

Hot Routes: Developing a New Technique for the Spatial Analysis of Crime

Lisa Tompson

Jill Dando Institute of Crime Science, University College London

Henry Partridge

Transport for London; Jill Dando Institute of Crime Science, University College London

Naomi Shepherd

Transport for London

Keywords: Mapping, linear, network, rate, hotspot, transport

Note: A presentation of the methodology and its application was delivered at the 6th National Crime Mapping Conference in 2008 entitled 'Hot routes – developing a technique for the spatial analysis of bus crime in London'. See <http://www.cscs.ucl.ac.uk/events-1/cmc/ncmc2008/hot-routes>

Acknowledgments: The authors gratefully acknowledge the assistance of Claire Bromige, Rachel Carson, Spencer Chainey, Andy Gill, Brian Hall, Edgerton Rodgers and Lucía Summers during the development of the Hot Routes technique. We are also thankful to the anonymous reviewers for their constructive comments.

Abstract

The use of hotspot mapping techniques such as KDE to represent the geographical spread of linear events can be problematic. Network-constrained data (for example transport-related crime) require a different approach to visualize concentration. We propose a methodology called Hot Routes, which measures the risk distribution of crime along a linear network by calculating the rate of crimes per section of road. This method has been designed for everyday crime analysts, and requires only a Geographical Information System (GIS), and suitable data to calculate. A demonstration is provided using crime data collected from London bus routes.

The identification of crime and disorder hotspots is standard practice among community safety practitioners. The development of such hotspot techniques has been primarily directed at crime events that occur as unique points in unconstrained planar space (e.g., residential burglary). Less attention has been devoted to advancing visualization methods where crime - or other unwanted behavior - occurs along linear networks (such as street configurations or transport layouts). Hence, the analytical tools available to crime analysts interested in linear phenomena are currently in their formative years.

This paper considers this developing area and proposes a refinement of the current techniques applied by United Kingdom (UK) practitioners to visualize crime distributions on linear networks. The presented methodology has been devised with the everyday crime analyst in mind – it requires no more than a standard GIS software application with suitable data.

The article proceeds as follows. We begin with a brief overview of how hotspots are used among UK Community Safety practitioners. Second we introduce data that are constrained by networks. Third we provide a critical commentary on conventional hotspot identification methods (for examining both first- and second-order properties) and how these are ill-suited to linear data configurations.

Fourth we review existing spatial analyzes techniques adapted for network-constrained data. Fifth we present our Hot Routes methodology. And finally, we discuss the advantages of using the Hot Routes methodology over conventional techniques such as KDE.

BRIEF OVERVIEW OF BACKGROUND LITERATURE

THE ROLE OF HOTSPOTS IN COMMUNITY SAFETY

Hotspots are generally understood to be areas of higher than average crime concentration (Eck, Chainey, Cameron, Leitner & Wilson, 2005) that can be analyzed owing to the inherent geographical qualities within crime (Chainey & Ratcliffe, 2005). Hotspot analysis is routinely performed by agencies responsible for community safety, as a means of better defining and understanding crime problems. Hotspot maps are a central part of crime analysis (see e.g., Boba, 2005) and are frequently used in various stages of the problem-solving process (Clarke & Eck, 2003). In addition to their visual role, hotspot maps are habitually generated to stimulate debate between practitioners, with the overarching aim of informing decision-makers about the best course of action to take.

Driven by contemporary UK policing philosophies (such as intelligence-led, neighborhood and problem-oriented policing) and fiscal constraints, place-specific interventions are ingrained into operational policing response strategies. The identification of hotspots therefore plays a pivotal role in how agencies choose to respond to crime problems. Identifying where crime concentrates facilitates crime reduction by focusing attention and resources on the places that experience a disproportionate amount of crime or other deviant behavior. Hotspot patrolling has thus become a key tactic to inhibit offending in particular areal sub-units (Braga, 2001; 2005).

In recent years, evaluation literature has materialized which lends support for ‘hotspot policing’ (Sherman, Gartin & Buerger, 1989; Weisburd, Mayer, Sherman, Buerger, Cohn & Petrisino, 1993) as an effective means of reducing high crime counts (or rates) in small, specific locations. For example, in Jersey City, New Jersey, policing was directed at known narcotics hotspots. This reduced the volume of disorder-related emergency calls for service in the hotspot areas and also appeared to have had a diffusion of benefits effect for surrounding areas (Weisburd & Green, 1995). Likewise, in the UK, ‘high visibility patrolling’ or ‘saturation patrols’ are a common – if not the default - tactical response to crime problems.

NETWORK CONSTRAINTS ON CRIME

Traditionally, hotspot analysis has been directed at crimes that are assumed to be situated across an infinite homogeneous environment (e.g., theft of motor vehicle), with little awareness of perceptible geographical restrictions. There has been increasing recognition in recent years that the spatial existence of many phenomena is constrained by networks. Much activity is aligned to the linear networks created by civilization, and crime is no exception.

A superficial consideration of the network types that could be influenced by crime events presents the most obvious candidates: road and rail networks. However, a more imaginative approach might include other types of network: environmental crimes could exist along waterways such as streams, canals, and rivers; and thefts of metal could occur along utility networks such as pipelines. Those sociologically inclined might be able to offer more examples in the way of interpersonal networks.

Crime Pattern Theory (Brantingham & Brantingham, 1984) contends that offenders are influenced by the geographical nature of their daily routine activities (Cohen & Felson, 1979) when they search for crime opportunities. Offenders build up an awareness space that serves as a ‘cognitive map’ of the places they commute through and visit on a daily basis. These theories assist us when imagining what types of networks may be used by offenders, as we can postulate the types of routes they may take to navigate their awareness space.

Offenders use transport in the same way as the non-offending population, and their choice of transportation depends on the geographical setting (e.g., rural or urban); the access they have to

public or private transport; their personal circumstances (e.g., whether they hold a legitimate driving license); and the activities they choose to undertake. In particular, the geographical setting will largely determine the locational behavior of offenders, as activities will be constrained by spatial network configurations such as the road network.

Networks are represented in Geographical Information Systems by a collection of point objects (commonly to denote nodes or intersections) and line objects. Lines can take on multiple forms in a GIS: They can be straight or curved entities. The consequence of this is that their representation in a digital database is "...intrinsically more complex than the storage of point objects where lists of coordinate pairs suffice for most purposes" (O'Sullivan & Unwin, 2003, p. 137). Linear objects therefore require a more sophisticated understanding by those seeking to manipulate and analyze them. This is, perhaps, why many engaged in the routine analysis of crime and disorder are oblivious to the complexities and nuances of performing calculations and applying techniques to linear networks.

CONVENTIONAL HOTSPOT IDENTIFICATION TECHNIQUES AND THEIR SUITABILITY TO LINEAR DISTRIBUTIONS OF CRIME

Second-order processes and the Euclidean distance problem

Crime analysis developed in tandem with data integrity. The widespread availability of point pattern crime data in the 1990's facilitated a proliferation of techniques to assess the second-order properties of such data sets. Methods for analyzing the second-order or global properties of data simply confirm the statistical presence of clustering, or hotspots, without any presumptions about their locations. This incorporates the basic principles of hypothesis testing and classical statistics, where it is assumed (the null hypothesis) that the distribution of point events will follow a pattern of complete spatial randomness (CSR) (see e.g., Kidner, Higgs & White, 2002, p 23). When seeking to define the level of clustering in a point data set, traditional practices involved generating a random point pattern (known as the expected distribution) with which to compare the actual point pattern (known as the observed distribution). Many techniques for identifying clustering or dispersion in data sets are founded on this principle and include Ripley's K-function (Ripley, 1976; 1981; Bailey & Gatrell, 1995; Levine, Watchs & Schirazi, 1986; Jones, Langford & Bentham, 1996) and Nearest Neighbor Distance (Clark & Evans, 1954; Levine, 2004). These tests are recommended prior to the geographical analysis of crime patterns (Eck et al., 2005), as they determine whether or not there is value in pursuing hotspot visualization.

The processes articulated above are well suited for crime data that are arranged across an unconstrained planar space. That is, an infinitely continuous Euclidean space where crime events can occur at any point. It has long since been acknowledged that this is inappropriate when a set of points are constrained or restricted because of a linear network (Krause, 1975). This is because Euclidean distance measurements take the shortest direct path between points - 'as the crow flies' - and uses the hypotenuse of the minimum bounding rectangle, even if the network dictates a longer path (Lu & Chen, 2007 provide an excellent summary of this concept). One example might be two people located at opposing corners of a block with a multi-story building in the center. The distance between them is underestimated using Euclidean distance because to get to one another they would have to walk around two lengths of the block's perimeter - it cannot be traversed another way unless there is an internal path through the building.

Network-derived activities such as transportation take place along a one-dimensional subset of Euclidean space (Miller, 1994). A random pattern on a network may then be wrongly interpreted as a clustered pattern, because the network itself restricts the positioning of the point pattern (Yamada & Thill, 2007). Most applications of geographical crime analysis rely on Euclidean distance to represent the spatial separation between points. The appropriateness of conventional analytical techniques therefore becomes questionable when applying them to linear networks.

The disparity between urban street networks and the Euclidean distance measurement can be overcome by incorporating Manhattan distance into spatial calculations. This is where the distance between two points is the sum of the (absolute) differences of their coordinates. Levine (2004) is one of the few researchers to recognize the limitations of applying Euclidean distance to crime pattern analysis when phenomena are distributed along networks, and thus incorporated Manhattan functionality into his spatial statistics software product known as CrimeStat.

Two flaws prevail though when considering Manhattan distance a worthy substitute for Euclidean distance. First, it may still underestimate the distance between two points on a gridded network, as the absolute difference between the coordinates could fall short of the total network distance between the points. Second, it is naïve to assume that an urban space will contain purely gridded street networks. Most often, there will be a blend of different street networks, sometimes chaotically if they are artifacts of ancient settlements (e.g., London, UK).

First-Order Processes

One of the most germane concepts in spatial analysis is density which is the expression of a number of discrete objects per unit of area. This is known as a first-order process which identifies the locations of data concentration and is relatively easy to measure using polygon areas such as quadrats or administrative boundaries. Consider, for example, calculating the population density for a state or county. An enumeration of the number of events is derived and then divided by the area that the polygon encloses. These types of count (or rate) maps are known as thematic or choropleth maps, as each polygon is thematically shaded corresponding to the count of crimes within its boundary. These approaches have some important limitations pertaining to the Modifiable Areal Unit Problem (we shall not review them here, but for a thorough treatment please see Openshaw, 1984, and Chainey & Ratcliffe, 2005) which led to the development of more sophisticated visualization tools.

Line density analysis in contrast is less intuitive to the layperson and can take multiple forms. de Smith, Goodchild and Longley (2007) maintain that line density statistics can include line frequency (i.e., the number of line segments per unit area), the length of line segments per unit area, the number of line intersections per unit area, or the intersection density (i.e., the number of line sections per unit of length). These measures do rest on just one geographical entity, rather than being concerned with multiple ones that are combined in analysis (e.g., discrete points such as crimes on lines).

KDE is one of the most popular hotspot techniques employed by UK crime analysts (Weir and Bangs, 2007). This takes point data and aggregates it in the GIS within a specified 'moving window' (known as the bandwidth) to create a smooth surface that represents the density of events across an area (more information on this process is available in Eck et al., 2005). The popularity of KDE with practitioners centers on several factors. First, it overcomes the Modifiable Areal Unit Problem. KDE is not reliant on administrative boundaries and is consequently more sympathetic to the underlying distribution of crime. Second, it is easy to generate. The most prominent GIS software companies have devised extensions that allow the user to create KDE maps with a few clicks of the mouse. Third, it is easy to interpret. Hotspots are intuitively represented in red or other warm colors and areas with lower concentrations of crime are shown in cooler colors such as light yellow or blue. Fourth, it is visually pleasing. Compared with point or grid thematic mapping, KDE maps are easy on the eye. Fifth, it is statistically robust. A consistent algorithm is applied to data in a KDE, irrespective of the distribution or nature of that data. Moreover, further statistical tests can be administered to a KDE map output (such as the GI and GI* statistics, see Getis & Ord, 1996) that incorporate statistical significance into hotspot identification.

There are, however, a number of weaknesses in KDE mapping. Foremost is the human tendency to get caught in its 'visual lure' (Eck, et al., 2005). To a layperson or inexperienced crime analyst, the KDE map offers a statistically sophisticated geographic that produces a visually pleasing output. Because it looks nice, people are more inclined to attribute a sense of accuracy to it, rather than critically reviewing the map and questioning its validity or statistical robustness.

Related to this is the influence of thematic classification. A KDE map is more than just a statistical algorithm, and there are several display options that a map maker is able to use. The thematic class classification system (e.g., natural breaks, quantiles, and standard deviations) is one way of introducing ambiguity into maps, as different choices can produce vastly different map outputs. Careful range setting is therefore vitally important to the credibility of the map and crime analysts responsible for making KDE maps are sometimes not aware of this nor alert to the problems it can cause.

KDE requires two parameters to be set prior to running – the grid cell size and the bandwidth (also known as the search radius). Ratcliffe (1999) suggests a useful starting point for determining a suitable grid cell size, and bandwidth choice guidance is offered by Williamson, McLafferty, Goldsmith, Mallenkopf & McGuire (1999). Contentions exist with these parameters though, as the guidelines for setting them are not widely known among crime analysts. Instead poorly trained operational staff are inclined to use the GIS software default options for KDE and are unaware of the influence of these settings.

KDE maps are not popular in all quarters. For example, there is a perception in some UK Police Forces that KDE maps are inaccurate and consequently discourage their crime analysts from generating them. Feeding into this perception is the reality that KDE can, and often does, ‘over-smooth’ the intensity values. Depending on the parameters chosen by the map user, the resulting map output can entirely distort the underlying point pattern on which it is based. For example, an inappropriate choice of bandwidth can result in a map that presents a high intensity of crime in areas that are improbable because of the over-smoothing tendencies inherent in the methodology.

Of the weaknesses pertaining to the KDE technique, perhaps the most important one in the context of linearity is the smoothing effect. KDE is not sympathetic to linear networks and represents crime hotspots as smoothed areas, rather than lines, due to the way it is calculated (with a grid). These areas often comprise geography far beyond the linear network of interest, and therefore hotspots include areas where crime simply could not have occurred. This presents a problem for anyone hoping to gain a deeper understanding of the nature of the crime in these situations, as the mapping output will not be directly relevant to the circumstances. Similarly, if one attempts to represent spatial linear patterns of crime by thematically mapping them by administrative area, the same issues will arise. In recognition of this, Eck et al. state that, “Most clustering algorithms, unfortunately, will show areas of concentration even when a line is the most appropriate dimension” (2005, p. 8).

SPATIAL ANALYSIS TECHNIQUES ADAPTED FOR NETWORK SPATIAL PHENOMENA

A group of Japanese researchers instrumental in this field are based at the University of Tokyo and led by Atsuyuki Okabe. Okabe, Okunuki and Shiode (2006) acknowledge that using planar spatial analysis to study network spatial phenomena can result in false conclusions. In response to this they have developed a toolbox that is compatible with GISs called SANET (Spatial Analysis on a NETwork) which is the culmination of over a decade’s research findings.

Within the SANET toolbox is the ability to perform network K-function methods and network cross K-function methods. These were first proposed in Okabe and Yamada (2001) and later applied by Spooner, Lunt, Okabe and Shiode (2004). Yamada and Thill (2004) also adopted these methods and compared them with the usual planar K-functions for patterns of vehicle accidents in Buffalo, New York. Findings from Yamada and Thill indicated that planar K-function methods overestimated clustering tendencies in comparison to the network adapted methods. Lu and Chen (2007), while taking a different stance on Yamada and Thill’s use of the term ‘false positive’, confirm their general findings when similar analysis was undertaken on vehicle thefts in San Antonio, Texas.

Yamada and Thill (2007) took the network K-function methods a step further to devise a new exploratory spatial data analysis method that they dubbed ‘local indicators of network-constrained clusters’ (LINCS). The purpose of this method is to detect local clusters of events that are constrained

on a network and is founded on the LISA statistics (Local Indicators of Spatial Association) devised by Anselin (1995). Other researchers applied this logic to adapting the LISA statistics for link-attribute-based phenomena which include the local G statistic (Berglund & Karlstrom, 1999) and the local I statistic (Steenberghen, Dufays, Thomas, & Flahaut, 2004).

Conventional KDE, which is typically used for creating hotspot maps for continuous surfaces, applies a bandwidth that above all highlights 'circular' data clusters (Borruso, 2008). Consequently, it is an inadequate way to represent linear data clusters. Innovators such as Borruso (2005; 2008) have countered this problem by developing new methods of estimating density across a network – termed Network Density Estimation (NDE). This involves a modification of the KDE bandwidth based on Euclidean distances to a network-based one where the shortest path departing from each grid cell's center is calculated (following the different line intersections in the network). The bounding polygon of the shortest path distances is then generated to provide an alternative search function for the density analysis (known as the 'network service area'). No weighting functions are imposed during the NDE methodology, so the density value is expressed in terms of both 'events per (linear) kilometer' and 'events per square kilometer' for each reference cell. An application of the NDE methodology to the locations of bank branches and insurance companies in two municipalities led Borruso to conclude that "linear densities seem to be better [at] representing the 'philosophy' of a network-driven analysis" (2008, p. 391)

Also within the SANET toolbox is the ability to use nearest neighbor distance methods that relate to networks (Okabe, Yomono & Kitamura, 1995; Okabe, Okunuki & Shiode, 2004). Further tools are available in SANET, in conjunction with data manipulation tools for network spatial phenomena, and the capacity to generate a random point pattern within a given network. This toolbox makes an important contribution to the analytical community, but only if they are initially aware of it, and of equal importance, if their organizational IT department grants them permission to install it on their internal server (many UK policing agencies can be pernickety about the software that is used).

Practitioners have developed this research area by alternative means. Standard ways of visualising network spatial phenomena have focussed on representing the count of events per street. However, longer street segments will have a greater chance, statistically speaking, of experiencing more events. The resulting map will therefore draw attention to the longest segments, rather than the segments that are of most interest from a risk perspective. For example, LeBeau (2000) used the technique, 'Safe Streets', to map intoxicated pedestrians and drug arrests in Charlotte-Mecklenburg, North Carolina. This involved altering the size and/or color of street segments according to the density of incidents. Van Patten (2007) used a similar method to investigate street robberies in Roanoke, Virginia. He named this technique 'Hot Lanes'. Chainey and Desyllas (2004) took a different approach and used pedestrian footfall data to calculate the risk of being a victim of robbery along certain streets in the West End of London. These examples serve to illustrate that innovation in this field is required to overcome the drawbacks of more traditional methods when sophisticated network-based tools are not available to practitioners.

Crime analysis, and by relation crime hotspot identification, is conducted to inform decision-making by managers. If the crime hotspots do not accurately reflect the spatial patterning of linear crime then these maps are likely to mislead crime control and policy making efforts. For this reason it is important that academics and practitioners collaborate to provide more useful tools for analysis of linear data.

Offered below is a technique that we call 'Hot Routes', an acknowledgment that crime can and does occur on routes between places, as opposed to fixed geographical points. This methodology was born out of a need to develop a better way of visualizing, and consequently understanding, crime that occurred on bus routes that traversed London, UK. Until now, conventional hotspot techniques such as KDE did not provide an output which was sufficiently contextual to derive meaningful inferences from. We therefore devised a methodology that would account for crime risk across the bus network

and facilitate a better understanding of the mechanisms driving crime. We believe that it will appeal to crime analysts and other practitioners as it is simple to calculate, and requires no more than appropriate data and a Geographical Information System. This technique bridges the gap between more sophisticated tools, currently the preserve of academics, and the standard analytical techniques used by practitioners.

THE HOT ROUTES METHODOLOGY

This section documents the steps taken to perform the Hot Routes methodology in a GIS. These instructions are offered alongside an application of the methodology to Driver Incident Reports (DIRs) made on London buses.

Step 1 – DATA REQUIREMENTS

Three primary files are requisite for the Hot Routes analysis: A point pattern data set representing events; a geospatial layer depicting the linear configuration of the network of interest; and a user-defined grid layer with each cell possessing a unique identifying number (e.g., A1, A2, A3 and so on). It is assumed that the linear network layer includes a unique topographic identifier which relates to each individual line. If this is not the case, then a GIS should be used to assign such identifiers.

For present purposes, one year of Driver Incident Reports (DIRs) for a single bus route were extracted ($n = 471$) to be used as the point pattern layer. DIRs are a Transport for London (TfL) data source featuring calls from operational staff, mainly bus drivers, when an emergency response is required. Reports relate to a range of situations on stationary and moving buses. Drivers are connected to London's 24-hour bus control centre (CentreComm) which records the location of incidents including the road name and a set of XY coordinates. These coordinates go through a geocoding process at a later date to improve the accuracy of the data points so a more precise location can be identified. As the point data accuracy cannot be fully guaranteed we have developed steps one through nine (additional to this is an alternative step) to ensure the most accurate final output. If a high geocoding precision is available within a point pattern set, then steps six through eight can be bypassed and the user can perform the alternative step nine in their place.

The address details of the DIR data were manually cleaned - and where appropriate, recoded - by the authors so that each road name was spelled correctly and nearest junctions were noted. These data were specifically recorded as 'disturbances', which is a category that includes a variety of anti-social behaviors such as alcohol and fare related disturbances.

The network layer utilized for the following example was Ordnance Survey's (OS) Mastermap Integrated Transport Network (ITN). This is a georeferenced data set that includes every navigable road - from motorways to alleyways - in the UK. Each road section in these data has a unique topographic identifier, or TOID, and relates to a single row in the attribute table. These road sections start and end at irregular intervals, largely, yet not exclusively at intersections with other roads (i.e., road junctions).

The final data layer was a 250m2 grid, so chosen by the authors because it is the approximate distance between bus stops, and is the standard grid cell size used by Transport for London's partner agencies (such as London's Metropolitan Police Service). Using a point layer of bus stops was considered as an alternative means of splitting the road segments, but discarded as being unsuitable for three reasons. First, although a point layer of bus stop locations exists, these bus stops do not have routing information appended to them, and so discerning which bus stops are used on which route is problematic. Second, some bus stops are only visited when a passenger hails it from the street to get on, or, if already on the bus, signals to the driver that they wish to disembark. Finally, this methodology was devised to be applicable to both single route analysis and multiple route analysis. Using a grid cell layer standardized the units of analysis and was consequently appropriate for both these applications.

Step 2 – PARTITIONING THE NETWORK LAYER WITH THE GRID LAYER

In many scenarios, the linear network of interest will consist of a link-to-node configuration. Succinctly put, this represents a collection of varying length line segments that are delineated by nodes (usually line intersections but not always). The varying lengths may, in some situations, influence the level of precision achievable when later computing the Hot Routes methodology (e.g., longer lines will have a greater chance of events occurring on/near them than shorter lines). To reduce the impact of this problem, we suggest that a GIS is used to further split the linear network when it intersects with a grid cell boundary. This functionality is commonly referred to in a GIS as the ‘split’ command and will result in new rows being created in the attribute table to account for the recently partitioned lines.

For the present case study the 250m² grid layer mentioned in Step 1 was superimposed onto the ITN layer and the split command invoked in the GIS. Each individual road section in the ITN layer was thus partitioned where it intersected with a grid cell boundary. Figure A illustrates this process.

Figure A: Road spanning three grid cells, with the ITN layer in grey and the 250m² grid in black.

For example, the road in the center-left of Figure A that has its road section breaks shown as dotted black lines is First Avenue. First Avenue will be represented in the ITN attribute table as one row before Step 2 is undertaken. After the road sections are partitioned by the grid cell layer, three line segments are produced, each spanning a separate grid cell (see Figure B). Thus, three rows will represent First Avenue in the updated attribute table (see Table 1).

Table 1: Road split at grid cells - attribute table

Legend	TOID	Road Name
Local street single carriageway	4000000030102440	FIRST AVENUE
Local street single carriageway	4000000030102440	FIRST AVENUE
Local street single carriageway	4000000030102440	FIRST AVENUE



Figure A: Road spanning three grid cells, with the ITN layer in grey and the 250m² grid in black.

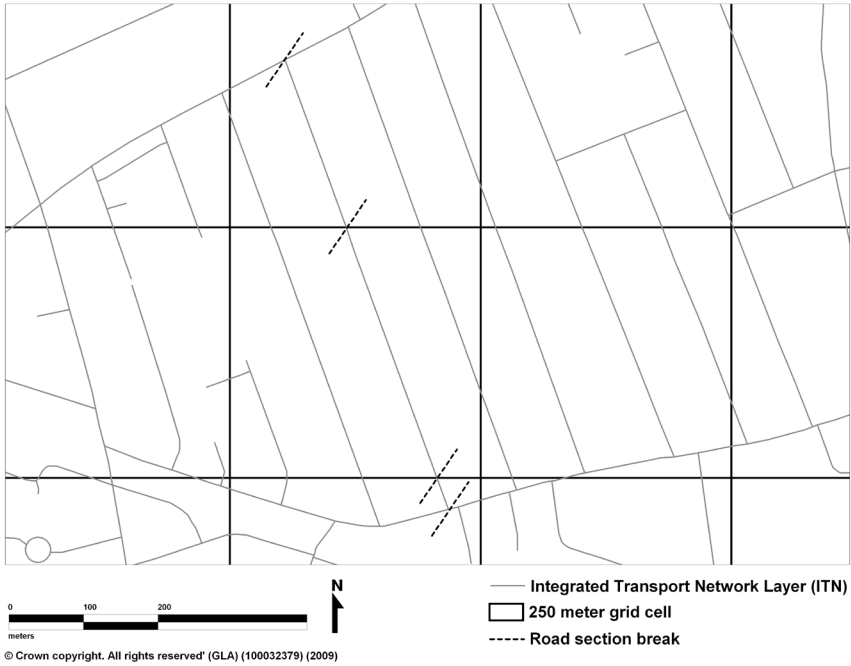


Figure B: Road split at grid cells, with the ITN layer in grey and the 250m² grid in black.

Step 3 – CREATING UNIQUE IDENTIFYING REFERENCES (UIR)s FOR THE NETWORK LAYER

Once Step 2 has been concluded, each line in the network should be represented by a separate row in the network layer attribute table, but has yet to be assigned a unique identifying reference (UIR). This UIR is subsequently used as the linking variable to join to the point pattern data.

To create the UIR in our example, a spatial join between the grid layer and the ITN (network) layer was performed, so that the grid cell reference that related to each line segment was populated in a new field on the ITN layer. This new field with the grid cell reference was then concatenated (the joining of two fields in the attribute table) in the GIS with another unique reference – the road name – to produce the UIR (see Table 2).

Table 2: Multiple road sections with the same name in one grid cell- attribute table

Legend	TOID	RoadName	GridCellRef	UIR
A road single carriageway	4000000030156130	HIGH STREET NORTH	420852	HIGH STREET NORTH - 420852
A road single carriageway	4000000030256340	HIGH STREET NORTH	420852	HIGH STREET NORTH - 420852
A road single carriageway	4000000030360130	HIGH STREET NORTH	420852	HIGH STREET NORTH - 420852
A road single carriageway	4000000030360130	HIGH STREET NORTH	420852	HIGH STREET NORTH - 420852
A road single carriageway	4000000030360130	HIGH STREET NORTH	420852	HIGH STREET NORTH - 420852



Figure C: Multiple road section with the same name in one grid cell

Step 4 – MERGING DUPLICATE UNIQUE IDENTIFYING REFERENCES (UIRs) IN THE NETWORK LAYER

Depending on the nature of the network layer, a problem may be encountered at this stage where a line that has been split by a grid cell contains internal splits (usually but not always defined by intersections with other lines) within the grid cell itself. Thus, multiple lines in the network layer attribute file will have the same UIR and cause problems later in the methodology. The way of overcoming this issue is to get the GIS to perform a merge based on the field containing the UIR. Once this is achieved, each row in the network layer attribute table should refer to a unique line that does not span past the boundaries of a single grid cell. This is demonstrated in Figure C where the dotted black lines distinguish the four different line segments. Table B further reflects this example, and we see from the rows in the attribute table of the ITN layer that three duplicates are present, in addition to the main entry. We called upon the GIS to merge these line entities, so that only one remained, with its corresponding UIR.

Step 5 – CALCULATING THE LINE SEGMENT LENGTHS IN THE NETWORK LAYER

The length of the line segments will be later used as a denominator to construct a rate of events per linear distance (meters in our example). It is therefore necessary to get the GIS to calculate the lengths of each line segment in a new field of the network layer attribute table. Most mainstream GISs have functionality that calculates this, and we utilized this command to produce a length in meters of each of the line segments we had created. This concludes the data manipulation for the ITN (network) layer.

Step 6 – CREATING UNIQUE IDENTIFYING REFERENCES (UIRs) FOR THE POINT PATTERN LAYER

We now turn our attention to the point pattern layer. We needed to create the same UIR in the point pattern layer so it could be used as the linking variable to subsequently join the network layer (ITN) and the point pattern layer (DIRs).

This was achieved by performing a spatial join between the grid layer and the point pattern layer (DIR), so that the grid cell reference that related to each data point was populated in a new field on the point pattern layer. This new field with the grid cell reference was then concatenated (the joining of two fields in the attribute table) in the GIS with another unique reference – the road name – to produce the UIR, in the same way as for the ITN (network) layer.

Step 7 – UPDATING THE POINT PATTERN LAYER WITH THE LINE LENGTH THE DATA POINT OCCURRED ON

A new field should next be created in the point pattern layer and subsequently populated with the length in meters of the line it falls nearest to. Another spatial join is performed, similar to Step 6, in that the line segment length from the network layer (ITN) is appended to each row in the point pattern layer (DIRs) where the UIR in the ITN (network) layer, matches the UIR in the point pattern layer. Hence, once Step 6 and 7 are complete, each data point (and thus row in the attribute table) should have the additional fields of UIR and line length.

Step 8 – WEIGHTING EACH POINT BY THE LENGTH OF THE LINE SEGMENT

Two new fields should now be created in the point pattern layer (DIR) that will be used to calculate a weighted rate of events per linear meter. The first new field should have the value of '1' inserted for each row in the attribute table. Using the GIS, the second field should then contain the results of a calculation whereby the first field (of 1) is divided by the line length that the point occurred on (created in step 7). For example, if an event occurred on a line that was 10 meters long, then the weighted score in the second field would be calculated as 0.1 – reflecting the rate per meter of events.

Step 9 – VISUALIZING THE AGGREGATE EVENT RATE PER LINEAR METER BY THEMATIC SHADING

Finally, once Steps 1-8 are completed, it is now possible to visualize the results by thematically shading each line segment with a color that corresponds to ranges of the incident rate per linear meter. A thematic map should be created of the network layer, using the sum of the event rate per meter field from the point pattern layer (DIR) that is available after step 8 to reflect the spatial variation of the crime rate per line segment. An example of this is shown in Figure D, where red is used to indicate the presence of a high rate of incidents, and yellow a lower incident rate. The thematic map can be formatted to suit the individual user, and a detailed map, or other background information can be layered underneath to provide any necessary contextual information.

Because Steps 1-9 are independent of X and Y coordinates, poorly geocoded points do not skew the analysis. If an incorrectly geocoded point falls outside the correct grid cell, even if it is on the correct road, then it will not be included in the analysis. Similarly, if a data point is in the correct grid, but not geocoded to the right part of the road, then this will be included but will not skew the final map. This and the absence of any smoothing effect means that areas with high event concentrations are clearly and accurately mapped to the linear network.

Alternative step – AN ALTERNATIVE MEANS OF DERIVING THE EVENT RATE PER LINEAR METER

Should the point pattern layer have a high geocoding precision then Steps 1-5 can be adopted

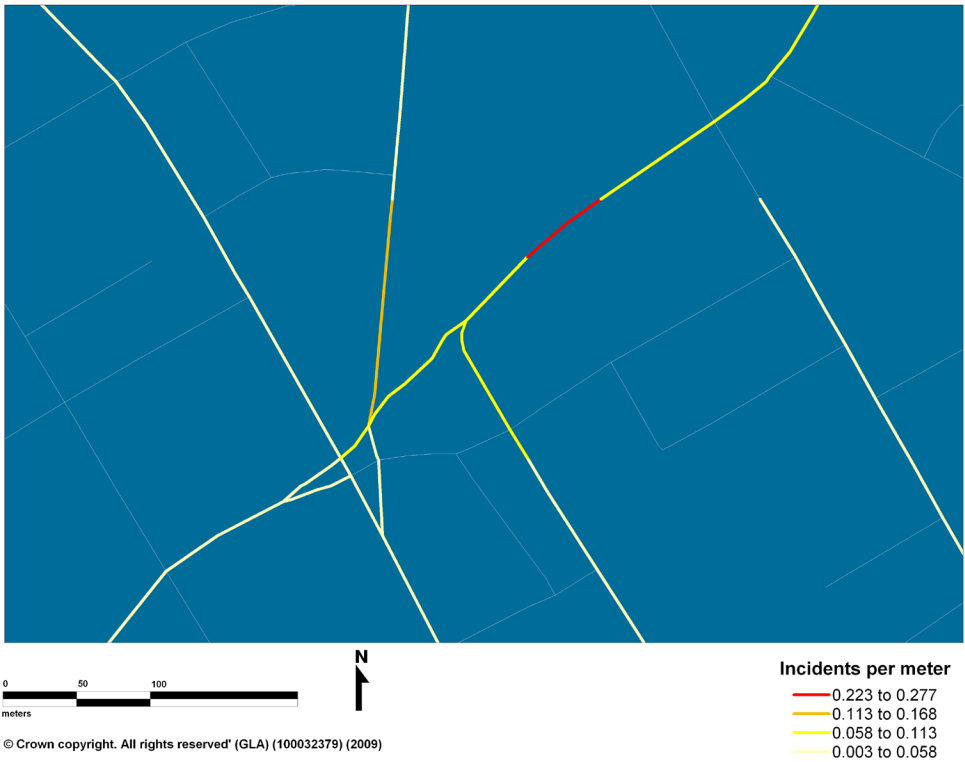


Figure D: Example of a Hot Routes Map, the legend denotes the rate of incidents per linear meter

and then a spatial join undertaken to derive the count of events per line segment (a point-to-line join) for the network layer. Once this has been completed a new field in the network layer should then be created. In this field perform a calculation that divides the aggregated count of events per line segment by the length of the specific line segment (to derive the event rate per linear m). Step 9 can then be resumed to complete the visualization of the results.

THE ADDED VALUE OF HOT ROUTES OVER CONVENTIONAL HOTSPOT IDENTIFICATION METHODS

Hot Routes was developed as a technique that would overcome the problems inherent in using conventional hotspot detection methods such as thematic mapping of administrative boundaries and KDE. We believe that Hot Routes is better suited to highlighting linear clusters of events than other techniques, and present the follow demonstration of the advantages that Hot Route has over conventional KDE.

Figure E shows the distribution of DIRs of disturbances (a dataset used earlier in the methodology section) along this bus route, as reported by bus drivers. This shows an uneven spread of data points across the network, with a small number situated some way off the route (which can be explained by the bus being on diversion or poor geocoding). Some parts of the bus route are free of incidents, but these segments are not large relative to the whole bus route.

With the same data a KDE map was generated. Figure F shows the results of this and illustrates a clear area of high intensity towards the east end of the bus route. In fact, no other part of the bus route appears to have a high intensity from this map. Figure F demonstrates a point we made earlier; that

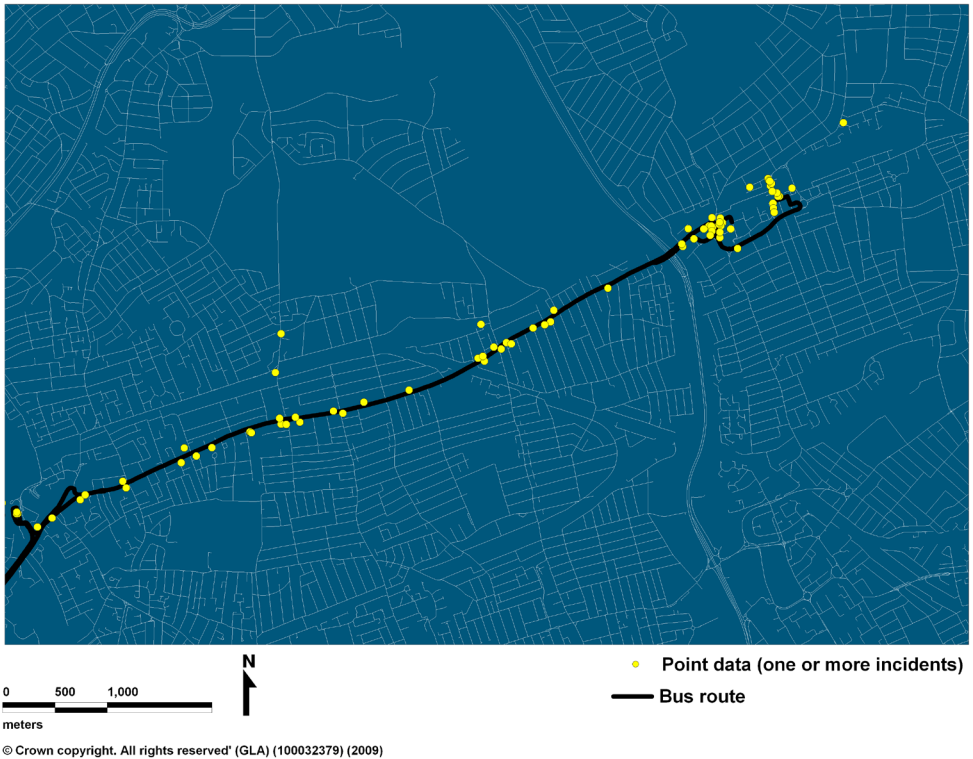


Figure E: Driver Incident Report point data superimposed onto a bus route

KDE is particularly good at detecting circular clusters of data. When we compare Figures E and F, it is noticeable that the points in Figure E which fall in the vicinity of the high intensity area shown in Figure F are at two distinct parts of the bus route, with a section of the route in between that has no offenses. Thus the KDE map has a tendency to over-smooth across two parts of the bus route that experience a high volume of incidents, but fails to represent other clusters of incidents on the route. The KDE map does not assist the reader in understanding the true spatial pattern of the bus-related crime.

Figure G utilizes the Hot Routes methodology to visualize the incident data used in Figures E and F and is thematically shaded according to the incident rate per route meter. The thematic shading parameters have been deliberately chosen to mirror those used in the KDE map in Figure F to facilitate comparison. Visual inspection of Figure G reveals a distinctly red section at the eastern end of the route (indicating a high rate of incidents), which corresponds with what the KDE map showed us. In addition to this information though, it is clear that incidents exist at other parts of the bus route (denoted by the yellow and orange shaded lines) and not along other parts of the route (the black sections of road with no thematic shading). Furthermore, there are a couple of instances where thematically shaded road sections appear away from the bus route – informing us that something unusual has occurred, such as a diversion. A lot more information is thus discernible from the Hot Routes map than the comparator KDE map.

Figure H shows the eastern KDE hotspot in closer detail. The most intense area of the hotspot lies away from the line of the bus route, indicative of the ‘over-smoothing’ limitations raised previously. Indeed, the red area with the highest intensity of data points sits north of the bus route and does not appear to follow the linear network at all.

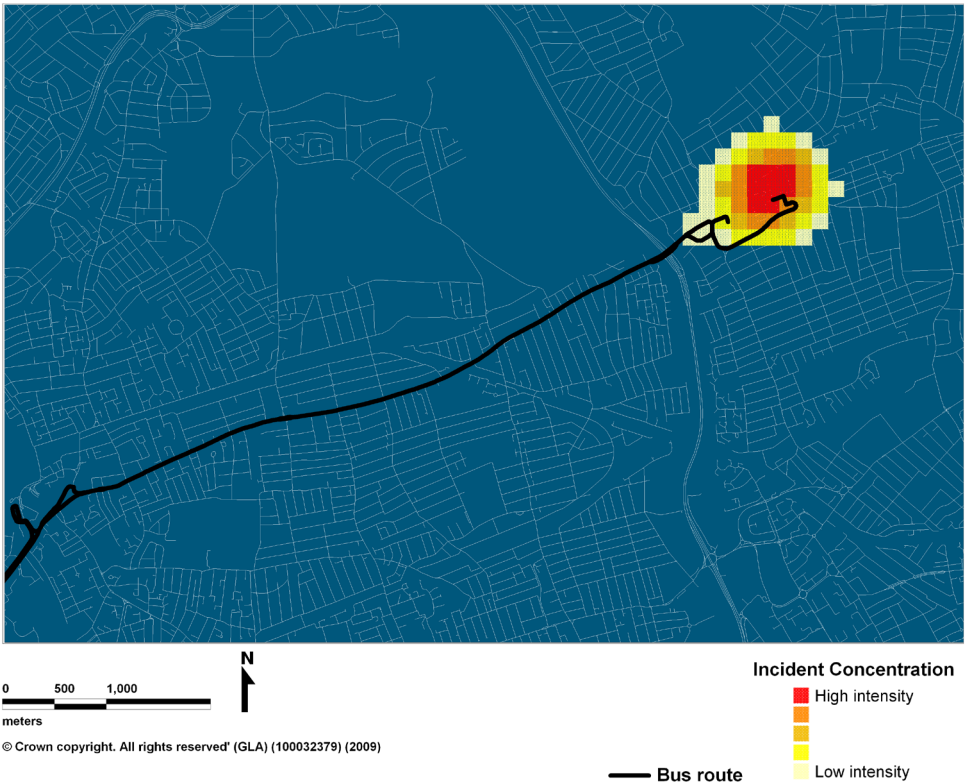


Figure F: A conventional KDE map of the Driver Incident Report data

Figure I presents the same level of magnification as Figure H, but shows the intensity of the incident rate using the Hot Routes method. This clearly demonstrates the ability of the Hot Routes method to ‘zoom in’ on an area and still provide a meaningful picture of the events at a level of detail appropriate to the magnification. Two roads are identifiable as having the highest concentration of incidents. However, one section follows the line of the bus route (shown with the black line below it) and the other runs north, away from the bus route. Further inspection of the data reveals that a bus depot lies at the northern end of the route. Such an inference could be made from the KDE hotspot map but with much less precision, and considerably more data drilling and interrogation by practitioners. Notice that a road in the extreme top-right hand corner of Figure I is thematically shaded. The incidents underlying this have not affected the visualization of the Hot Routes map because the method does not rely on coordinates. However, it is plausible to assume that this might have skewed the KDE map.

A CRITIQUE OF THE HOT ROUTES METHODOLOGY

The Hot Routes methodology is not intended to imitate the complex statistical algorithms of academically available spatial analyzes tools (e.g., Network Density Analysis as outlined previously). Instead, it is a practical, real world solution for everyday crime analysts working within tight operational deadlines. We do though acknowledge that Hot Routes suffers from the same thematic classification problem that plagues other techniques. Other limitations are documented to provide a critique of the Hot Routes methodology. We invite others to contribute to this debate.

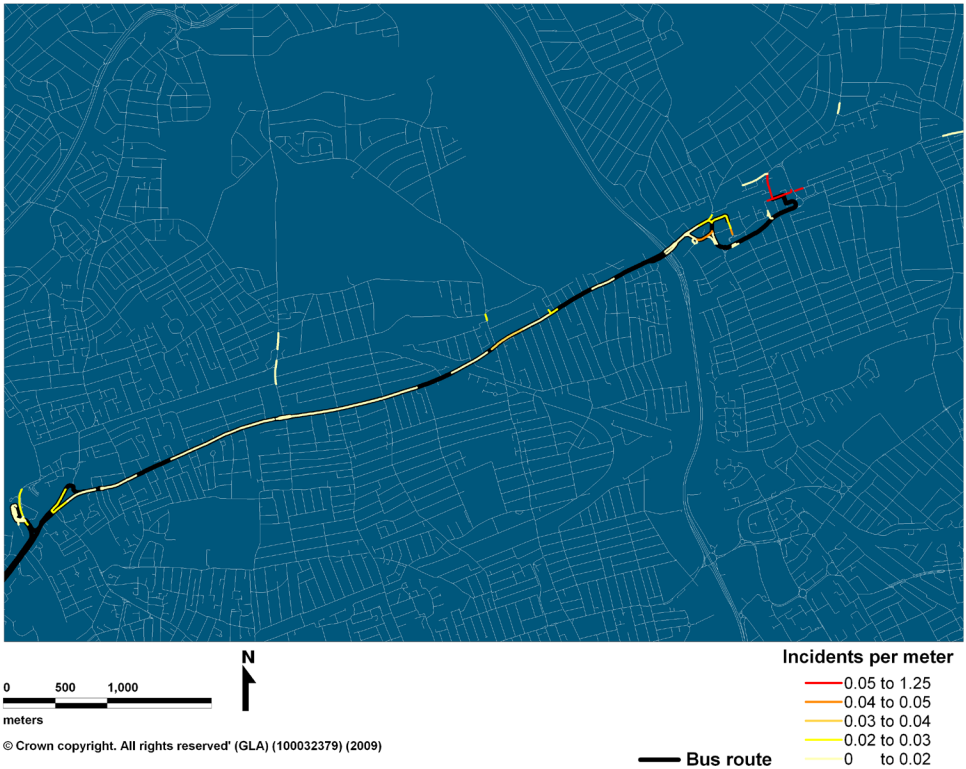


Figure G: A Hot Routes map of the Driver Incident Report data

The weaknesses of the Hot Routes method largely stem from the nature of the data that is used by the authors in their organization. This includes the data used for the purpose of this paper – Driver Incident Reports – collected from London bus drivers. These data problems can be seen to fall into two general categories:

- a) Data recording practices
- b) Unknown crime locations

Discussing these in turn, data recording practices relating to crime and unwanted behavior on London's transport network are still being refined for analytical purposes. For example, sometimes a (bus or train) station name is recorded for an incident rather than a road name. This then impacts on the geocoding precision of such incidents, and at present overcomplicates the Hot Routes methodology. The authors are currently progressing work relating to a simple gazetteer which will automatically replace the location name with a road name. This will facilitate the joining of previously unmatched data and remove the current requirement for the data to be manually cleaned and recoded.

Secondly, transport-related crime suffers from the same limitation as many other types of crime – that of an unknown location. For example, a victim of pick pocketing may only become aware that their wallet has been taken long after they have alighted from a bus. They will only know the two points (which represent stations, bus stops, etc.) between which it could have occurred. In this situation the crime will likely be recorded randomly or at some default location along the route. This poses a significant problem for the accuracy and confidence of the crime distribution. If crimes are either arbitrarily or repeatedly assigned to the same default location a potentially misleading concentration of offenses could appear to occur along one section of road.

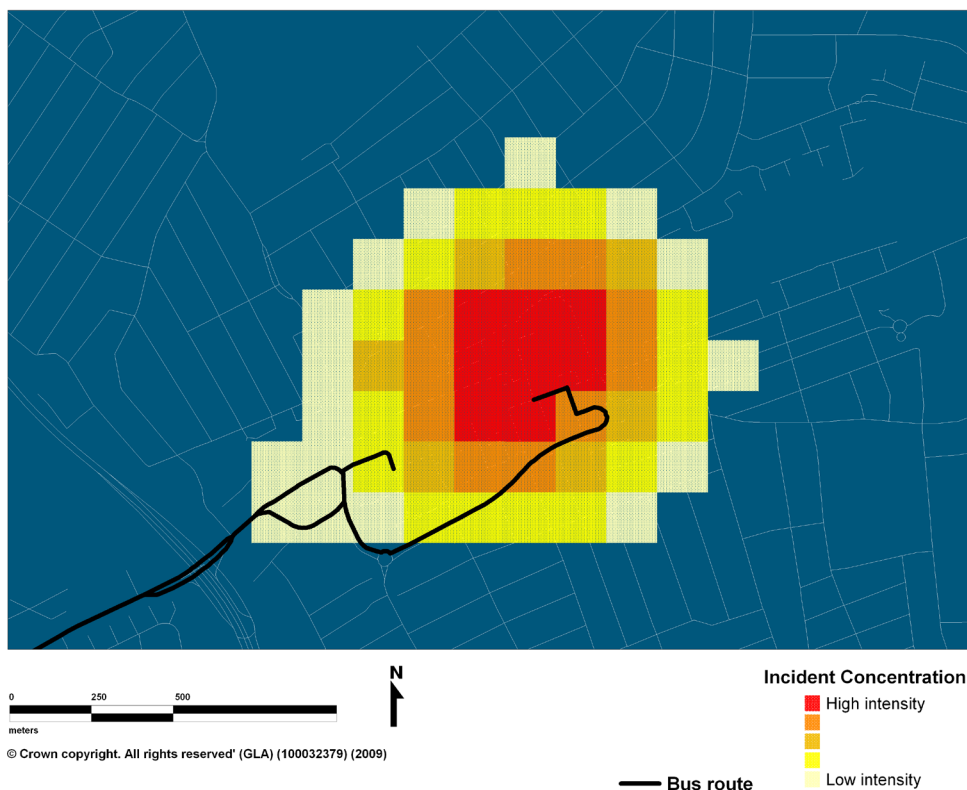


Figure H: KDE map, zoomed extent of the hotspot

One way of overcoming this problem could be to weight incidents that occur on roads that cross multiple 250m² grid squares by the probability that they occurred on a particular section of those roads. This would involve a similar method to aoristic network analysis (Gill, 2007, based on the aoristic temporal analysis technique devised by Ratcliffe, 2000) which seeks to overcome the problem of not knowing the precise location of a crime event. To demonstrate, in our ‘Hot Routes’ example,

The combined length of Oxford Street is 450m

Section 1 (Grid Ref 567897, length 100m)

– An incident has a 22 per cent chance of occurring along this section

Section 2 (Grid Ref 567898, length 250m)

– An incident has a 55 per cent chance of occurring along this section

Section 3 (Grid Ref 567899, length 100m)

– An incident has a 22 per cent chance of occurring along this section

The final limitation of the Hot Routes methodology relates to the resolution it is viewed at. At small, local scales it is entirely appropriate, and depicts small sections of the bus route which are operationally useful. However, when one zooms out to view the thematic map at a pan-London resolution, the detail is compromised by viewing the thousands of road sections en masse. The patterns that the Hot Routes methodology portrays are not suitable, nor meaningful, viewed at a city-wide level. The authors are currently working on a solution to this limitation.

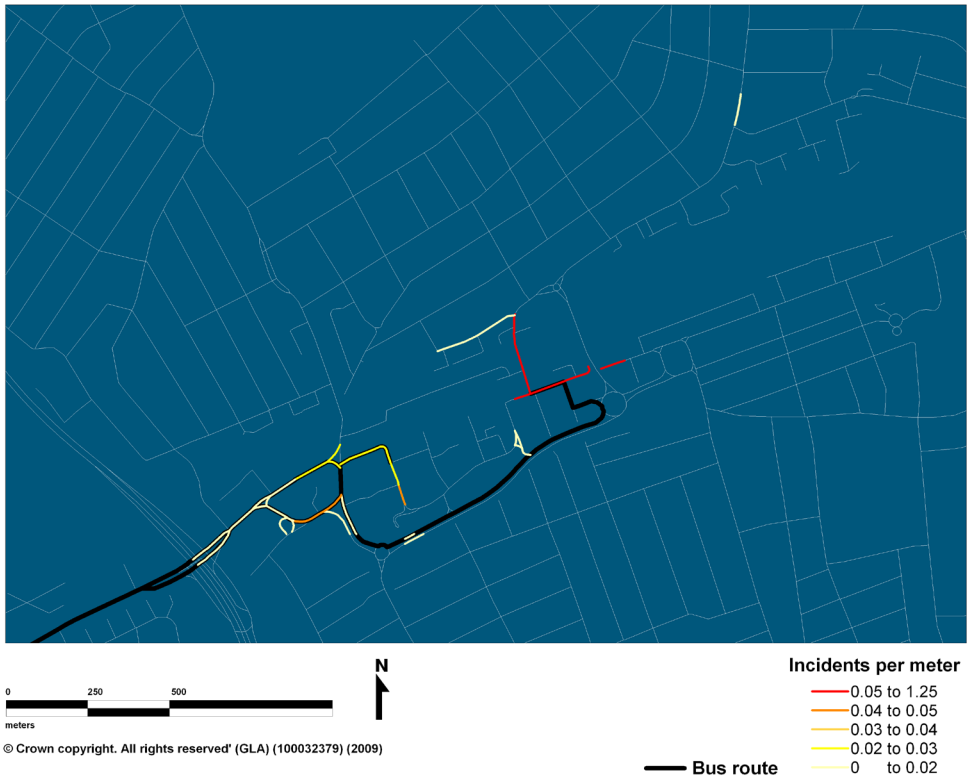


Figure 1: Hot Routes map, zoomed extent of the hotspot

DISCUSSION

The limitations of conventional hotspot mapping (e.g., KDE) as a method of representing linear events, particularly incidents along a bus route, were the stimulus for the creation of the Hot Routes technique. This methodology calculates the rate of crimes per linear meter of a bus route in order to represent risk distribution along a linear network. However, the technique itself is not limited to this application. It requires no great leap of the imagination to conceive of its employment in the spatial analysis of personal robberies, street drug dealing, road traffic collisions, the illegal poaching of fish populations, theft of metal along utility networks, contamination of waterways due to the illegal dumping of toxic waste, and even insurgent activity (e.g., roadside bombs) in Iraq.

The Hot Routes technique taps into a current debate in geographical criminology which refers to units of analysis (Weisburd, Bernasco & Bruinsma, 2009). The scholarly fascination with micro-units of place is a contemporary issue, but one that is likely to be augmented by developments such as the technique presented here. The Hot Routes method fulfils the desire to understand crime variation at a small unit of analysis, as the accuracy of the method is not compromised when the geographical resolution is small.

To practitioners, the benefits of achieving local level understanding of risk distribution along a linear network are manifold. Following the production of Hot Routes thematic maps, Safer Transport Teams (teams of police officers) can be deployed along specific sections of a route that are highlighted as being at high risk of crime occurring. This level of precision focuses resources in a precise way,

rather than instructing them to target a particular bus route, which may span many miles across the city. Moreover, these maps can be shown to partner agencies to encourage discussion about the possible causes for high-risk sections of the route.

The Hot Routes method allows other mapping techniques to be used simultaneously to augment understanding and to test hypotheses. For example, general crime (not related to the bus network) could be mapped with KDE and used to compare levels of crime on a bus with the areas it traverses through. Newton (2008) argues that incidents of bus crime tend to be highest on bus routes that travel through high crime areas. In London approximately 80% of bus crime in 2003 occurred in the 20% of areas with the highest levels of crime generally (Newton, 2008). While plotting the course of a bus route through high crime areas enables the viewer to see which sections are at highest risk of crime based on this inference, the Hot Routes method would allow the viewer to identify concentrations of bus crime along actual sections of the route, thus testing this hypothesis. This would enable the viewer to determine whether concentrations of bus crime are unique to the internal environment of the bus or influenced by the external environment the bus passes through. In addition, one could compare levels of crime and disorder on a section of route with the physical and socio-demographic characteristics of the surrounding area.

CONCLUSION

This paper presents a method of visualizing incidents that follow a linear pattern (such as street robberies, car crashes, bus related crime and railway cabling theft) by linking the address detail of point data to road segments. The reliance of the Hot Routes method on street names and grid references rather than on the distribution of data points overcomes a number of the weaknesses associated with conventional hotspot mapping techniques. These include poor geocoding, over-smoothing, and unclear range settings. It also enhances the visualization of data. We have shown that KDE hotspot maps provide a 'global' picture of the spread and concentration of incidents, but the Hot Routes technique gives a more localized, street level view which is superior for the purpose of small-scale analysis.

The Hot Routes method is easy to perform and versatile in its application. GIS software is widely available and any linear based data can be visualized using the technique. The authors have identified a number of academics and practitioners who have experimented with Hot Routes type methods in the past. The work presented here is a development of these ideas and further refinements should enhance the functionality of the technique and perhaps facilitate its wider use.

To borrow the conclusion from Lu and Chen (2007, p 630)

...the network K-function is not yet available in the commonly used software packages for crime analysis. Developing the appropriate crime analysis tools to incorporate the network K-function and probably many other important spatial analysis techniques should be on the top of the to-do list to promote the correct application of spatial statistics to crime analysis.

We suggest that until this day comes, the Hot Routes methodology offers an important contribution to the crime analysis community.

REFERENCES

- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93–115.
- Bailey, T. C. & Gatrell, A.C. (1995). *Interactive Spatial Data Analysis*. Burt Hill, UK: Longman Scientific and Technical.
- Berglund, S. & Karlström, A. (1999). Identifying Local Spatial Association in Flow Data. *Journal of Geographical Systems*, 1, 219–36.
- Boba, R. (2005). *Crime Analysis and Crime Mapping: An Introduction*. Thousand Oaks, CA: Sage Publications.
- Borruso, G. (2005) Network Density Estimation: Analysis of point patterns over a network. In Gervasi, O., Gavrilova, M. L., Kumar, V., Laganà, A., Lee, H.P., Mun, Y., Taniar, D. & Tan, C.J.K. (eds). *Computational Science and Its Applications*. (ICCSA 2005). Berlin, Springer Lecture Notes in Computer Science. No 3482, 126–32.
- Borruso, G. (2008). Network Density Estimation: A GIS Approach for Analyzing Point Patterns in a Network Space. *Transactions in GIS*, 12(3), 377–402.
- Braga, A. (2001). The effects of hot spot policing on crime. *The ANNALS of the American Academy of Political and Social Science*, 578, 104–25.
- Braga, A. (2005). Hot spots policing and crime prevention: A systematic review of randomized controlled trials. *Journal of Experimental Criminology*, 1, 317–42.
- Brantingham, P. L. & Brantingham, P. J. (1984). *Patterns in Crime*. New York: Macmillan.
- Chainey, S.P. & Desyllas, J. (2008). Modelling pedestrian movement to measure on-street crime risk. In Liu, L. & Eck, J. (ed.). *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Hershey: Idea Group Inc.
- Chainey, S. & Ratcliffe, J. (2005). *GIS and Crime Mapping*. West Sussex: Wiley.
- Clark, P.J. & Evans, F.C. (1954). Distance to nearest Neighbor as a measure of spatial relationships in population. *Ecology*, 35, 45–453
- Clarke, R.V. & Eck, J. (2003). *Become a Problem-Solving Crime Analyst: In 55 Small Steps*. London: UCL Jill Dando Institute of Crime Science.
- Cohen, L. E. & Felson, M. (1979). Social Change and crime rate trends: A routine activity approach. *American Sociological Review*, 44, 588–608.
- de Smith, M., Goodchild, M.F. & Longley, P.A. (2007). *Geospatial analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Troubador, Leicester.
- Eck, J., Chainey, S., Cameron, J., Leitner, M., & Wilson, R. (Eds.) (2005). *Mapping Crime: Understanding Hot Spots*. USA: National Institute of Justice. Accessed from <http://www.ncjrs.gov/pdffiles1/nij/209393.pdf> on 15 September 2008.
- Getis, A. & Ord, J.K. (1996). Local spatial statistics: an overview. In Longley, P. & Batty, M. (ed). *Spatial analysis: modelling in a GIS environment*. GeoInformation International, London.
- Gill, A., (2007). Developing aoristic network analysis upon London's transport system. Presentation delivered at the National Crime Mapping Conference 2007. Accessed from <http://www.cscs.ucl.ac.uk/events-1/cmc/cmc-2007/developing-aoristic-network-analysis-upon-london2019s-transport-system> on 27 November 2008.
- Jones, A.P., Langford, I.H., & Bentham, G. (1996). The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. *Social Science and Medicine*, 42(6), 879–885.
- Kidner, D.B., Higgs, G. & White, S.D. (2002). *Socio-economic Applications in Geographical Information Science*. Innovations in GIS 9. London, Taylor and Francis.
- Krause, E.F. (1975). *Taxicab Geometry*. Addison-Wesley Publishing Company, Inc., USA.
- LeBeau, J. L. (2000). *Demonstrating the Analytical Utility of GIS for Police Operations: A Final Report*. US: NCJRS. Accessed from <http://www.ncjrs.gov/App/Publications/abstract.aspx?ID=187104> on 1 October 2008
- Levine, N. (2004). *CrimeStat: a spatial statistics program for the analysis of crime incident locations* (version 3.0). Ned Levine and Associates, Houston, TX and the National Institute of Justice, Washington, DC.
- Levine, N., Watchs, M. & Schirazi, E. (1986). Crime at bus stops: a study of environmental factors. *Journal of Architectural and Planning Research*, 3(4), 339–361.
- Lu, Y. & Chen, X. (2007). On the false alarm of planar K-function when analyzing urban crime distributed along streets. *Social Science Research*, 36(2), 611–632.
- Miller, H.J. (1994). Market area delimitation within networks using geographic information systems. *Journal of Geographic Systems*, 1(2), 157–173.
- Newton, A. (2008). A Study of Bus Route Crime Risk in Urban Areas: The Changing Environs of a Bus Journey. *Built Environment*, 34(1), 88–103.
- Okabe, A., Yomono, H. & Kitamura, M. (1995). Statistical analysis of the distribution of points on a network. *Geographical Analysis*, 27(2), 152–175.

- Okabe, A. & Yamada, I. (2001) The K-function method on a network and its computational implementation. *Geographical Analysis*, 33(3), 271–290.
- Okabe, A., Okunuki, K. & Shiode, S. (2004). SANET: A Toolbox for Spatial Analysis on a Network. Tokyo, Japan: Center for Spatial Information Science, University of Tokyo.
- Okabe, A., Okunuki, K. & Shiode, S. (2006). SANET: A Toolbox for Spatial Analysis on a Network. *Geographical Analysis*, 38, 57–66.
- Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, 38, 41.
- O'Sullivan, D. & Unwin, D. J. (Eds.) (2003). *Geographic Information Analysis*. New Jersey. Wiley and Sons Ltd.
- Ratcliffe, J. H. (1999). Hotspot Detective for MapInfo Helpfile Version 1.0. Accessed from <http://jrattcliffe.net/hsd/index.htm> on 1 October 2008.
- Ratcliffe, J. H. (2000) Aoristic Analysis: the spatial interpretation of unspecific temporal events. *International Journal of Geographical Information Science*, 14(7), 669–679.
- Ripley, B.D. (1976). The Second-order Analysis of Stationary Point Process. *Journal of Applied Probability*, 13, 255–66.
- Ripley, B.D. (1981). *Spatial statistics*. Wiley, New York.
- Sherman, L.W., Gartin, P. & Buerger, M.E. (1989). Hot Spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27(1), 27–55.
- Spooner, P., Lunt, I., Okabe, A. & Shiode, S. (2004). Spatial analysis of point patterns on road networks using the network K-function. *Landscape Ecology*, 19(5), 491–499.
- Steenberghen, T., Dufays, T., Thomas, I. & Flahaut, B. (2004). Intra-Urban Location and Clustering of Road Accidents Using GIS: A Belgian Example. *International Journal of Geographical Information Science*, 18(2), 169–81.
- Van Patten, I. (2007). Robbery Hot Lanes: Southeast Roanoke. Accessed from <http://www.radford.edu/~ivanpatt/crimemaps/HotLanes.pdf> on 1 October 2008.
- Weir, R. & Bangs, M. (2007). *The Use of Geographic Information Systems by Crime Analysts in England and Wales*. Home Office Online Report Series. London: Home Office.
- Weisburd, D., Bernasco, W. & Bruinsma, G.J.N. (Eds.). (2009). *Putting Crime in its Place: Units of Analysis in Geographic Criminology*. Springer: New York
- Weisburd, D. & Green, L. (1995). Policing drug hot spots: The Jersey City drug market analysis experiment. *Justice Quarterly*, 12(4), 711–735.
- Weisburd, D., Maher, L., Sherman, L., Buerger, M., Cohn, E. & Petrisino, A. (1993). Contrasting crime general and crime specific theory: The case of hot spots of crime. In Alder, F. & Laufer, W.S. (Eds.), *New directions in criminological theory* (45–70). London: Transaction publishers.
- Williamson, D., McLafferty, S., Goldsmith, J., Mallenkopf, J. & McGuire, P. (1999). A Better Method to Smooth Crime Incident Data. *ESRI ArcUser Magazine*. January-March, 1–5. Accessed from <http://www.esri.com/news/arcuser/0199/crimedata.html> on 1 October 2008.
- Yamada, I. & Thill, J. (2004). Comparison of planar and network K-functions in traffic accident analysis. *Journal of Transport Geography*, 12, 149–158.
- Yamada, I. & Thill, J. (2007). Local indicators of network-constrained clusters in spatial point patterns. *Geographical Analysis*, 39, 268–292.