D graphics tool which can be viewed at `http://www.stats.otago.ac.nz/~tdavies/wacbd_hen.html`. Characteristically, we see a smoother adaptive estimate in areas of relatively low point density when compared to the fixed bandwidth estimate, with taller peaks than the fixed bandwidth estimate in the most dense areas.

5.8. Piecewise constant estimators

Several recently-developed methods for linear networks produce intensity estimates which are piecewise constant, with the aim of improving behaviour. The Voronoi estimator (Barr and Schoenberg, 2010; Moradi et al., 2019) assigns a constant intensity value on each tile of the Dirichlet-Voronoi tessellation induced by the data points. It is highly adaptable to changes of spatial scale, but has unacceptably high variance, which can be reduced by a bootstrap resample smoothing procedure (Moradi et al., 2019).

Fused density estimates (Bassett and Sharpnack, 2019) are defined as the solution of a mathematical extremal problem. They are constant except for

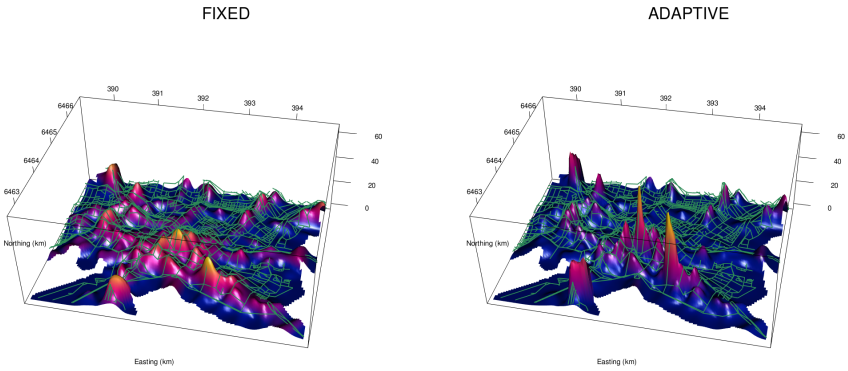

Figure 13: A screenshot of the fixed and adaptive intensity estimates of the Perth CBD data shown as interactive HEN plots. Accessible at the URL noted in the text.

discontinuities at the data points and the network vertices. To our knowledge, fused density estimation is the only technique which prevents kernel mass from "leaking" between adjacent network segments. For example, in Figures 10 and 11 the high density of accidents along the freeway running through the Perth CBD is partly leaking into neighbouring streets which are much quieter. This would not occur with a fused density estimate.

5.9. Relative risk

If $\mathbf{x}$ and $\mathbf{y}$ are two point patterns on the same network L, it may be important to estimate the ratio of intensities $r(u) = \lambda_{\mathbf{Y}}(u)/\lambda_{\mathbf{X}}(u)$ of the underlying point processes $\mathbf{X}$ and $\mathbf{Y}$. This ratio could represent the spatially-varying relative risk of two different types of crimes, the relative abundance of two types of trees, and so on. Estimates of $r(u)$ are typically computed by applying one of the kernel smoothing methods listed above to both patterns, and plugging-in to obtain $\widehat{r}(u) = \widehat{\lambda}_{\mathbf{Y}}(u)/\widehat{\lambda}_{\mathbf{X}}(u)$. Bandwidth selection methods are studied in McSwiggan et al. (2019) with the strong advice that the two estimates $\widehat{\lambda}_{\mathbf{Y}}(u)$, $\widehat{\lambda}_{\mathbf{X}}(u)$ should be computed using identical bandwidths.

In the Chicago data of Figure 4, due to the small sample size, we grouped the crime types into "personal" (assault, robbery) and "property" (burglary, car theft, damage, theft and trespass) crimes. Figure 14 shows a plug-in estimate of relative risk for "personal" versus "property" crimes, normalised by total counts. There is a hint of spatial variation in the type of crime.

Adaptive estimation of relative risk for two-dimensional point patterns was developed by Davies and Hazelton (2010), including asymptotic tolerance contours. Davies et al. (2016) advocated the use of equal bandwidths in estimating the numerator and denominator of risk, extending Kelsall and Diggle (1995) to the adaptive case. These results were extended to linear networks by Rakshit et al. (2019b).

6. Intensity depending on a covariate

The effect of an explanatory variable on the density of points can be investigated using nonparametric curve estimation techniques.

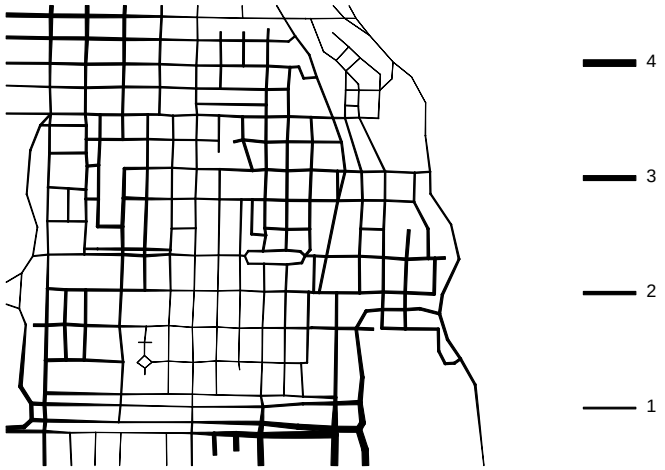

Figure 14: Estimated relative risk of crimes against the person versus property crimes for the Chicago crimes data, normalised.

Suppose that Z is a spatial covariate function, and we believe that the intensity of the points depends only on Z through a relationship

$$\lambda(u) = \rho(Z(u)) \tag{26}$$

where $\rho(z)$ is an unknown function that is to be estimated. For example if $\rho(z)$ is a decreasing function of z, the relationship (26) specifies that the density of points will be lower in those parts of the network where $Z(u)$ has a larger value. In ecological applications, $Z(u)$ could measure the local concentration of a resource such as water, and $\rho(z)$ is a 'resource selection function' indicating a species' habitat preferences (Manly et al., 2004). In geology, $Z(u)$ could be a geochemical variable, and $\rho(z)$ is a 'prospectivity index' representing the predicted spatial density of gold deposits as a function of geochemistry.

Nonparametric techniques for estimating the function $\rho(z)$ have been discussed by Manly et al. (2004); Guan (2008); Baddeley et al. (2012). These techniques compare the relative distribution (Handcock and Morris, 1999) of the values of Z at the observed data points with the values of Z at all locations. Although the techniques were developed for spatial point patterns in two dimensions, they depend only on the space of values of Z, so they apply to point patterns in any space (Baddeley et al., 2012), in particular to networks (Baddeley et al., 2015, Sec. 17.4.3).

The left panel of Figure 15 shows an estimate of intensity for the "thin" dendritic spines (Figure 6) as a function of shortest-path distance to the root of the dendritic tree, using the "ratio" method of Baddeley et al. (2012). It suggests that the thin spines have higher density at the far reaches of the tree, which is consistent with current understanding of dendrite growth (Baddeley et al., 2014).

The right panel of Figure 15 shows an estimate of intensity for the Geelong road accidents (Figure 7) as a function of traffic volume, using the "reweighted"

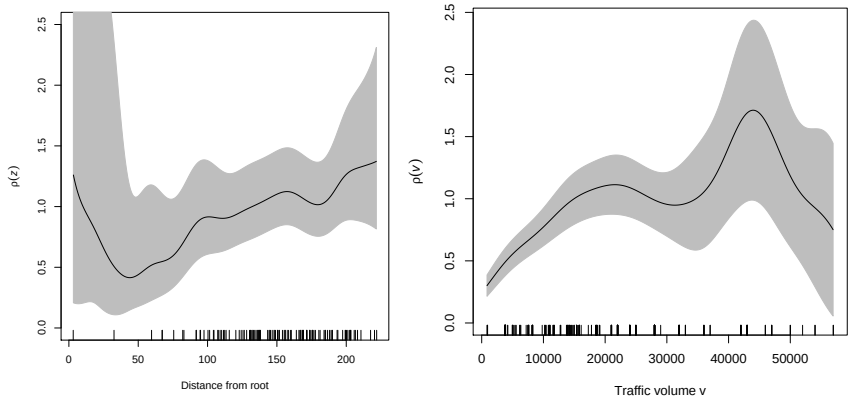

Figure 15: Estimates of intensity depending on a covariate. *Left:* Intensity of dendritic spines ("thin" type) as a function of distance from the root of the dendritic tree. *Right:* Intensity of traffic accidents in Geelong as a function of traffic volume. Solid lines: estimate of $\rho(z)$; grey shading: pointwise 95% confidence interval for $\rho(z)$.

method of Baddeley et al. (2012). This is broadly consistent with the expected relationship that accident intensity should increase with the square root of traffic volume (Jurewicz and Bennett, 2010).

Hypothesis tests for the (non-)dependence of intensity on a spatial covariate, such as Berman's tests and spatial cumulative distribution tests (Baddeley et al., 2015, Sec. 6.7.2, 10.5.2) likewise depend only on the values of the covariate; so they can also be applied to data on a linear network.

7. Parametric models and model-fitting

A major goal of statistical analysis is to formulate and fit parametric models to point pattern data on a network. The models are point processes which depend on explanatory spatial variables. The primary aim is usually to model the dependence of the intensity on the covariates, taking into account any stochastic dependence between different points.

The general theory of point processes (Daley and Vere-Jones, 2003) easily handles the definition, construction and characterisation of parametric point process models on a linear network, as well as space-time point processes. However, geometrical inhomogeneity hampers the construction of models with desired properties. Consequently, explicit point process modelling on a network is somewhat under-developed.

7.1. Poisson models

The simplest and most important reference model is the *Poisson point process* which effectively assumes that individual random points are independent of each other. The Poisson process with intensity function $\lambda(u)$ is formally characterised by the properties that for any line segment $B \subset L$, the number of points falling in B has a Poisson distribution with mean $\mu(B) = \int_B \lambda(u)\,\mathrm{d}u$, while events occurring in disjoint line segments $B_1, \ldots, B_m \subset L$ are independent. Simulated realisations of the Poisson point process can be generated using these properties, or using the thinning algorithm of Lewis and Shedler (1979).

A parametric model would postulate that the point process is Poisson with an intensity $\lambda_\theta(u)$ which depends explicitly on a parameter θ to be estimated, according to any desired functional form, which may depend implicitly on spatial covariates. For example, the *log-linear model*

$$\lambda_\theta(u) = \exp(\theta^\top Z(u)) = \exp(\theta_1 Z_1(u) + \ldots + \theta_p Z_p(u)) \tag{27}$$

specifies the intensity as a function of p explanatory variables $Z_1(u), \ldots, Z_p(u)$ and a p-dimensional vector of parameters $\theta = (\theta_1, \ldots, \theta_p)$.

The Poisson process with a parametrically specified intensity $\lambda_\theta(u)$ can be fitted to point pattern data by maximising the log likelihood

$$\log L(\theta) = \sum_{i=1}^{n} \log \lambda_\theta(x_i) - \int_L \lambda_\theta(u)\, \mathrm{d}u \tag{28}$$

as a function of the model parameter θ. McSwiggan (2019) developed algorithms for maximum likelihood estimation on a linear network, using versions of the Berman-Turner device (Berman and Turner, 1992; Baddeley and Turner, 2000). Models have been fitted to traffic accident data including covariates such as traffic volume, speed limit, and distance to nearest road intersection (McSwiggan, 2019). Results include predicted accident rate, confidence intervals for the parameters and for the fitted accident rate, and model selection using analysis of deviance.

Figure 16 shows the fitted intensities of Poisson point process models fitted separately to each panel of the spider webs data (Figure 5) assuming the log intensity is a quadratic function of the Cartesian coordinates.

Even when the Poisson process is not an appropriate model, experience with two-dimensional point pattern data suggests that the Poisson likelihood may still be an appropriate tool for estimating the intensity (Guan et al., 2015; Waagepetersen and Guan, 2009).

Model selection or variable selection methods are needed when there are many explanatory variables under consideration. Some methods of variable selection are available for point process models in two dimensional space, including sufficient dimension reduction (Guan and Wang, 2010) and penalised maximum likelihood (Yue and Loh, 2015) as well as classical hypothesis tests and Akaike information criteria (Baddeley et al., 2015, pp. 335–338, 371–378, 512–513). Recently Rakshit et al. (2019c) adapted penalised maximum likelihood methods, including the lasso, ridge regression and elastic net, to point process models on a linear network.

7.2. Clustered point process models

Point process models which exhibit clustering (positive association between points), such as Poisson cluster processes and Cox processes (Møller and Waagepetersen, 2004, Chap. 5, Baddeley et al., 2015, Chap. 12), can easily be constructed on a linear network, and moment properties can be calculated by standard means.

The standard methods for fitting Cox and cluster models to point patterns in two dimensions are based on first and second moments (Diggle, 1978; Diggle and Gratton, 1984; Waagepetersen, 2007; Jalilian et al., 2013; Guan, 2006; Tanaka et al., 2008). The pair correlation function of the model must be invariant under translation so that it may be estimated nonparametrically from data (Baddeley

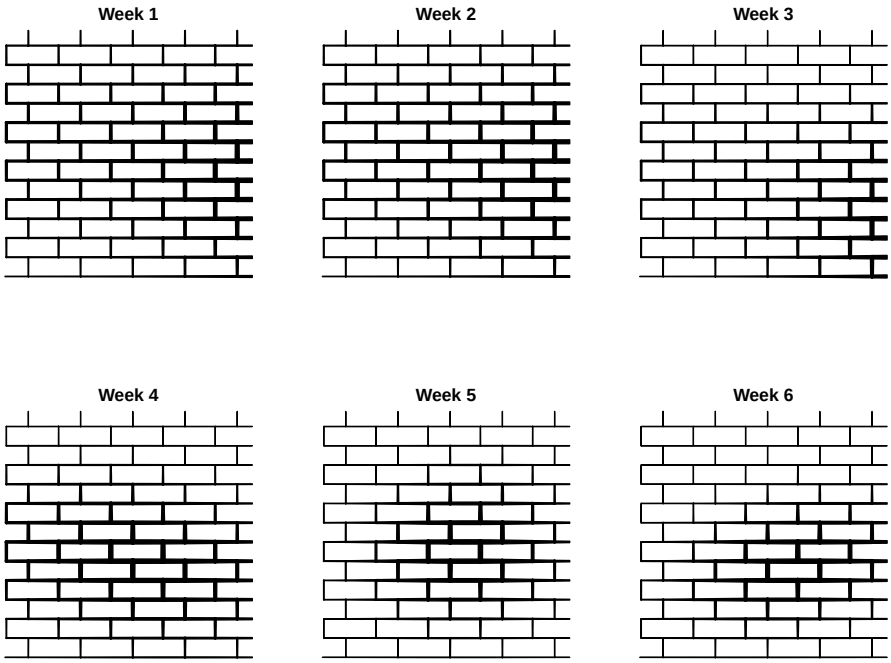

Figure 16: Fitted intensities of Poisson models for the spider webs data assuming a different log-quadratic intensity for each pattern.

et al., 2000), and it must be explicitly known as a function of the parameters, so that the model can be fitted to the nonparametric estimate by curve fitting (Diggle and Gratton, 1984).

It is more complicated to adapt these model-fitting methods to a linear network. The concept of translation invariance is not applicable; instead we assume that the pair correlation function depends only on distance between points. On many networks there do not exist Cox processes with a pair correlation function which depends only on shortest-path distance (Anderes et al., 2017). However, there do exist Cox processes whose pair correlation depends only on Euclidean distance (Baddeley et al., 2017) and these could be fitted to data using appropriate estimators of the pair correlation function (Rakshit et al., 2017).

7.3. Negatively associated point processes

Two-dimensional point process models which exhibit negative association between points include dependent thinnings of a Poisson process, such as the Matérn models (Matérn, 1986), and Markov and Gibbs point processes (van Lieshout, 2000, Møller and Waagepetersen, 2004, Chap. 6, Baddeley et al., 2015, Chap. 13). Counterparts of these models on a linear network can be defined, subject to technical conditions such as integrability and stability, which can be difficult to verify for some models (Geyer, 1999). Some nearest-neighbour Markov processes have been developed (van Lieshout, 2018). Existing model-fitting techniques for Gibbs models could also be adapted to linear networks.

7.4. Replicated point patterns

The successive panels in the spider web data, Figure 5, can be treated as replicated observations of the same spatial point process on a linear network. Replication allows a much wider range of statistical tools to be applied (Baddeley et al., 2015, Chap. 16; Baddeley et al., 1993; Bell and Grunwald, 2004; Diggle et al., 1991, 2000; Mateu, 2001; Myllymäki et al., 2013; Wager et al., 2004; Webster et al., 2005).

We have recently extended the model-fitting algorithm described in Section 7.1 to apply to replicated point patterns on a network. We fitted a Poisson point process model to the six spider web patterns assuming that the log intensity is a quadratic function of the Cartesian coordinates. This implicitly assumes that the patterns in successive weeks can be treated as independent but not necessarily identically distributed. We tested the null hypothesis that all six panels had the same coefficients of the quadratic function, against the alternative that the coefficients may be different in different panels (represented by the separate fitted intensities in Figure 16). Differences were not significant according to the Likelihood Ratio Test (deviance difference 36.9 on 30 degrees of freedom; p-value 0.18 using asymptotic χ^2 approximation). That is, our analysis suggests the spiders' habitat preferences are not changing over time. Details will be reported elsewhere.

8. K–function and pair correlation function

Correlation is a widely-used statistical measure of "dependence" or "association" between variables. For spatial point patterns, the K–function and the pair correlation function are correlation measures of association between points in the pattern. They have served a valuable role in the analysis of spatial point patterns in two and three dimensions. The task is to adapt these methods to a linear network.

In the last two decades, substantial research effort has been addressed to this problem by Prof. A. Okabe and collaborators (Okabe and Yamada, 2001; Shiode and Shiode, 2009; Warden, 2008; Yamada and Thill, 2007; Okabe et al., 2009, 2008, 2000; Shiode, 2008; Okabe and Satoh, 2006; Okabe et al., 1995, 2006b,a). Their work is surveyed in Okabe and Satoh (2009). Related work is by Borruso (2005, 2008); Downs and Horner (2007a); Jones et al. (1996a). More recently the statistical community has made contributions (Ang et al., 2012; Baddeley et al., 2014; Rakshit et al., 2017, 2019a).

8.1. K–function in two-dimensional space

The usual estimate of the K–function from an observed pattern of points $x_1, \ldots, x_n$ in a two-dimensional study region W is (Ripley, 1976, 1981, 1988)

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_i \sum_{j \neq i} e_{ij} \mathbf{1}\{\|x_i - x_j\| \leq r\}, \quad r \geq 0, \tag{29}$$

where $|W|$ is the area of the study region, e_{ij} is a correction for boundary effects, and $\mathbf{1}\{\|x_i - x_j\| \leq r\}$ is equal to 1 if the Euclidean distance between the points x_i and x_j is at most r, and equal to 0 otherwise.

We emphasise that (29) is not the definition of the K–function. Rather, (29) is an *estimate*, from the observed point pattern, of the true K–function of the point process which generated the pattern (in the same way that an average of observed numbers is an estimate of the true population mean). The true K–function $K(r)$ of a stationary point process $\mathbf{X}$ is defined as the (normalised) expected number of random points lying within a distance r of a typical random point (Ripley, 1977; Baddeley et al., 2015, Sec. 7.3.2, eq. (7.4), (7.6)):

$$K(r) = \frac{1}{\lambda}\mathbb{E}[\sum_{x_i \in \mathbf{X}} \mathbf{1}\{0 < \|x_i - x_0\| \leq r\} \mid x_0 \in \mathbf{X}], \quad r \geq 0, \tag{30}$$

where λ is the intensity of $\mathbf{X}$, and x_0 is any fixed location. The assumption of a stationary process is needed in order for this definition to be meaningful, and it also guarantees that (29) is a good estimate of the true value $K(r)$.

The K–function is a useful index of association between points in the pattern. A completely random pattern would have $K(r) = \pi r^2$. Values of $\hat{K}(r)$ exceeding this benchmark value suggest that the pattern is clustered. Statistically significant departures from a completely random pattern can be detected by comparing the empirical estimate $\hat{K}(r)$ with the envelopes of K–functions obtained from simulated completely random point patterns (Ripley, 1977; Baddeley et al., 2015, Sec. 7.8, 10.7).

Some writers advocate using the *pair correlation function* $g(r)$ instead of the K–function (Illian et al., 2008). The pair correlation function is related to the K–function through $g(r) = K'(r)/(2\pi r)$, where $K'(r)$ is the derivative of $K(r)$. A completely random pattern would have $g(r) = 1$. The pair correlation has an appealing interpretation in terms of the probability of observing a pair of random points separated by a distance r (Baddeley et al., 2015, p. 226).

The K–function and pair correlation function have been extended to two-dimensional point processes which have inhomogeneous intensity $\lambda(u)$, but have a stationary correlation structure (Baddeley et al., 2000; Baddeley et al., 2015, Sec. 7.10).

8.2. Two-dimensional K–function of points on a network

For a point pattern on a network, it would be feasible to ignore the network, extract the spatial coordinates of the points, and calculate $\hat{K}(r)$ according to (29). However, the interpretation of $\hat{K}(r)$ would now be different. It would not be correct to declare the pattern to be "clustered" if we find that $\hat{K}(r) > \pi r^2$, the benchmark for a completely random point process in two dimensions. A completely random point pattern *on the network* could produce values $\hat{K}(r) > \pi r^2$ for small r, because nearby points are constrained to lie on the same one-dimensional line. Real data examples are given by Yamada and Thill (2004); Lu and Chen (2007).

Indeed, in this context it would no longer be correct to refer to $\hat{K}(r)$ as an "estimate" of a well-defined quantity. Ripley's K–function formally assumes the point process is stationary in two dimensions, and this requirement is violated by point processes on a network, so the K–function is not well-defined.

It would nevertheless be possible to use $\hat{K}(r)$ as a summary statistic, and to perform simulation-based inference by comparing the value of $\hat{K}(r)$ calculated from the data with the envelopes of $\hat{K}(r)$ computed from simulated patterns generated according to a null model on the network. A similar argument is often

used for two-dimensional data. The main problem is that $\hat{K}(r)$ no longer has a clear interpretation if the null hypothesis is rejected.

8.3. Network K–function

Okabe and Yamada (2001) introduced a modification of the K–function procedure in which the Euclidean distance is replaced by the shortest-path distance. Suppose events have been observed to occur at the locations $x_1, \ldots, x_n$ on a network L. The 'network K function' is (Okabe and Yamada, 2001)

$$\widehat{K}_{\text{net}}(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{d_L(x_i, x_j) \leq r\}, \quad r \geq 0 \tag{31}$$

where $|L|$ denotes the total length of the linear network.

Note that $\widehat{K}_{\text{net}}$ is an *estimated* K–function (29). The corresponding 'theoretical' curve — the expected value of $\widehat{K}_{\text{net}}(r)$ for a completely random point process on the network — is not a simple function of r, and depends on the network geometry. In a homogeneous Poisson process on the network, the expected number of points which fall within a distance r of the location u (measured by the shortest path) is proportional to the total length of all line segments within distance r — that is, the total length of the segments making up the disc $b_L(u, r)$. See Figure 9. The 'theoretical' expected value of $\widehat{K}_{\text{net}}(r)$ is the average, over all locations u on the network, of the length of the disc of radius r centred at u:

$$\frac{\mathbb{E}\left[N(N-1)\widehat{K}_{\text{net}}(r)\right]}{\mathbb{E}[N(N-1)]} = \frac{1}{|L|} \int_L |b_L(u, r)| \, \mathrm{d}u. \tag{32}$$

That is, for a completely random point pattern, the expected value of the Okabe-Yamada network K–function is approximately equal to the average length of the balls of radius r centred at all possible locations on the network. Calculation of this 'theoretical' curve is a complicated task in itself. Network K–functions obtained from different networks are not directly comparable. For example, it is difficult to compare the spatial patterns of crime in two different cities using the respective network K–functions.

Nevertheless it is possible to perform simulation-based inference using the Okabe-Yamada empirical K–function (31).

8.4. Geometrically corrected K–function

Ang (2010); Ang et al. (2012) developed an adjusted network K–function that intrinsically corrects for the inhomogeneous geometry of the network. The geometrically-corrected empirical K–function is

$$\hat{K}_{\mathsf{L}}(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1\{d_L(x_i, x_j) \leq r\}}{m(x_i, d_L(x_i, x_j))}, \tag{33}$$

where $m(u, t) = \#\{v \in L : d_L(u, v) = t\}$ is the number of points of the network lying at path distance t exactly from the location u. That is, the contribution to (33) from each pair of points (x_i, x_j) is weighted by the reciprocal of the number of points $u \in L$ that are situated at the same distance from x_i as x_j is.

The weighting factor compensates exactly for the geometry of the network: for a completely random point pattern, on any network, the corrected K–function is always $K(r) = r$. This provides a simple benchmark for completely random point patterns on a linear network. It also permits comparison between the corrected K–functions obtained from different point patterns on different networks.

Ang's geometrical correction is formally analogous to the isotropic edge correction of Ripley (1977, 1981), which removes geometry-dependent bias in the two-dimensional case. The geometrical correction restores many natural properties of K, including its direct relationship to the pair correlation function. The corrected K–function estimator has approximately constant variance as a function of r. Bias and variance are calculated by Ang et al. (2012).

This approach extends to the spatially *inhomogeneous* case, yielding a K–function for inhomogeneous point processes on a network (Ang et al., 2012):

$$\hat{K}_{\mathsf{LI}}(r) = \frac{1}{|L|} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1\{d_L(x_i, x_j) \leq r\}}{\hat{\lambda}(x_i)\hat{\lambda}(x_j) m(x_i, d_L(x_i, x_j))}, \tag{34}$$

where $\hat{\lambda}(u)$ is an estimate of the intensity function. Statistical performance is improved if the denominator $|L|$ is replaced by a data-dependent estimate $\sum_i 1/\hat{\lambda}(x_i)$, which is positively correlated with the double sum in (34).

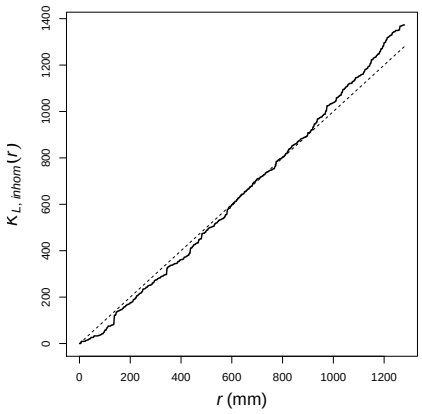

Figure 17: Estimated inhomogeneous K–function of the first spider web pattern assuming a log-quadratic intensity.

Figure 17 shows the estimated inhomogeneous K–function of the spider webs pattern in the top left panel of Figure 5, using the fitted intensity of the log-quadratic Poisson model described in Section 7. There is no evidence of departure from an inhomogeneous Poisson process — that is, no evidence that the individual spider web locations are dependent on each other, after allowing for location preferences.

For the investigation of spatial interaction, the geometrically corrected K–function (33) and pair correlation function have the strong advantage that they permit the range of interaction to be identified.

For "multitype" point patterns in which the points are classified into several different categories, such as the Chicago data (Figure 4) and dendritic spines

data (Figure 6), extensions of the geometrically corrected K–function (33)–(34) were developed by Baddeley et al. (2014) and applied to the dendritic spines data. When investigating the dependence between types, it may also be useful to estimate the network version of the *mark connection function* (Illian et al., 2008) which is a combination of these functions (Baddeley et al., 2014).

9. Construction problem and non-existence

Many standard methods for analysing spatial point patterns assume that the underlying point process is stationary. This gives access to a powerful statistical methodology, embracing nonparametric characteristics such as the K–function and pair correlation function (Illian et al., 2008), as well as parametric modelling and inference (Møller and Waagepetersen, 2004; Baddeley et al., 2015).

Much of this methodology *cannot be extended to a linear network*, because the network itself is not homogeneous. Different neighbourhoods have different geometry, so there are no non-trivial geometrical transformations which preserve the network, and it is not possible to define stationary processes (in the strict sense) on a linear network. This creates a fundamental problem for the standard methodology.

A fallback strategy is to analyse only the first and second moments. In two dimensions, the K–function and pair correlation function g remain well-defined if the point process is only *second order stationary*, i.e. if its first two moments are invariant under translation.

The K–function and pair correlation function have been adapted to point patterns on a linear network, as discussed in Section 8 above. The point process is required to be *correlation-stationary*, meaning that the pair correlation function depends only on the shortest-path distance between points (Ang et al., 2012).

Baddeley et al. (2017) described some simple constructions of point processes on a network which are correlation-stationary as defined above. However, the findings suggest that such processes may be quite rare, except when the network is a tree (a graph without cycles). Anderes et al. (2017) subsequently proved that stationary correlation functions (with respect to the shortest-path distance) do not exist on networks which contain certain kinds of loops.

This severely weakens the rationale for using the K–function and pair correlation function on a network (based on shortest-path distance) in real applications. Modelling and inference also become much more complicated.

The exception is when the network is a tree (acyclic graph), which is highly relevant in applications to river and stream networks (O'Donnell et al., 2014; Ver Hoef et al., 2006; Ver Hoef and Peterson, 2010) and to the dendritic trees of neurons (Baddeley et al., 2014; Jammalamadaka et al., 2013). A simple construction in Baddeley et al. (2017) shows that point processes with exponential pair correlation, $g(r) = \alpha \exp(-\beta r)$, exist on any tree.

10. Alternative distance metrics

Okabe and Sugihara (2012, pp. 7–8) explain carefully that, while it is often sensible to measure distances in a network by the shortest path, this is not obligatory, and may occasionally be inappropriate to the application.

Alternative metrics include the Euclidean distance between points in two-dimensional space, and the "resistance distance" defined by treating the lines of the network as electrical resistors (Klein and Randić, 1993; Bapat, 2004). Shortest-path distance may be modified by assigning a different cost per unit length on each segment of the network, where cost is determined by stream flow in a river network (Ver Hoef et al., 2006), by landscape topography (Foltête et al., 2008) or by expected travel time across a road network.

The Euclidean distance, for example, may be appropriate when studying the influence of weather on road accident risk. Locations on different roads which are spatially close together will have similar weather.

Amongst the metrics mentioned, the Euclidean distance is the only metric which is not affected by changes to the network, including clipping to an observation region and addition of new segments to an existing network.

Baddeley et al. (2017); Rakshit et al. (2017) investigated the statistical implications of using a distance metric other than the shortest-path distance. For any given metric δ, a point process is defined to be δ-*correlated* if its pair correlation function depends only on δ-distance, $g^{(2)}(u,v) = g_\delta(\delta(u,v))$, for some function g_δ. Baddeley et al. (2017) showed there is a rich class of point process models which are δ-correlated with respect to Euclidean distance: this includes Cox processes driven by any stationary random field on the plane, counterparts of Switzer's (1965) model and the cell process of Baddeley and Silverman (1984), and many others.

Rakshit et al. (2017) adapted the K–function and pair correlation function to a general distance metric δ. If the point process is δ-correlated, then a version of the K–function based on δ distances is well-defined. In order to estimate it from data, we need the Jacobian

$$J_\delta(u,v) = \left| \frac{\partial \delta(u,v)}{\partial v} \right|, \tag{35}$$

the rate of change of the δ-distance between u and v as v moves along the network at unit speed. For the shortest-path metric, $J = 1$. For the Euclidean metric, $J(u,v) = |\sin\theta|$ where θ is the angle of incidence between the segment containing v and the Euclidean circle centred at u that passes through v.

An empirical estimator of the K–function $K_\delta(r)$ is

$$\hat{K}_\delta(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{\delta_{ij} \leq r\} \frac{\widetilde{J}_\delta(x_i, \delta_{ij})}{m_\delta(x_i, \delta_{ij})}, \tag{36}$$

where $\delta_{ij} = \delta(x_i, x_j)$ for $i \neq j$ and

$$\widetilde{J}_\delta(u,r) = \left[\frac{1}{m_\delta(u,r)} \sum_{v \in b_\delta(u,r)} \frac{1}{J_\delta(u,v)} \right]^{-1}, \tag{37}$$

where $m_\delta(u,r)$ is the circumference of the disc $b_\delta(u,r)$ of radius r in the metric δ. That is, $\widetilde{J}_\delta(u,r)$ is the harmonic mean of the Jacobian values $J_\delta(u,v)$ at all locations v lying exactly r units away from u according to the metric δ.

Using this estimator and the corresponding estimator for the pair correlation function, Rakshit et al. (2017, Sec. 7) re-analysed several datasets, and reported

conflicting conclusions from analysis of the dendritic spines data using different metrics.

The Poisson process is δ-correlated for any metric δ. Thus, whatever the choice of metric δ, we may use the same functional form $K(r) = r$ as the benchmark of "complete randomness". While this result is very useful for data analysis, it also indicates that there is no right or wrong choice of the metric δ when the point process is completely random. The metric is "unidentifiable" under the null hypothesis of complete randomness, cf. Davies (1977, 1987).

Software

All analyses were performed using the libraries `spatstat` (Baddeley and Turner, 2005; Baddeley et al., 2015) and `spatstat.Knet` (Rakshit et al., 2019a) which are contributed extension packages for the R statistical software system (R Development Core Team, 2018). They can be downloaded from `cran.r-project.org`.

Acknowledgements

This article includes summaries of the results of joint research with the Perth Spatial Point Processes Group (Adrian Baddeley, Gopalan Nair, Suman Rakshit), with former members (Ya-Mei Chang, Andrew Hardegen, Thomas Lawrence, and Yong Song) and with Spatial Analysis Group Otago (Tilman Davies, Martin Hazelton, Adrian Baddeley), as well as collaborators Ege Rubak, Rob Foxall and Rolf Turner.

This work was supported by Australian Research Council grant DP130102322, "Statistical methodology for events on a network, with application to road safety" in which Mark Handcock, Martin Hazelton, Jakob Gulddahl Rasmussen and Valerie Isham participated. We thank CSIRO Data61 for computing resources and related support.

References

Abramson, I., 1982. On bandwidth estimation in kernel estimates – a square root law. Annals of Statistics 10, 1217–1223.

Anderes, E., Møller, J., Rasmussen, J., 2017. Isotropic covariance functions on graphs and their edges. Technical Report. Centre for Stochastic Geometry and Bioimaging. Aalborg, Denmark.

Andersen, B., 1990. Methodological Errors in Medical Research. Blackwell Scientific, Oxford.

Anderson, T., 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis and Prevention 41, 359–364.

Ang, Q., 2010. Statistical methodology for events on a network. Master's thesis. School of Mathematics and Statistics, University of Western Australia.

Ang, Q., 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, 591–617.

Anselin, L., 1995. Local indicators of spatial association – LISA. Geographical Analysis 27, 93–115.

Baddeley, A., 2017. Local composite likelihood for spatial point processes. Spatial Statistics 22, 261–295. doi: 10.1016/j.spasta.2017.03.001.

Baddeley, A., Chang, Y., Song, Y., Turner, R., 2012. Nonparametric estimation of the dependence of a spatial point process on a spatial covariate. Statistics and its Interface 5, 221–236.

Baddeley, A., Jammalamadaka, A., Nair, G., 2014. Multitype point process analysis of spines on the dendrite network of a neuron. Applied Statistics (Journal of the Royal Statistical Society, Series C) 63, 673–694. doi:10.1111/rssc.12054.

Baddeley, A., Møller, J., Waagepetersen, R., 2000. Non- and semiparametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica 54, 329–350.

Baddeley, A., Moyeed, R., Howard, C., Boyde, A., 1993. Analysis of a three-dimensional point pattern with replication. Applied Statistics 42, 641–668.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., 2017. 'Stationary' point processes are uncommon on linear networks. STAT 6, 68–78.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC, London.

Baddeley, A., Silverman, B., 1984. A cautionary example on the use of second-order methods for analyzing point patterns. Biometrics 40, 1089–1094.

Baddeley, A., Turner, R., 2000. Practical maximum pseudolikelihood for spatial point patterns (with discussion). Australian and New Zealand Journal of Statistics 42, 283–322.

Baddeley, A., Turner, R., 2005. Spatstat: an R package for analyzing spatial point patterns. Journal of Statistical Software 12, 1–42. URL: www.jstatsoft.org, ISSN: 1548-7660.

Bapat, R., 2004. Resistance matrix of a weighted graph. Communications in Mathematical and in Computer Chemistry 50, 73–82.

Barr, C., Schoenberg, F., 2010. On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. Biometrika 97, 977–984.

Bassett, R., Sharpnack, J., 2019. Fused density estimation: theory and methods. Journal of the Royal Statistical Society, Series B In press.

Bell, M., Grunwald, G., 2004. Mixed models for the analysis of replicated spatial point patterns. Biostatistics 5, 633–648.

Berkson, J., 1946. Limitations of the application of fourfold table analysis to hospital data. Biometrics Bulletin 2, 47–53.

Berman, M., Turner, T., 1992. Approximating point process likelihoods with GLIM. Applied Statistics 41, 31–38.

Bithell, J., 1990. An application of density estimation to geographical epidemiology. Statistics in Medicine 9, 691–701.

Borruso, G., 2003. Network density and the delimitation of urban areas. Transactions in GIS 7, 177–191.

Borruso, G., 2005. Network density estimation: Analysis of point patterns over a network, in: Gervasi, O., Gavrilova, M., Kumar, V., Laganà, A., Lee, H., Mun, Y., Taniar, D., Tan, C. (Eds.), Computational Science and its Applications — ICCSA 2005. Springer, Berlin/Heidelberg. number 3482 in Lecture Notes in Computer Science, pp. 126–132.

Borruso, G., 2008. Network density estimation: A GIS approach for analysing point patterns in a network space. Transactions in GIS 12, 377–402.

Botev, Z., Grotowski, J., Kroese, D., 2010. Kernel density estimation via diffusion. Annals of Statistics 38, 2916–2957.

Briz-Redón, Á., Martínez-Ruiz, F., Montes, F., 2019. Spatial analysis of traffic accidents near and between road intersections in a directed linear network. Accident analysis and prevention 132. In press.

Chaudhuri, P., Marron, J., 2000. Scale space view of curve estimation. Annals of Statistics 28, 408–428.

Comber, A., Brunsdon, C., Green, E., 2008. Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. Landscape and Urban Planning 86, 103–114.

Cressie, N., Collins, L., 2001. Analysis of spatial point patterns using bundles of product density LISA functions. Journal of Agricultural, Biological and Environmental Statistics 6, 118–135.

Cressie, N., Frey, J., Harch, B., Smith, M., 2006. Spatial prediction on a river network. Journal of Agricultural, Biological and Environmental Statistics 11, 127–150.

Cressie, N., Majure, J., 1997. Spatio-temporal statistical modeling of livestock waste in streams. Journal of Agricultural, Biological and Environmental Statistics 2, 24–47.

Cressie, N., Wikle, C., 2011. Statistics for Spatio-Temporal Data. John Wiley and Sons, Hoboken, NJ.

Daley, D., Vere-Jones, D., 1988. An Introduction to the Theory of Point Processes. Springer-Verlag, New York.

Daley, D., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods. Second ed., Springer-Verlag, New York.

Daley, D., Vere-Jones, D., 2008. An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure. Second ed., Springer-Verlag, New York.

Davies, R., 1977. Testing the hypothesis that a point process is Poisson. Advances in Applied Probability 9, 724–746.

Davies, R., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Davies, T., Baddeley, A., 2018. Fast computation of spatially adaptive kernel estimates. Statistics and Computing 28, 937–956.

Davies, T., Hazelton, M., 2010. Adaptive kernel estimation of spatial relative risk. Statistics in Medicine 29, 2423–2437.

Davies, T., Jones, K., Hazelton, M., 2016. Symmetric adaptive smoothing regimens for estimation of the spatial relative risk function. Computational Statistics and Data Analysis 101, 12–28.

Davies, T., Marshall, J., Hazelton, M., 2018. Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. Statistics in Medicine 37, 1191–1221.

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

Diggle, P., 1978. On parameter estimation for spatial point processes. Journal of the Royal Statistical Society, Series B 40, 178–181.

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

Diggle, P., 2010. Nonparametric methods, in: Gelfand, A., Diggle, P., Fuentes, M., Guttorp, P. (Eds.), Handbook of Spatial Statistics. CRC Press, Boca Raton, FL. chapter 18, pp. 299–316.

Diggle, P., 2014. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. Third ed., Chapman and Hall/CRC, Boca Raton, FL.

Diggle, P., Gratton, R., 1984. Monte Carlo methods of inference for implicit statistical models (with discussion). Journal of the Royal Statistical Society, Series B 46, 193–227.

Diggle, P., Lange, N., Benes, F., 1991. Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. Journal of the American Statistical Association 86, 618–625.

Diggle, P., Mateu, J., Clough, H., 2000. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. Advances in Applied Probability (SGSA) 32, 331–343.

Downs, J., Horner, M., 2007a. Characterising linear point patterns, in: Winstanley, A. (Ed.), Proceedings of the GIScience Research UK Conference (GISRUK), Maynooth, Ireland, National University of Ireland Maynooth, County Kildare, Ireland. pp. 421–424.

Downs, J., Horner, M., 2007b. Network-based kernel density estimation for home range analysis, in: Proceedings of the 9th International Conference on Geocomputation, Maynooth, Ireland.

Downs, J., Horner, M., 2008. Spatially modelling pathways of migratory birds for nature reserve site selection. International Journal of Geographical Information Science 22, 687–702.

Duong, T., Hazelton, M., 2003. Plug-in bandwidth matrices for bivariate kernel density estimation. Journal of Nonparametric Statistics 15, 17–30.

Duong, T., Hazelton, M., 2005. Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation. Journal of Multivariate Analysis 93, 417–433.

Foltête, J.C., Berthier, K., Cosson, J.F., 2008. Cost distance defined by a topological function of landscape. Ecological Modelling 210, 104–114.

Fotheringham, A., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley and Sons.

Gaetan, C., Guyon, X., 2009. Spatial Statistics and Modeling. Springer. Translated by Kevin Bleakley.

Gelfand, A., Diggle, P., Fuentes, M., Guttorp, P. (Eds.), 2010. Handbook of Spatial Statistics. CRC Press, Boca Raton, FL.

Geyer, C., 1999. Likelihood inference for spatial point processes, in: Barndorff-Nielsen, O., Kendall, W., van Lieshout, M. (Eds.), Stochastic Geometry: Likelihood and Computation. Chapman and Hall / CRC, Boca Raton, FL. number 80 in Monographs on Statistics and Applied Probability. chapter 3, pp. 79–140.

Guan, Y., 2006. A composite likelihood approach in fitting spatial point process models. Journal of the American Statistical Association 101, 1502–1512.

Guan, Y., 2008. On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. Journal of the American Statistical Association 103, 1238–1247.

Guan, Y., Jalilian, A., Waagepetersen, R., 2015. Quasi-likelihood for spatial point processes. Journal of the Royal Statistical Society, Series B 77, 677–697.

Guan, Y., Wang, H., 2010. Sufficient dimension reduction for spatial point processes directed by Gaussian random fields. Journal of the Royal Statistical Society, Series B 72, 367–387.

Hall, P., Marron, J., 1988. Variable window width kernel estimation of probability densities. Probability Theory and Related Fields 80, 37–49.

Handcock, M., Morris, M., 1999. Relative Distribution Methods in the Social Sciences. Springer-Verlag, New York.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons, Chichester.

Isaak, D., Peterson, E., Ver Hoef, J., Wenger, S., Falke, J., Torgersen, C., Sowder, C., Steel, E., Fortin, M., Jordan, C., Ruesch, A., Som, N., Monestiez, P., 2014. Applications of spatial statistical network models to stream data. Wiley Interdisciplinary Reviews: Water 1, 277–294.

Jalilian, A., Guan, Y., Waagepetersen, R., 2013. Decomposition of variance for spatial Cox processes. Scandinavian Journal of Statistics 40, 119–137.

Jammalamadaka, A., Banerjee, S., Manjunath, B., Kosik, K., 2013. Statistical analysis of dendritic spine distributions in rat hippocampal cultures. BMC Bioinformatics 14.

Johnson, S.D., 2010. A brief history of the analysis of crime concentration. European Journal of Applied Mathematics 21, 349–370.

Jones, A.P., Langford, I.H., Bentham, G., 1996a. The application of k-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. Social Science and Medicine 42, 879–885.

Jones, M., 1993. Simple boundary corrections for kernel density estimation. Statistics and Computing 3, 135–146.

Jones, M., Marron, J., Sheather, S., 1996b. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91, 401–407.

Jurewicz, C., Bennett, P., 2010. Road Safety Engineering Risk Assessment Part 7: Crash Rates Database. Technical Report AP-T152/10. Austroads. Sydney, Australia. ISBN 978-1-921709-02-9.

Kelsall, J., Diggle, P., 1995. Kernel estimation of relative risk. Bernoulli 1, 3–16.

Klein, D., Randić, M., 1993. Resistance distance. Journal of Mathematical Chemistry 12, 81–95.

Kostrykin, V., Potthoff, J., Schrader, R., 2007. Heat kernels on metric graphs and a trace formula, in: Germinet, F., Hislop, P. (Eds.), Adventures in Mathematical Physics. American Mathematical Society, Providence, RI. number 447 in Contemporary Mathematics, pp. 175–198.

Kostrykin, V., Potthoff, J., Schrader, R., 2012. Brownian motions on metric graphs. Journal of Mathematical Physics 53. doi: 10.1063/1.4714661.

Lewis, P., Shedler, G., 1979. Simulation of non-homogeneous Poisson processes by thinning. Naval Logistics Quarterly 26, 406–413.

van Lieshout, M., 2000. Markov Point Processes and Their Applications. Imperial College Press, London.

van Lieshout, M.N.M., 2018. Nearest-neighbour Markov point processes on graphs with Euclidean edges. Advances in Applied Probability 50, 12751293. doi:10.1017/apr.2018.60.

Lord, D., Mannering, F., 2010. The statistical analysis of crash frequency data: a review and assessment of methodological alternatives. Transportation Research A 44, 291–305.

Lord, D., Washington, S., Ivan, J., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37, 35–46.

Lu, Y., Chen, X., 2007. On the false alarm of planar K-function when analyzing urban crime distributed along streets. Social Science Research 36, 611–632.

Manly, B., McDonald, L., Thomas, D., McDonald, T., Erickson, W., 2004. Resource Selection by Animals: Statistical Design and Analysis for Field Studies. Kluwer, New York and Dordrecht.

Marshall, J., Hazelton, M., 2010. Boundary kernels for adaptive density estimators on regions with irregular boundaries. Journal of Multivariate Analysis 101, 949–963.

Matérn, B., 1986. Spatial Variation. Number 36 in Lecture Notes in Statistics, Springer Verlag, New York.

Mateu, J., 2001. Parametric procedures in the analysis of replicated spatial point patterns. Biometrical Journal 43, 375–394.

McSwiggan, G., 2019. Spatial point process methods for linear networks with applications to road accident analysis. Ph.D. thesis. University of Western Australia.

McSwiggan, G., Baddeley, A., Nair, G., 2016. Kernel density estimation on a linear network. Scandinavian Journal of Statistics 44, 324–345.

McSwiggan, G., Baddeley, A., Nair, G., 2019. Estimation of relative risk for events on a linear network. Statistics and Computing 30, 469–484. doi:10.1007/s11222-019-09889-7. published online 24 august 2019.

Møller, J., Waagepetersen, R., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Boca Raton, FL.

Moradi, M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J., Baddeley, A., 2019. Resample-smoothing of Voronoi intensity estimators. Statistics and Computing 29, 995–1010. doi:10.1007/s11222-018-09850-0. published online 22 january 2019.

Moradi, M.M., Rodríguez-Cortés, F., Mateu, J., 2018. On kernel-based intensity estimation of spatial point patterns on linear networks. Journal of Computational and Graphical Statistics 27, 302–311. Doi: 10.1080/10618600.2017.1360782.

Myllymäki, M., Särkkä, A., Vehtari, A., 2013. Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates. Spatial Statistics 8, 104–121.

O'Donnell, D., Rushworth, A., Bowman, A., Scott, E., 2014. Flexible regression models over river networks. Applied Statistics (Journal of the Royal Statistical Society, Series C) 63, 47–63.

Okabe, A., Boots, B., Sugihara, K., Chiu, S., 2000. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley and Sons, Chichester.

Okabe, A., Kitamura, M., 1996. A computational method for market area analysis on a network. Geographical Analysis 28, 330–349.

Okabe, A., Okunuki, K., 2001. A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. Transactions in GIS 5, 209–220.

Okabe, A., Okunuki, K., Shiode, S., 2006a. SANET: a toolbox for spatial analysis on a network. Geographical Analysis 28, 57–66.

Okabe, A., Okunuki, K., Shiode, S., 2006b. The SANET toolbox: new methods for network spatial analysis. Transactions in GIS 10, 535–550.

Okabe, A., Sato, T., Furuta, T., Suzuki, A., Okano, K., 2008. Generalized network Voronoi diagrams: Concepts, computational methods, and applications. International Journal of Geographical Information Science 22, 1–30.

Okabe, A., Satoh, T., 2006. Uniform network transformation for points pattern analysis on a non-uniform network. Journal of Geographical Systems 8, 25–37.

Okabe, A., Satoh, T., 2009. Spatial analysis on a network, in: Fotheringham, A., Rogers, P. (Eds.), The SAGE Handbook on Spatial Analysis. SAGE Publications, London. chapter 23, pp. 443–464.

Okabe, A., Satoh, T., Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. International Journal of Geographical Information Science 23, 7–32.

Okabe, A., Sugihara, K., 2012. Spatial Analysis Along Networks. John Wiley and Sons, New York.

Okabe, A., Yamada, I., 2001. The K-function method on a network and its computational implementation. Geographical Analysis 33, 271–290.

Okabe, A., Yomono, H., Kitamura, M., 1995. Statistical analysis of the distribution of points on a network. Geographical Analysis 27, 152–175.

Okunuki, K., Okabe, A., 2003. Solving the Huff-based competitive location model on a network with link-based demand. Annals of Operations Research 111, 239–252.

R Development Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/. ISBN 3-900051-07-0.

Rakshit, S., Baddeley, A., Nair, G., 2019a. Efficient code for second-order analysis of events on a linear network. Journal of Statistical Software 90, 1–37. URL: https://www.jstatsoft.org/v090/i01, doi:10.18637/jss.v090.i01.

Rakshit, S., Davies, T., Moradi, M., McSwiggan, G., Nair, G., Mateu, J., Baddeley, A., 2019b. Fast kernel smoothing of point patterns on a large network using 2D convolution. International Statistical Review 87, 531–556. doi:10.1111/insr.12327. published online 06 June 2019.

Rakshit, S., McSwiggan, G., Nair, G., Baddeley, A., 2019c. Variable selection on a linear network. Submitted for publication.

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.

Ripley, B., 1976. The second-order analysis of stationary point processes. Journal of Applied Probability 13, 255–266.

Ripley, B., 1977. Modelling spatial patterns (with discussion). Journal of the Royal Statistical Society, Series B 39, 172–212.

Ripley, B., 1981. Spatial Statistics. John Wiley and Sons, New York.

Ripley, B., 1988. Statistical Inference for Spatial Processes. Cambridge University Press.

Sain, S., Baggerly, K., Scott, D., 1994. Cross-validation of multivariate densities. Journal of the American Statistical Association 89, 807–817.

Scott, D., 1992. Multivariate Density Estimation. Theory, Practice and Visualization. John Wiley and Sons, New York.

Shiode, S., 2008. Analysis of a distribution of point events using the network-based quadrat method. Geographical Analysis 40, 401–422.

Shiode, S., Shiode, N., 2009. Detection of hierarchical point agglomerations by the network-based variable clumping method. International Journal of Geographical Information Science 23, 75–92.

Silverman, B., 1982. Kernel density estimation using the fast Fourier transform. Applied Statistics 31, 93–99.

Silverman, B., 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. J. R. Statist. Soc. B 47, 1–52.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Snow, J., 1855. On the Mode of Communication of Cholera. John Churchill, New Burlington Street, London, England.

Som, N., Monestiez, P., Ver Hoef, J., Zimmerman, D., Peterson, E., 2014. Spatial sampling on streams: principles for inference on aquatic networks. Environmetrics 25, 306–323.

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

Sugihara, K., Satoh, T., Okabe, A., 2010. Simple and unbiased kernel function for network analysis, in: ISCIT 2010 (International Symposium on Communication and Information Technologies), IEEE. pp. 827–832. doi:10.1109/ISCIT.2010.5665101.

Switzer, P., 1965. A random set process in the plane with a Markovian property. Annals of Mathematical Statistics 36, 1859–1863.

Tanaka, U., Ogata, Y., Stoyan, D., 2008. Parameter estimation and model selection for Neyman-Scott point processes. Biometrical Journal 50, 43–57.

Terrell, G., 1990. The maximal smoothing principle in density estimation. Journal of the American Statistical Association 85, 470–476.

Vandenbulcke-Plasschaert, G., 2011. Spatial analysis of bicycle use and accident risks for cyclists. Ph.D. thesis. Université Catholique de Louvain. ISBN 978-2-87558-019-1. Available from www.i6doc.com.

Ver Hoef, J., Peterson, E., 2010. A moving average approach for spatial statistical models of stream networks. Journal of the American Statistical Association 105, 6–18.

Ver Hoef, J., Peterson, E., Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environmental and Ecological Statistics 13, 449–464.

Voss, S., 1999. Habitat preferences and spatial dynamics of the urban wall spider: Oecobius annulipes Lucas. Honours Thesis. Department of Zoology, University of Western Australia.

Voss, S., Main, B., Dadour, I., 2007. Habitat preferences of the urban wall spider *Oecobius navus* (Araneae, Oecobiidae). Australian Journal of Entomology 46, 261–268.

Waagepetersen, R., 2007. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. Biometrics 63, 252–258.

Waagepetersen, R., Guan, Y., 2009. Two-step estimation for inhomogeneous spatial point processes. Journal of the Royal Statistical Society, Series B 71, 685–702.

Wager, C., Coull, B., Lange, N., 2004. Modelling spatial intensity for replicated inhomogeneous point patterns in brain imaging. Journal of the Royal Statistical Society, Series B 66, 429–446.

Wand, M., Jones, M., 1995. Kernel Smoothing. Chapman and Hall.

Warden, C., 2008. Comparison of Poisson and Bernoulli spatial cluster analyses of pediatric injuries in a fire district. International Journal of Health Geographics 7, 51–67.

Webster, S., Diggle, P., Clough, H., Green, R., French, N., 2005. Strain-typing transmissible spongiform encephalopathies using replicated spatial data, in: Baddeley, A., Gregori, P., Mateu, J., Stoica, R., Stoyan, D. (Eds.), Case Studies in Spatial Point Process Modeling. Springer, New York. number 185 in Lecture Notes in Statistics, pp. 197–214.

Xie, Z., Yan, J., 2008. Kernel density estimation of traffic accidents in a network space. Computers, Environment and Urban Systems 32, 396–406.

Yadav, A., Gao, Y., Rodriguez, A., Dickstein, D., Wearne, S., Luebke, J., Hof, P., Weaver, C., 2012. Morphologic evidence for spatially clustered spines in apical dendrites of monkey neocortical pyramidal cells. Journal of Comparative Neurology 520, 2888–2902.

Yamada, I., Thill, J.C., 2004. Comparison of planar and network K-functions in traffic accident analysis. Journal of Transport Geography 12, 149–158.

Yamada, I., Thill, J.C., 2007. Local indicators of network-constrained clusters in spatial point patterns. Geographical Analysis 39, 268–292.

Yue, Y., Loh, J., 2015. Variable selection for inhomogeneous spatial point process models. Canadian Journal of Statistics 43, 288–305.

Yule, G.H., 1903. Notes on the theory of association of attributes in Statistics. Biometrika 2, 121–134.