Accident Analysis and Prevention 41 (2009) 359–364

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Kernel density estimation and K-means clustering to profile road accident hotspots

Tessa K. Anderson

University of Queensland, School of Geography, Planning and Environmental Management, Brisbane, QLD 4072, Australia

ARTICLE INFO

Article history:
Received 10 September 2008
Received in revised form 17 November 2008
Accepted 4 December 2008

Keywords:
Accident
GIS
Clustering
Policy

ABSTRACT

Identifying road accident hotspots is a key role in determining effective strategies for the reduction of high density areas of accidents. This paper presents (1) a methodology using Geographical Information Systems (GIS) and Kernel Density Estimation to study the spatial patterns of injury related road accidents in London, UK and (2) a clustering methodology using environmental data and results from the first section in order to create a classification of road accident hotspots. The use of this methodology will be illustrated using the London area in the UK. Road accident data collected by the Metropolitan Police from 1999 to 2003 was used. A kernel density estimation map was created and subsequently disaggregated by cell density to create a basic spatial unit of an accident hotspot. Appended environmental data was then added to the hotspot cells and using K-means clustering, an outcome of similar hotspots was deciphered. Five groups and 15 clusters were created based on collision and attribute data. These clusters are discussed and evaluated according to their robustness and potential uses in road safety campaigning.

1. Introduction

Road accidents are a consequence of the increased mobility of today's society. The impact of road traffic accidents in terms of injuries and fatalities is a social and public health challenge. Road safety is of great concern to the UK government highlighted by the most recent report 'Tomorrow's Roads: Safer for Everyone' (UK Department of Transport, 2000) which outlines the current trends and statistics in road accidents. The report offers projections for the future and a best practice policy guideline to reduce these numbers. Every year approximately 3500 people are killed on Britain's roads and 40,000 are seriously injured (UK Department for Transport, 2000). The World Health Organisation estimates 1.18 million people were killed in 2002 in a road collision which is 2.1% of the global mortality (WHO, 2004). Road traffic accidents according to the WHO are the leading injury related cause of death among people aged 10–24.

Identifying accident hotspots and appending value added data to understand the processes occurring in these hotspots is important for the appropriate allocation of resources for safety improvements. By identifying road accident hotspots, using Geographical Information Systems (GIS) and appending value added data, a more robust understanding can be gained, with regards to indicators of casual effects. GIS is a technology for managing and processing locational and related information (Longley et al., 2005). Using GIS as a platform to perform this research is fundamental as it enables the efficient manipulation, analysis and visualisation of spatial data.

The foci of this paper are (1) to present a methodology for the identification of high density accident zones using GIS and kernel density estimation (KDE) (2) to add attribute data to the accident zones, and (3) to identify similar zones using a K-means clustering algorithm with regards to attribute data and compare. This methodology and clustering technique uses 5 years (1999–2003) of road accident data for the metropolitan area of London as the study region.

2. Background

The road accident literature provides no universally accepted definition of a road accident 'hotspot'. Hauer (1997) describes how researchers rank locations according to accident rate while other researchers use accident frequencies (accident per road kilometre). Road accident hotspot analysis has traditionally centred on road segments or specific junctions (Thomas, 1996) while area wide hotspots and the spread of risk which is produced from a collision is somewhat neglected. Traditionally since the late 1970s statistical models have been applied to road accident analysis, however the early models where flawed in such that they would assume accidents to be normally distributed (Oppe, 1979; Ceder and Livneh, 1982). The next stage was to accommodate this statistical drawback was to use Poisson log linear regression to account for the randomness of accidents in time and space (Blower et al., 1993). Many authors such as Hauer and Persaud (1987), Miaou (1994),

E-mail address: t.anderson3@uq.edu.au.

0001-4575/$ – see front matter
doi:10.1016/j.aap.2008.12.014

Shankar et al. (1995), Maher and Summersgill (1996), Abdel-Aty and Radwan (2000) have used negative binominal regression models. In all of these models only the reported number of accidents in the observed time period is used and locational characteristics are therefore modelled as constant within a given time period.

This paper in contrast is based on the assumption that road accidents occurring in a similar area, not just taking into account the road network or junctions, are spatially dependent because of the increased density of accidents in a specific area. This dependence is argued to be the result of a shared common cause(s) between the accidents, albeit of varying intensity (Flahaut et al., 2003; Flahaut, 2004). While human error and mechanical failure can be causes of road traffic accidents the importance of spatial factors has been 'grossly underestimated' (Whitelegg, 1987). This paper is based on the assumption that road accidents occurring in similar areas are spatially dependent. The existence of hotspots comes from the awareness of the evident spatial interaction existing between contiguous accident locations (Flahaut et al., 2003). Accident hotspots expose concentrations and therefore suggest spatial dependence between individual occurrences which may be due to one or several common causes.

Road accident hotspot analysis requires a comprehensive understanding of the vehicle accident involvement process, the severity of resultant injuries, and the surrounding road environment. A GIS platform is particularly suited to this type of problem because it provides an efficient system of linking a large number of disparate data bases. It provides a spatial referencing system for reporting output at different levels of aggregation, it allows input of both historical and statistical accident experience in estimating accident risk at different locations and times, and it allows controls on a myriad of risk factors explaining variations in accident involvement and injury severity. The most straightforward use of GIS for accident analysis is the examination of spatial characteristics of accident locations (Steenberghen et al., 2004).

Classifications of road accident hotspots are generally based on the available data associated with the accident itself (namely, time of day, type of victim, type of vehicle). This can often limit the scope of understanding of the complexities of road accident hotspots which can be the outcome of a number of environmental, social and economic factors neglected by the standard accident data collection. The use of 'traditional' typologies in road accident analysis has been prolific throughout the literature. Examples include delineating road users (Pietro, 2001; Oxley et al., 2005) and classifying by temporal and spatial measures and also types of accident (Levine et al., 1995). A small proportion of the research has focused on interpreting the spatial element into the typologies most notably Pulugurtha et al. (2003) who focused on ranking pedestrian hotspots. This paper presents a spatial methodology which investigates the variables which might be present at dependent spatial accidents in order to develop a robust understanding of the processes at work.

In general, methods which determine road accident hotspots rarely incorporate variables from other sources relating to land use and the environment. The literature review did not identify any research which identifies hotspots and goes on to cluster the hotspots and create a typology based on spatial indicators. Therefore this paper identifies an original method which uses GIS and kernel density estimation to create a basic spatial unit of a hotspot and statistical analysis to cluster the hotspots according to attribute data of the accident including the surrounding area.

3. Methodology to identify high density accident zones

There are a variety of spatial tools developed to assist the understanding of the changing geographies of point patterns. The most promising of these tools is kernel density estimation (Chainey and Ratcliffe, 2005; Sabel, 2006). There are many advantages of kernel density estimation (KDE) as opposed to statistical hotspot and clustering techniques such as K-means. The main advantage for this method lies in determining the spread of risk of an accident. The spread of risk can be defined as the area around a defined cluster in which there is an increased likelihood for an accident to occur based on spatial dependency. Secondly by using this density method, an arbitrary spatial unit of analysis can be defined and be homogenous for the whole area which makes comparison and ultimately a taxonomy possible.

Kernel density estimation involves placing a symmetrical surface over each point and then evaluating the distance from the point to a reference location based on a mathematical function and then summing the value for all the surfaces for that reference location. This procedure is repeated for successive points. This therefore allows us to place a kernel over each observation, and summing these individual kernels gives us the density estimate for the distribution of accident points (Fotheringham et al., 2000).

$$f(x,y)=\frac{1}{nh^2}\sum_{i=1}^{n}K\left(\frac{d_i}{h}\right) \tag{1}$$

where $f(x, y)$ is the density estimate at the location (x, y); n is the number of observations, h is the bandwidth or kernel size, K is the kernel function, and d_i is the distance between the location (x, y) and the location of the ith observation. The effect of placing these humps or kernels over the points is to create a smooth and continuous surface. The method is known as KDE because around each point at which the indicator is observed a circular area (the kernel) of defined bandwidth is created. This takes the value of the indicator at that point spread into it according to some appropriate function. Summing all of these values at all places, including those at which no incidences of the indicator variable were recorded, gives a surface of density estimates. Density can be measured by two methods; simple and kernel. The simple method divides the entire study area to predetermined number of cells and draws a circular neighbourhood around each cell to calculate the individual cell density values, which is the ratio of number of features that fall within the search area to the size of the area. Radius of the circular neighbourhood affects the resulting density map. If the radius is, increased there is a possibility that the circular neighbourhood would include more feature points which results in a smoother density surface (Silverman, 1986). The kernel method divides the entire study area into predetermined number of cells. Rather than considering a circular neighbourhood around each cell (the simple method), the kernel method draws a circular neighbourhood around each feature point (the accident) and then a mathematical equation is applied that goes from 1 at the position of the feature point to 0 at the neighbourhood boundary (see Fig. 1). Road

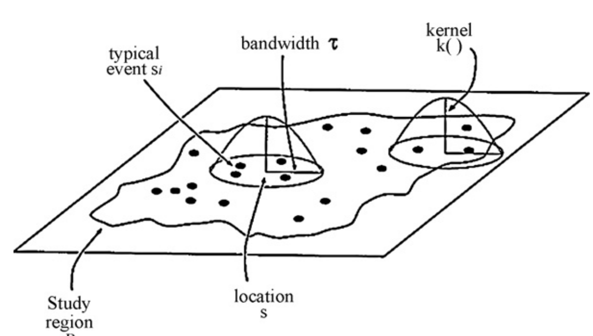

Fig. 1. Diagram of how the quadratic kernel density method works and is the basis for the density method used for this study (source: Bailey and Gatrell, 1995).

traffic accident data from 1999–2003 was used covering the geographical area of London, UK. This data is called Stats19 data and is the official statistics collected by the Department of Transport and Police on personal injury road accidents. The accuracy and credibility of the Stats19 collection process depends upon close co-operation between central government, local government and police forces. In terms of spatial accuracy, the location of the accident is recorded using a GPS by the attending police officer to a 10 m resolution.

The resulting surface is a selection of grid cells (2290) which have an extensively high density. These grid cells are scattered evenly across London and a large proportion are grouped together indicating differing sizes in high density hotspots. The two parameters which affect the outcome of the KDE are bandwidth (sometimes known as search radius) and cell size. Arguably the most important criterion for determining the most appropriate density surface is the bandwidth (Silverman, 1986; Bailey and Gatrell, 1995; Fotheringham et al., 2000). The choice of bandwidth will affect the outcome of the hotspots, for example the larger the bandwidth the larger the hotspots will be. The limited range of studies which have documented parameters for road accident density measurements means that the process of deciding the bandwidth and grid cell size is somewhat subjective. The final choice was based on taking a search radius which is two times the size of the grid cell, therefore for this study the bandwidth is 200 m and the grid cell size is 100 m.

4. Adding data and K-means clustering

The methods used in the previous section created a surface whereby the grid cells represent the hotspots based on the density measure. The number of grid cells in a hotspot varies, showing the hotspots are not uniform in size or shape. To be part of a hotspot a grid cell has to have an accident density level which is over a specified threshold, indicating these are the areas in London where the accident density prevalence is at its most intense. This grid surface provides the basis for collating the accidents which occur within these grid based hotspots. The result is a database of hotspots whereby the nature of the database means that there is no analysis of individual accidents, but an analysis of groups of accidents which share a common nearby spatial location, implying a common casual factor. By ascertaining the nature of this similarity, comparisons between hotspots can be made on a 'like by like' basis.

Fig. 2 shows the different spatial levels of road accidents. The accidents which share a spatial commonality within the grid cell are selected. These hotspots are then classified using the clustering process and are organised into classes (or clusters) based on similar attributes. These clusters are then organised into groups, based on the similarity of the clusters. This hierarchical process allows a classification of spatial hotspots based on similarity of either the characteristics of the accidents within the hotspots or of the environmental and land use information associated with the hotspot area itself or indeed of both.

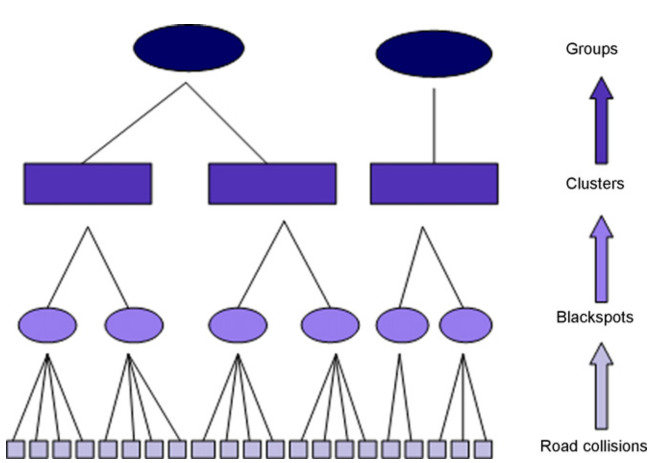

Fig. 2. How the hotspot classification method works.

Table 1
Environmental and land use data and associated sources.

Attribute	Source
Road length	Ordnance Survey Mastermap™
Cycle lane length	London Cycle Network
Pedestrian crossings	Transport for London
London underground stations	Transport for London
Traffic lights	Transport for London
Bus stops	Transport for London
Schools (primary and secondary)	Department for Education
Speed cameras	London Safety Camera Partnership

When determining the database used to build the classification it is important to assess the type of data which would be collected and would have the potential of having impact on accident density. Therefore it was important to consider not just the attributes of the accidents themselves but environmental and land use data which were found in the vicinity of the hotspots.

To select the hotspots, various rules were established to make the process simpler. Hotspots were established by the linking of cells. A hotspot could be made up of one single cell of many cells. The rule was that the cells had to join sides, and no diagonally linked cells would constitute being part of the hotspot. Overall a total of 428 hotspots were selected as being over a predetermined density threshold, Table 1 shows an example of attribute data and corresponding source.

The accident data added into the hotspot database was in count format, relating to the accidents occurring within the hotspot. For example, the severity of each accident (fatal, serious or slight) within the hotspot is added to the database. Each of these counts including the added non-Stats19 counts needed to be related to a corresponding base count. The main objective in creating a logical classification is to create an accurate representation of the data; therefore the data must have a suitable basis for comparison.

The data was normalised by one of two variables, the number of accidents in each hotspot or the number of grid cells that make up each hotspot. The reason for this, was to accommodate the two different types of data being used (Stats19 and environmental/road network). The Stats19 accident type counts were divided by the number of accidents within the hotspot, and the environmental/road network data were divided by the number of grid cells as these data were related to the area within the grid cell. The variables were also weighted for the purpose of the clustering methodology.

5. Clustering process

This procedure represents an attempt to classify each of the 428 hotspots into relatively homogenous types based on their environmental characteristics. The algorithm starts by defining K hotspots, one for each of the predefined clusters. These are selected on a random basis but proportional to the number of grid cells in the hotspot. The first stage is to find the most similar' pair of clusters, i.e. the pair that could be combined for the least incremental loss of variance of the original data; these are merged into a parent cluster and the process repeated until all the clusters have been joined together. The similarity between clusters is defined in terms of the distance between clusters weighted by the number of grid cells. The success of the clustering process is based of what is referred to as the percentage of variance which is explained by the clustering programme. In this instance the variance explained is 34%. Although this appears to be a low figure, previous studies indicate that this figure is an 'acceptable'. A study by Naveh and Marcus (2002) who tried to depict the causes of fatal road accidents and injury only road accidents found the variance for fatal accidents, 2% and for injury only accidents, 12%.

6. Results and discussion

The clustering process created 5 groups and 15 clusters (these numbers were predetermined). Fig. 3 shows the numbers of hotspots within each cluster and the number of clusters within each group.

The groups created from the clustering process vary in size considerable, with groups A1 and A4 containing the majority of the clusters. The major difference between the groups and the clusters within them is the variance of the constituent variables. The variance for the variables in each group is considerably lower, because of the hierarchical structure of the clustering results (Table 2). Therefore the description of the pen portraits is much less conclusive compared to the clusters. To give an indication of the characteristics of the hotspot clusters, two clusters are discussed in more depth; A5 Illicit late night Zone 1 pedestrians; C10 Cyclists in Westminster.

6.1. Cluster A5: Illicit late night Zone 1 pedestrians

This cluster consists of 48 of the total number of records (428). This cluster is characterised by having a higher than average number of accidents occurring between midnight and 3 a.m. and early on a Sunday morning. A high proportion of these occur at or near pedestrian crossings and bus stops resulting in serious injury. Table 3 demonstrates variables with the highest and lowest index scores.

The London boroughs with high proportions of A5 include Islington (5), Southwark (5), Camden (4), Haringey (4) and Kensington (4). This shows a strong Central London distribution, which supports the idea of pedestrians who have been drinking in Central London and then try to cross busy roads in order to get home or to a bus stop. There is a prominent Camden hotspot associated with high numbers of bars and clubs in this area (see Fig. 4).

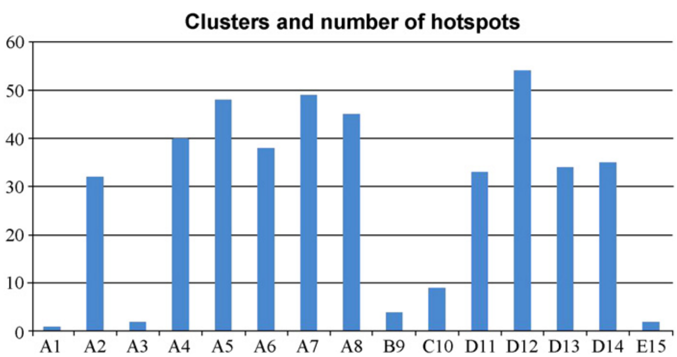

Fig. 3. Clusters and groups.

Table 2
Groups and associated clusters.

Groups	Clusters
Group A: Central London Pedestrians	A1: Late Night Football Supporters
	A2: Inner city pedestrian risk takers
	A3: Saturday morning leisure time hotspots
	A4: Sunday afternoon car drivers
	A5: Illicit late night Zone 1 pedestrians
	A6: Pedestrians in the dark
	A7: Weekday rush hour pedestrian hotspots
	A8: Morning commuting cyclists at rush hour
Group B: High density vehicle damage	B9: High density careless weekend drivers
Group C: Cyclists in danger	C10: Cyclists in Westminster
Group D: Multiple main road accidents	D11: Dual carriageway joy riders
	D12: Main road multiple victim accidents in Outer London
	D13: Sunday afternoon multiple casualties
	D14: Risk taking early risers
Group E: Weekend risk takers	E15: Sunday morning pedestrian risk takers

Table 3
Characteristics of Cluster Type A5.

High			Low		
Variable	Index	Variance	Variable	Index	Variance
Casualties = 6	190	4.72	Rain with wind	67	7.77
Pedestrians	161	62.11	Cyclist	66	75.69
Tube stations	138	4.03	Casualties = 3	65	71.81
Bus stops	121	5.65	Vehicles = 4	62	5.2
Vehicles = 1	121	57.42	Number of cells	30	12.57
Time 0000–0359	117	33.83	Number of accidents	28	83.02
Sunday	114	40.7	Vehicles = 5	14	82.83

6.2. Cluster C10 Cyclists in Westminster

The characteristics of cluster C10 are unique. It has 10 hotspots and the average number of grid cells per hotspot is 39.3 which is the highest out of all the clusters. This is due to the nature of the cluster covering a large proportion of Central London. This cluster, unlike all the others has a significantly high index score for both the number of accidents and number of cells. This indicates therefore that the hotspots which occur in this cluster have a high number of accidents and are spatially extensive. Many of these cluster 10 hotspots occur in Central London, particularly Westminster where there is a large traffic flow and the hotspots cover large areas (Fig. 5). The cluster also has a high number of accidents involving only one vehicle, likely to result in a fatal injury. However, what is missed from the index scores is the unusually low propensity for cyclists involved. From the index scores, cyclists are underrepresented within this cluster (Table 4): however the count data reveal that this cluster account for 37% of all cyclists' accidents. It also consists of nearly 17% of all the fatal accidents, which is significantly higher than average.

The cluster types have led to some interesting patterns across time and space, particularly the strong divide of clusters between those that involve pedestrians and cyclists and those that do not. Predominantly the cluster types involving pedestrians and cyclists occur in Central London while vehicle only cluster types are more likely to occur on the larger more arterial roads around Central London. The outcome of this methodology is a database of hotspots of varying size and density and associated collisions which fall

Table 4
Characteristics of Cluster Type C10.

High			Low		
Variable	Index	Variance	Variable	Index	Variance
Number accidents	524	83.02	Casualties = 4	11	6.95
Number cells	516	82.83	Rain with wind	5	7.77
Other weather	178	12.12	Vehicles = 4	4	5.2
Severity = Fatal	162	9.16	Snowing	0	3.5
Vehicles = 1	149	57.42	Vehicles = 5	0	12.57
Unknown	120	14.57	Casualties = 5	0	8.24
Tube stations	118	4.03	Casualties = 6	0	4.72

Fig. 4. A5 'Illicit late night Zone 1 pedestrians (Camden)'.

within the boundaries of the hotspot. The nature of the database means that there is no analysis of individual collisions, but an analysis of the set of the collisions which share a common nearby spatial location, implying a common and linking casual factor. Although hotspots may be treated as though each is unique, they may share similar characteristics such as the proportion of pedestrians or cyclists or an increased number of collisions in certain weather types or they may occur at a certain time of day or particular day of week. By ascertaining the nature of this similarity, comparisons between hotspots can be made on a 'like by like' basis.

Using kernel density estimation to investigate the spatial clustering of road accidents is not new. Previous work by Flahaut et al. (2003), Steenberghen et al. (2004) suggest that this is a successful spatial clustering method. However the study presented here suggests a innovative method adding value to the kernel density estimation method by using it not only to determine high density accident areas but to select sites for further investigation and append data to them. It is this last point which is of importance in the results and outcome of this paper. Previous empirical research has been limited to only the accident event itself. In a dense urban environment such as London, accident locations are based on proximity characteristics, examples might include accident concentrations near schools or underground station exits. This research is unique because it uses a spatial interpolation technique to differentiate the different factors that influence road accident rates from a social-spatial perspective. This is largely because of the unique nature of the urban setting and of the clusters identified. Studies can support individual findings of the clusters, for example the characteristics of cluster C10 indicated high numbers of accidents involving cyclists. This is generally not a new finding, with evidence from xxx suggesting similar findings in an urban setting. Cyclists are a vulnerable road user however; the nature of the cluster indicates other factors which are influencing the high density

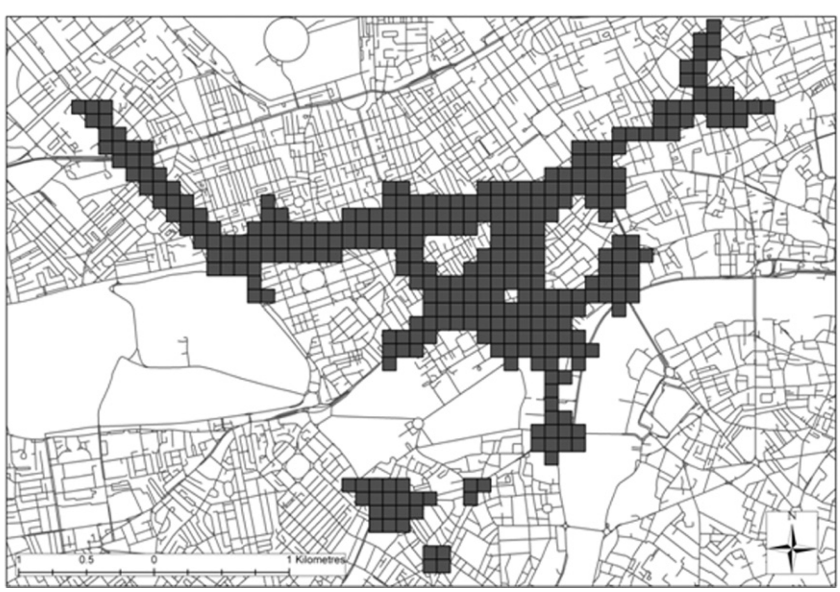

Fig. 5. C10 'Cyclists in Westminster'.

of road accidents. The very nature of these results provides further thought into the spatial interaction of different factors within these clusters.

7. Conclusion

This paper presents a methodology to identify high density accident hotspots and in turn create a clustering technique which determines casual indicators more likely to be present at certain clusters, therefore being able to compare like with like across time and space. The kernel density estimation tool enabled an overarching visualisation and manipulation of the accidents based on density which was used in turn to create the basic spatial unit for the hotspot clustering method. The classification of road accident hotspots in road safety still remains an important and yet under developed theme. These typologies provide a snapshot of the processes which are occurring at these sites and the people upon whom they impact. This information can lead road safety professionals to a better understanding, not only of the types of hotspots but their patterns across London. There are some evident potential policy implications for certain clusters. For example, C10 highlights the need for safety of cyclists in central London, whether this is the mandatory use of cycle helmets, better cycle lane provision or better cyclist/driver education. One of the most important recommendations which reflects the current local governmental policy is the focus on community and neighbourhood. By drilling down to specific accident clusters in specific areas, allows for a greater neighbourhood participation in understanding people' road user risk.

Some critiques of KDE challenge the fact that its treats discrete events as a continuous surface. However this paper is also concerned with the spread of risk, the risk of having an accident, geographically will occur not just at a single point but over a given area. KDE offers a method which takes into account this notion of spread of accident risk. However one main drawback reoccurs, which relates to determining the statistical significance of the resulting clusters. This is an area of research which is something to investigate in further studies. This study represents a large scale model for road accident clustering. Further study needs to be conducted in a number of areas. Firstly, there needs to be development of a method for testing for statistical significance of the kernel density output. Secondly, there needs to be investigation into the changing dynamics of the clusters over different temporal and spatial scales. Thirdly, a policy led investigation needs to be conducted into the suggestions made from the cluster outcomes.

Road accident analysis particularly the spatial patterns of road accidents requires further attention. This study aims to highlight some of the gaps in the research with particular attention to spatial clustering of road accidents. This is one of a handful of research papers which addresses the nature of supplementary data to examine the road accident hotspots. Traditionally research has relied on raw statistics alone without examining the potential indicators found in complimentary datasets such as those referring to the environment, land use, accident victims and road furniture for statistical clustering. This paper adds significant value to the research on the delineation of road accident hotspots and the complex nature of how we measure road accident hotspots.

Acknowledgements

The author was funded by ESRC CASE award number PTA-033-2002-00025 and gratefully acknowledges the support of The Metropolitan Police and ABM-Software. A special note of gratitude is also extended to Professor Richard Webber and Professor Paul Longley for their assistance and guidance.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modelling traffic accident occurrence and involvement. Accident Analysis and Prevention 32 (5), 633–642.
Bailey, T.C., Gatrell, A.C., 1995. Interactive Spatial Data Analysis. John Wiley and Sons, New York, NY.
Blower, D., Campbell, K., Green, P., 1993. Accident rates for heavy truck-tractors in Michigan. Accident Analysis and Prevention 25 (3), 307–321.
Ceder, A., Livneh, M., 1982. Relationships between road accidents and hourly traffic flow. Accident Analysis and Prevention 14 (1), 19–34.
Chainey, S., Ratcliffe, J., 2005. GIS and Crime Mapping. John Wiley and Sons, UK.
Department of Transport, 2000. Tomorrow's Roads: Safer for Everyone. Department of Transport, UK.
Flahaut, B., Mouchart, M., San Martin, E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accident Analysis and Prevention 35 (6), 991–1004.
Flahaut, B., 2004. Impact of infrastructure and local environment on road unsafety logistic modelling with spatial autocorrelation. Accident Analysis and Prevention 36 (6), 1055–1066.
Fotheringham, S., Brunsdon, C., Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Data Analysis. Sage, Thousand Oaks, CA.
Hauer, E., Persaud, B.N., 1987. How to estimate the safety of rail-highway grade crossing and the effects of warning devices. Transportation Research Record 1114, 131–140.
Hauer, E., 1997. Observational before-after studies in road safety. Pergamon, Oxford.
Levine, N., Kim, K., Nitz, L., 1995. Daily fluctuations in Honolulu motor vehicle accidents. Accident Analysis and Prevention 27 (6), 785–796.
Longley, et al., 2005. Geographic Information Systems and Science. John Wiley and Sons, Chichester.
Maher, M., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accident Analysis and Prevention 28 (3), 281–296.
Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention 26 (4), 471–482.
Naveh, E., Marcus, A., 2002. A Re-Assessment of Road Accident Data-Analysis Policy. Center for Transportation Studies, University of Minnesota.
Oppe, S., 1979. The use of multiplicative models for analysis of road safety data. Accident Analysis and Prevention 11, 101–115.
Oxley, J., Charlton, J., Fildes, B., Koppel, S., Scully, J., Congiu. M., Moore, K., 2005. Crash Risk of Older Female Drivers, Report no. 245, Monash University Accident Research Centre.
Pietro, P., 2001. High Risk Road Users. NRMA ACT Road Safety Trust, Australia.
Pulugurtha, S., Vanjeeswaren, K., Nambisan, S., 2003. Development of Criteria to Identify Pedestrian High Crash Locations in Nevada, Quarterly Progress Report, University of Nevada, USA.
Sabel, C., 2006. Kernel Density Estimation as a Spatial-Temporal Data Mining Tool: Exploring Road Traffic Accident Trends, GISRUK 2006, University of Nottingham.
Silverman, B., 1986. Density Estimation for Statistics and Data Analysis, 1st ed. Chapman and Hall, London.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention 27 (3), 371–389.
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), 169–181.
Thomas, I., 1996. Spatial data aggregation: exploratory analysis of road accidents. Accident Analysis and Prevention 28 (2), 251–264.
Whitelegg, J., 1987. A geography of road traffic accidents. Transactions of the Institute of British Geographers. 12, 161–176.
World Health Organisation, 2004. World Report on road traffic injury prevention, World Health Day Publication.