

Road Traffic Accident Simulation Modelling - A Kernel Estimation Approach

Dr Clive E Sabel¹, Dr Simon Kingham¹, Prof Alan Nicholson², Phil Bartie¹

¹Department of Geography
University of Canterbury, Christchurch, New Zealand
Phone: +64 3 364 2987 Fax: +64 3 364 2907
Email: clive.sabel@canterbury.ac.nz

²Department of Civil Engineering
University of Canterbury, Christchurch, New Zealand
Phone: +64 3 364 2233 Fax: +64 3 364 2758
Email: alan.nicholson@canterbury.ac.nz

Presented at SIRC 2005 – The 17th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
November 24th-25th 2005

ABSTRACT

This paper outlines the development of a method for using Kernel Estimation cluster analysis techniques to automatically identify road traffic accident ‘black spots’ and ‘black areas’. Christchurch, New Zealand, was selected as the study area and data from the LTNZ crash database used to trial the technique. A GIS and Python scripting was used to implement the solution, combining spatial data for average traffic flows with the recorded accident locations. Kernel Estimation was able to quickly identify the accident clusters, and when used in conjunction with Monte Carlo simulation techniques, was able to identify statistically significant clusters.

Keywords and phrases: kernel estimation, spatial clusters, road traffic accidents, Monte Carlo simulation, spatial data mining.

1.0 INTRODUCTION

Improvements in car safety, road engineering, and driver education have brought about a decrease in the number of car accidents in Christchurch in the last decade. Guidelines exist for targeting ‘black’ sites, routes, and areas (Institute of Highways and Transportation, 1986) with there being a natural progression from site plans, to route plans, to area plans (Nicholson, 1990).

The digital era has brought with it a wealth of data stored in computer systems, pertaining to easy retrieval. Incident reporting systems are increasingly being recognised as offering analysts the capability to data mine for trends which may be subtly related, in the hope that reoccurrence of incidents can be reduced (Cassidy et al 2003). Current practices in the spatial analysis of road traffic accidents mostly rely on a visual examination, and are therefore highly subjective depending upon the observer (Cressie, 1991)

Road traffic accidents are complicated to analyse as they cross the boundaries of engineering, geography, and human behaviour. It is not desirable to reduce the dimensionality of data through aggregation as statistical trends can be obscured, or even reversed, a problem formally known as Simpson’s paradox (Tunaru, 2001). It is also possible that attempting to join too many disparate datasets may introduce errors (Peled, 1993).

There is therefore a need for a more systematic approach which can automatically detect statistically significant spatial accident clusters, offering repeatable results by removing subjectivity. This research paper outlines the development of ongoing research into how spatial techniques can be used to fulfil this requirement.

2.0 RESEARCH METHODOLOGY

Vehicle crash records hold both temporal and spatial trends. For this research we have concentrated on the spatial element, examining the global trends and applying the knowledge at local level to find anomalies. Whilst human error and mechanical failure can be causes of road traffic accidents the importance of spatial factors has been ‘grossly underestimated’ (Whitelegg, 1987).

2.1 Study Area

Christchurch city was chosen as the study area, where both crash data and average traffic flow data were available to the researchers. The data recorded in the Land Transport New Zealand (LTNZ) accident system included all 28,645 reported minor, serious, and fatal accidents in Christchurch from 1980 to 2004. For this research a subset of approximately 3061 accidents from 2000 to 2004 was used to coincide with the time period for which high quality road flow data was available, and to reduce effects from any road layout changes during the study period.

2.2 Cluster Detection

In this exploratory, data-mining study, we adopted an Exploratory Spatial Data Analysis (ESDA) approach. ESDA can be broadly defined as the collection of techniques to describe and visualise spatial distributions, identify atypical locations, or spatial outliers, and discover patterns of spatial association, clusters or hotspots and suggest spatial regimes or other forms of spatial heterogeneity (Anselin, 1999) with a view to develop hypotheses.

A number of spatial tools have been developed recently that help in understanding the geography, and changing geographies, of point-patterns. For our purposes the most promising of these is Kernel Estimation (KE), whereby a distribution of discrete point ‘events’ is transformed into a continuous raster surface (Sabel et al. 2000) (Figure 1).

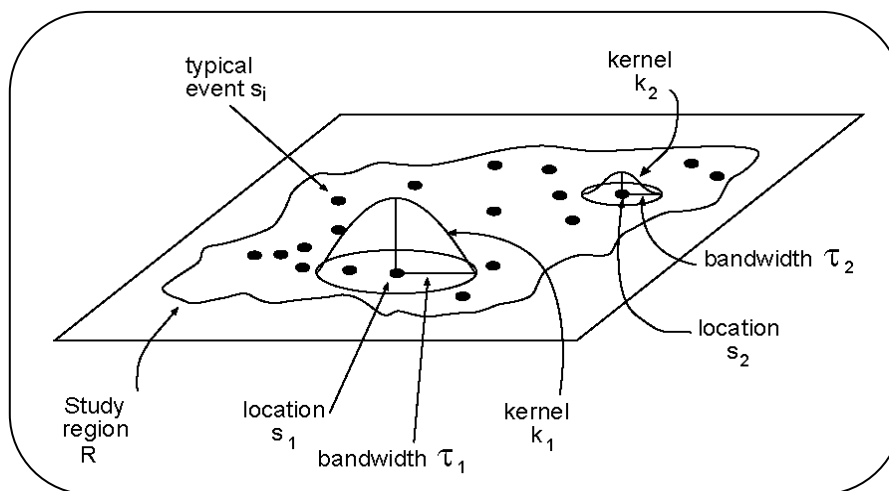


Figure 1: Principles of Kernel Estimation

2.3 Simulated Risk Input Surface – Model Input

Point analysis using KE is integrated functionality in modern GIS packages, such as ESRI’s ArcGIS, however determining the statistical significance is not. We coded this using Python scripting.

The risk of an accident occurrence is not equal across all road sections, with many sections having no recorded accidents within the last 25 years. Junctions are known to be accident ‘black spots’ within city regions,

accounting for 66% of all accidents in our dataset. Another key contributing factor for the relative accident risk is the road segment daily traffic flow.

Mapping the actual accident data on the modelled average daily traffic volume dataset makes it possible to record the corresponding estimate road flow for the location where the accident occurred. By grouping the road flow values to nearest 100 per day and summarising the number of occurrences globally it is possible to plot a graph of 'expected' crash frequency for a given road flow (Figure 2). The trend is for high volume motorways and low volume minor roads to have fewer crashes than the mixed usage medium flow volume urban arterial roads.

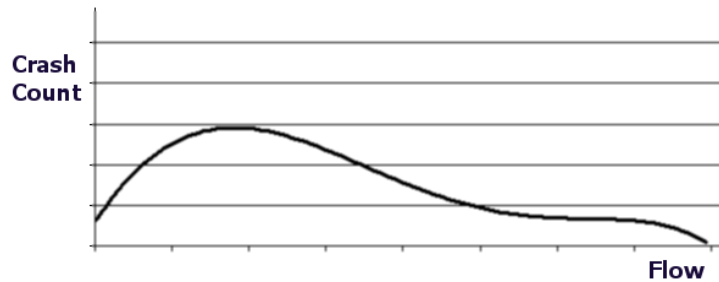


Figure 2: Number of Actual Crash Events for Estimated Road Flow

Using this global relationship it is possible to reclassify each road segment with an expected crash count. This does however assume an even distribution of accident risk along the length of the road section, whilst in reality there are many other spatial factors which can increase accident numbers in certain places.

Given their significance, junction data was introduced into the model. The distance from each crash to the nearest junction was calculated, establishing a global relationship which was then used to assign a crash risk rating to each square metre of road. By combining this with that of the flow data, an overall indication of the hazard level for each road section was established.

2.4 Monte Carlo Simulation

Monte Carlo simulation was employed to establish the statistical significance of clusters found using KE. Conceptually, we modified the methodology of Kelsall and Diggle (1995) to generate new synthetic data in a random allocation manner. If this process is then repeated m times, in a form of Monte-Carlo simulation, upper and lower simulation envelopes can be established and thereby an estimate of how unusual the observed pattern is obtained. If the observed pattern lies outside the simulation envelope, one can begin to speak of areas of significantly elevated, or reduced, risk.

The process involved generating the same number of random crash points as existed in the observed dataset, selecting random locations with a bias towards more hazardous areas as described by the Simulated Input Risk Surface. A KE surface was calculated from the random points and compared to that of the actual data, with areas of higher 'observed' than 'expected' being logged. This process was repeated 99,999 times, so that a cumulative log of those cells constantly scoring higher accident cluster rates than expected could later be examined.

The results of the simulations can be graphically displayed by constructing a 'p-value surface' which, for each \mathbf{x} (location), gives the proportion of values $\hat{s}_I(\mathbf{x})$, $I = 1, \dots, m$ which are less than the original estimate $\hat{s}_\theta(\mathbf{x})$. That is, it gives the proportion of simulated cells which are less than each observed cell, for each grid cell in the matrix. The 2.5% and 97.5% contours of this surface are effectively tolerance and they can then be draped over the original image of estimated risk, to highlight regions which correspond to significantly high or low risk.

3.0 RESULTS

Kernel Estimation is able to quickly visually identify 'clusters' from large datasets, and with the introduction of the Monte Carlo simulation it is possible to extract those clusters which are statistically significant. Our interim results demonstrated in figure 3 highlight some major intersections as well a general raised risk in the CBD of Christchurch.

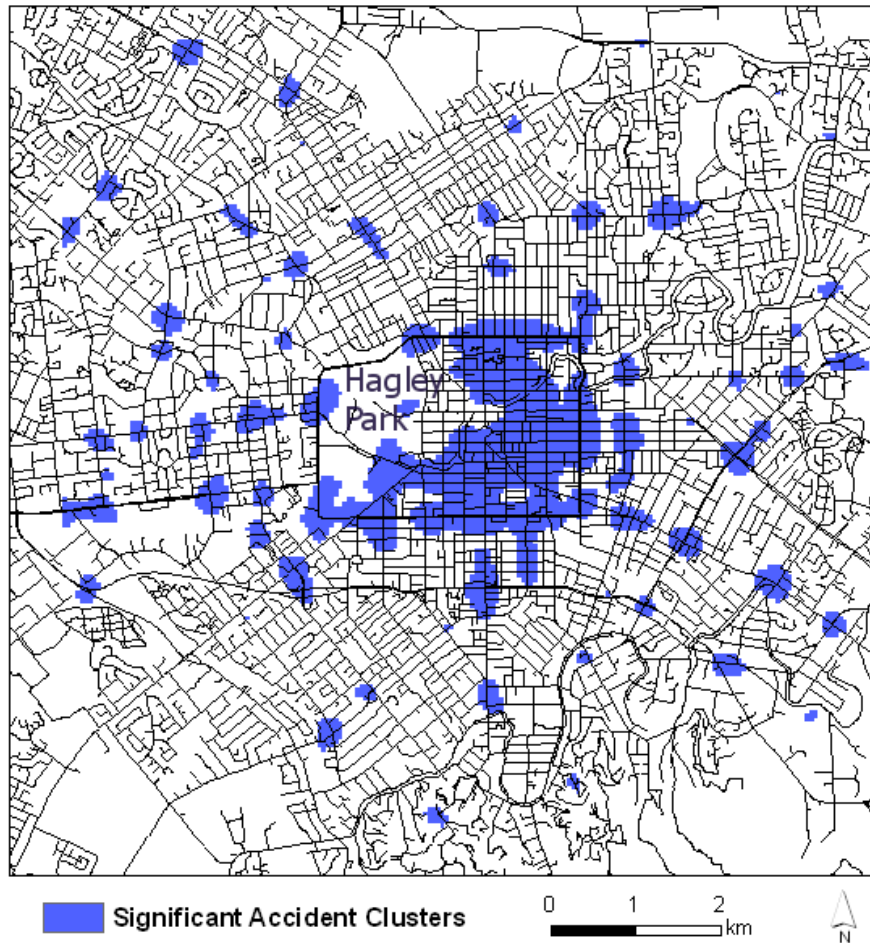


Figure 3: Significant Accident Clusters (top 2.5%) where Observed is higher than Expected - Christchurch, NZ.

4.0 DISCUSSION

Road traffic analysis is a very complex topic; the Police have almost 600 different cause codes which can be assigned to explain the origins of the accident in their report. It would be unrealistic to believe that our two input variables (flow and distance to junction) for a Simulated Input Risk Surface would form a complete answer to where expected accidents should be placed in the Monte Carlo simulation. During the life cycle of this research other factors have been tested, such as housing density, proximity to schools, supermarkets, and junction density. Balancing the relative importance of these variables is part of ongoing work.

Our results tend to focus on the statistically significant accident points and areas around junctions, which may be a side-effect of KE, being more suited to locating area clusters, than locating linear clusters ('black' routes). Adjustment to the KE bandwidth allows it to be adapted for use in locating 'black' spots to 'black' zones.

An issue for consideration when applying KE to road traffic accident data is that unrelated accidents on neighbouring roads can contribute to create apparent clusters (Steenbergen, T, *et al*, 2004) if the spatial bandwidth chosen is too large.

A future stage of this research might focus on those crashes which are considered to be of a spatial cause based on the traffic accident report, reducing the influence of randomness from human error (eg selecting wrong gear) or mechanical failure.

The risk of road travel changes at different temporal scales (time of day, season, school term time) and the modelling carried out in this research assumes a constant exists for the entire dataset. Despite this, the initial results are promising in terms of identifying areas which can become focal points for more detailed study, in this

respect this may form an automatic filtering stage for large accident crash datasets. These identified areas are already attracting interest from public authorities.

5.0 CONCLUSIONS

This research is ongoing and further refinements to the Simulated Input Risk Surface are being tested through an iterative design cycle. It has to be recognised that GIS will not be able to explain or predict all accidents, and that this methodology is only intended to draw attention to certain city areas which appear to be diverging from the general trends. The interim results have been demonstrated at a number of meetings with traffic engineers, who believe it will be a useful tool supplementing current techniques in the analysis of vehicle accident data.

ACKNOWLEDGEMENTS

The authors would like to thank LTNZ for permitting access to their data for this research.

REFERENCES

- Anselin, L., 1999, Interactive techniques and exploratory spatial data analysis. Geographical Information Systems: principles, techniques, management and applications. P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York, John Wiley: pp 253-266.
- Cassidy D, Carthy J, Drummond A, Dunning J, and Sheppard J, 2003, The Use Of Data Mining In The Design And Implementation Of An Incident Report Retrieval System, in, *Proceedings of the 2003 IEEE Systems and Information Engineering Design Symposium*, 24-25 April 2003 University of Virginia, Charlottesville, USA.
- Cressie, N.A.C., 1991, Statistics for Spatial Data. Wiley, New York.
- Institution of Highways and Transportation, 1986. Highway Safety: Guidelines for Accident Reduction and Prevention, 2nd Edition. I.H.T., London.
- Kelsall JE and Diggle PJ (1995) Kernel Estimation of Relative Risk, *Bernoulli*, 1, 3-16.
- Nicholson, A.J., 1990. Measures of accident clustering. In, *Transportation and Traffic Theory*, Elsevier, New York.
- Peled, A., Hakkert, A.S., 1993. A PC-oriented GIS application for road safety analysis and management. *Traffic Engineering and Control* 34 (7/8), 355-361.
- Sabel, C. E., A. C. Gatrell, et al. (2000). Modelling exposure opportunities: estimating relative risk for Motor Neurone Disease in Finland. *Social Science & Medicine* 50 (7&8): 1121-1137.
- Steenberghen.T, Dufays.T, Thomas.I and Flahaut.B, 2004, Intra-urban location and clustering of road accidents using GIS: a Belgian example, *IJGIS Volume.18*, pp. 169-18.
- Tunaru,R. 2001 Models of Association versus causal models for contingency tables, *The Statistician* vol 50, part 3 pp 257-269.
- Whitelegg, J, 1987 The geography of road accidents, *Transactions for the Institute of British Geographers*, Vol. 12.