Comparison of spatial methods for measuring road accident ‘hotspots’: a case study of London

TESSA ANDERSON

University of Canterbury, Department of Geography, Private Bag 4800, Christchurch 8020, New Zealand; tessa.anderson@canterbury.ac.nz

(Received 11th August 2006; Accepted 22nd November 2006)

Abstract: There is a continuing determination by academics and road professionals alike to investigate the most appropriate methods for identifying road accident hotspots particularly in urban areas. Increasingly this research has involved the use of GIS and spatial analysis in order to define both visually and statistically what can be defined as a road accident hotspot. Traditional methods of hotspot detection by road professionals have included comparing count data at different locations and rating the areas by severity. However the increase use of GIS has lead to academics using sophisticated methods to quantify hotspots. There is, however, no universal definition of a road accident hotspot which means that the definition of a hotspot is open to much speculation. This paper seeks to investigate the merits of three different spatial techniques for quantifying road accident hotspots. Kernel density estimation, network analysis and area wide analysis are used to demonstrate three methods. The methods are then reviewed. There is however an exhaustive list of hotspot detection techniques, not all of which can be outlined in this paper.
1. Introduction

Road accident hotspot analysis has traditionally centred on road segments or specific junctions (Thomas 1996, Cook et al 2001) while area wide hotspots are often neglected. The road accident literature provides no universally accepted definition of a road accident ‘hotspot’ (Brimicombe 2004). Although Hauer (1996) describes how some researchers rank locations according to accident rate (accidents per driven vehicle kilometre), while other researchers use accident frequencies (accident per road kilometre). Since the 1970s there has been a plethora of statistical models applied to the understanding of road accidents. However these models have had a tendency to neglect the spatial patterns of road accidents.

This paper compares three alternative spatial statistical methods for distinguishing road accident hotspots:

- Kernel density estimation
- Network analysis
- Census Output Area estimation

There are a wide range of approaches taken by researchers which are often limited to complex statistical modelling to identify accident hotspots. Only recently has GIS made a noticeable impact in road accident hotspot research and it brings with it a wide range of sophisticated spatial-statistical techniques to increase the accuracy and information of road accident hotspots. Road accident hotspots are often determined by a range of methods, however there is no universal method, which is why continuous comparison of methods is vital if road accident hotspots are to be determined in the most successful way.

The conceptual basis of this paper lies in the premise of the spread of risk that a collision generates over an area. In many circumstances, it has been acceptable to define a cluster of collision points in an area or on a line, but the risk of a collision occurring again will likely spread beyond the boundaries of the historical collision cluster. If the cluster of collisions is small and spiky the risk surrounding the cluster will be smaller, however if the collision cluster is quite flat but covers a wider area then the risk
surrounding the cluster will be larger but less intense than the cluster with the collisions closer together (Chainey and Ratcliffe 2005).

Considering road accidents as points on a map is not robust enough to distinguish areas of high numbers of accidents and implement effective policy decisions. This can be attributed to three reasons. Firstly, there are potential inaccuracies in the collection of road collision locational points (Austin 1995ab). Secondly, in a highly urban area such as London where there are many collisions it can be difficult to visually identify areas of high numbers of road traffic accidents. Thirdly, by just considering points on a map, there is a disregard for spatial or statistical patterns and it relies on personal and potentially bias interpretation. In recent years there have been many methods of measuring road accident hotspots that go beyond points on a map. Thomas (1996) for example focused on road accidents and network segments, specifically the most appropriate length. Whereas Sabel et al (2005) has used a surface based modelling approach for the quantification of accident hotspots using the interpolation technique, kernel density smoothing. Finally, Transport for London and other government institutions use aggregated accident data usually to boroughs, whereby rates of accidents are used to illustrate the differences between London boroughs.

2. Methods

Road accidents are inherently constrained by the road network and a considerable amount of research has been conducted into network spatial autocorrelation, relating to optimum length of road network to study and traffic flow (see Thomas 1996, Flahaut 2004, Geurts et al 2005). In other analyses, particular within boroughs road accidents have been linked to census information with regards to measuring socio-economic and areal characteristics of areas in relation to casualty home location. The advantages of using census areas involve the management of road safety programmes which are implemented at a borough level. Often boroughs are disaggregated into smaller units with which to provide solutions to road accident hotspots. Kernel density estimation is an interpolation technique, which is a method for generalising incident locations to an entire area. In short, interpolation techniques generalise the collisions over the study region.
A road collision data set dating from January 1998 to June 2002, supplied by the London Metropolitan Police, for a study area in North London was used (Figure 1). It consists of approximately 189,000 collisions, resulting in injury, and is georeferenced to a 10 metre resolution. Using OS MasterMap Integrated Transport Layer, collision point data were counted for each of the 374,000 road segment polylines for London. Collision counts for each Census 2001 Output Area in London were counted using ESRI’s ArcGIS.

The first map (Figure 2) shows an area of road network in North London, and associated accidents. In some cases this is determined as accident per road kilometre but in the instance the road segment is not specified, instead using the natural road length (consisting of a total of 374,000 polylines). It was compiled by spatially attaching road accidents to road segments. In this example, junctions were not included in the study. The second map (Figure 3) shows the aggregated measurements of accident rates for both Census Output Areas and boroughs in London. This map was compiled by appending road accidents to Census Output Areas and then normalised. The final map (Figure 4) uses kernel density estimation to create a density surface of the accidents by using the density function in ArcGIS Spatial Analyst. For this map the bandwidth was 200 metres and the cell size was 100 metres.

In order to concentrate on small areas within London and test how geographical constraints affect the locations of road accident hotspots, two boundaries have been chosen. The two alternative methods presented here represent two very different ways to analyse accident density with regards to boundary constraints. Both methods have a boundary constraint in the form of the London road network and Census output area boundaries.
Figure 1 Study area in North London.

Figure 2 A network analysis of road traffic collisions.
Figure 3 Road traffic collisions by 2001 census output areas.

Figure 4 Kernel density estimation of road traffic collisions.
3. Discussion

The use of density surfaces have been widely used in crime analysis (Brimicombe 2004, Chainey and Radcliffe 2005) and are slowly being adopted by road accident analysts (Flahaut et al 2003, Sabel et al 2005). Within this context it has been useful for crime analysis practitioners to apply surface interpolation to point crime data in order to determine hotspots over a given spatial area (Chainey and Ratcliffe 2005).

Chainey and Ratcliffe (2005) give a number of examples of using kernel density estimation to determine crime hotspots both in the UK and USA. The main difference between crime and road accident hotspots is that road accidents are constrained to the road network and crime hotspots occur anywhere, whether it is on a street or in a house, or park. Therefore the applicability of kernel density estimation to crime hotspots is somewhat more successful because of the lack of geographical constraint on the point dataset.

Kernel estimation is able to quickly and visually identify hotspots from large datasets and therefore provide a statistical and aesthetically satisfactory outcome. The advantages of these surface representations particularly of road accidents are that they can provide a more realistic continuous model of accident hotspot patterns reflecting the changes in density which are often difficult to represent using geographically constrained boundary based models such as the transport network or census tracts. Over the years there have been a number of spatial tools developed which help in the understanding of the changing geographies of point patterns. The most promising of these tools is kernel density estimation (Sabel et al 2005). There are many advantages to the use of kernel density estimation (KDE) as opposed to statistical hotspot and clustering techniques such as K-means. The main advantage for this particular method lies in determining the spread of risk of a accident, outlined earlier. In defining a cluster one is overlooking the buffer around it which will ultimately possess a degree of risk of accident for the people who enter it. This degree of risk would not be measured using the clustering techniques.
4. Conclusions

The use of kernel density estimation for determining road accident hotspots is dependant on the intended outcome. If the intention is to use only road accident data for the analysis of hotspots, then other methods may be more suitable. However kernel density estimation lends itself to the integration of supplementary datasets regarding the road environment and people involved in the road accident(s). This is because of the nature of a kernel density estimation being based on a fixed ‘cell’, whereby supplementary data can be allocated into each cell.

The advantages of the surface representations particularly of road collisions are that they can provide a more realistic continuous model of collision hotspot patterns reflecting the changes in density which are often difficult to represent using geographically constrained boundary based models such as the transport network or census tracts.

The purpose of this comparison is to determine not only the different spatial methods associated with defining road accident hotspots but evaluate their advantages and disadvantages. It is important to understand the contrasts between the three methods for a more robust understanding of the possibilities within road safety analysis and provide and unique research strand which has yet to be addressed. The key areas of focus here are accuracy and communication. In short defining what are the most accurate maps for the best visual communication for decision makers in road safety.

Software

The accompanying maps were compiled using ESRI’s ArcGIS 9.1 software including Spatial Analyst and Hawth’s Analysis Tools for ArcGIS.
References


