Spatial Statistical Modelling of Traffic Accidents

Yilin Liu[1] and David Jarrett[2]

[1]Transport Management Research Group
Middlesex University, London NW4 4BT
Email: y.liu@mdx.ac.uk

[2] Transport Management Research Group
Middlesex University, London NW4 4BT
Email:d.jarrett@mdx.ac.uk

1. Introduction

Traffic accidents cause injury and death, and result in a serious economic burden. In Great Britain in 2005, there were 3,201 road accident deaths and 198,735 reported injured accidents (Department for Transport 2006). In order to understand the causes of road accidents and achieve safer roads, sophisticated statistical methods are needed. Accident data are spatial data, and include locational information such as a geographical grid reference. This type of spatial information enables different kinds of spatial analysis of accident data, some of which use a geographical information system (GIS) as a tool for data preparation, geographical visualization and cluster identification (Schneider et al. 2004). Recent successes of using a conditional autoregressive (CAR) prior (Mollié 1996) to take account of the spatial correlation in accident models at the local authority level (by Miaou et. all 2003, Liu and Jarrett 2007) and previous findings of the existence of spatial correlation in accidents on a road network (for instance, Maher 1987, Loveday and Jarrett 1992) suggest the development of spatial models for accidents on a road network. This study aims to examine the spatial correlation in accident data for a road network and the M1 motorway in England, and to develop spatial models for accidents on junctions and links.

2. Data description and software used

Two sets of data are used in this study. The first includes accidents that occurred at major junctions in Coventry in a 5-year period. Other variables include the junction type, number of arms, and junction control detail. This set of data was obtained from the SPECTRUM accident database (Mott MacDonald Ltd). Another set of data consists of STATS19 (see Department for Transport 2006) accident data for the M1 in England from 1999 to 2005. Accidents are aggregated for each road link between adjacent motorway junctions. Other variables include the traffic flow data obtained from an on-line traffic database of DfT and the link length (Ordnance Survey). Road map data were obtained from SPECTRUM (Mott MacDonald Ltd) and Digimap, both of which are based on Ordnance Survey data.

In this study, accident data are aggregated and integrated with other data in ArcView-3. Two R packages relevant to geographical applications are used to measure the spatial correlation in the data and create graphical plots. They are ‘spdep’ (Bivand 2004) and ‘maptools’ (Lewin-Koh and Bivand 2004). Models are fitted in WinBUGS (Spiegelhalter et al. 2003), within which GeoBUGS (Thomas et al. 2004) is an add-on.

3. Spatial correlation in road accident data

For both of the data sets, spatial correlation in the accident data is measured by Moran's *I* (Upton and Fingleton 1985). This requires defining which sites are neighbours.

Fig. 1 is a node-link graph displaying the neighbours structure of junctions on A- and B- roads in Coventry. Each node represents a junction. 55 junctions are identified. A link between two nodes means there is a road joining the nodes. Nodes at each end of a link are identified as neighbours, but the extent of spatial dependence is assumed to depend on the length of the link. Moran's *I* for junction accidents is marginally significant. It has a value of 0.173 with a p-value of 0.06.

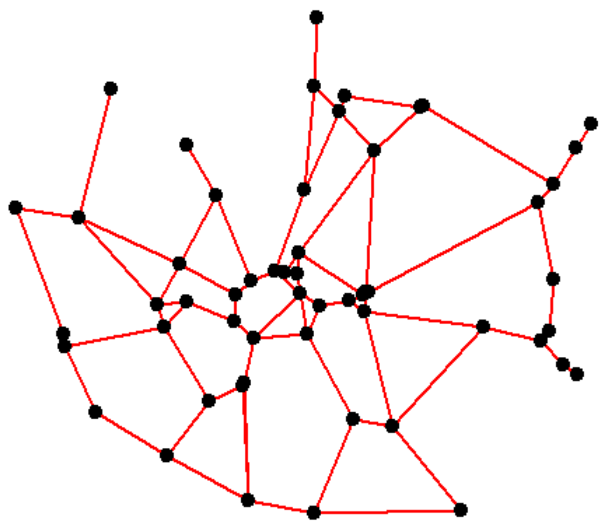

Figure 1 neighbours structure of major junctions in Coventry.

The M1 extends approximately in a north-south direction. Fig. 2 plots the locations of accidents on the M1 in 2005. This plot illustrates the road layout of M1. It is relatively easy to identify the structure of neighbours here. For accidents on road links, adjacent links are defined as links that join in a common node (here a motorway junction). Therefore, most links on the M1 have just two neighbours. High spatial correlation is found for road link accidents here. Moran's *I* has a value of 0.64. This suggests spatial correlation needs to be considered in models for accident on road links.

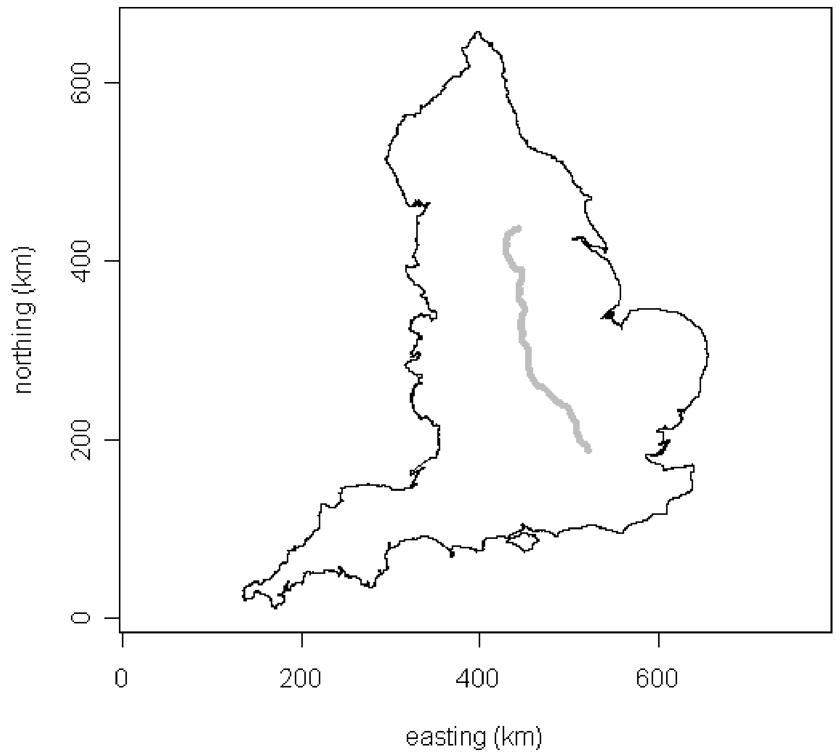

Figure 2. Plot of locations of accidents on the M1 in England.

4. Spatial models for accidents on a road network

Spatial models are developed for accidents at major junctions in Coventry and accidents on road links of the M1 in England respectively. In addition, temporal effects are considered for the M1 data. In the framework of Bayesian data analysis (Gelman et al. 2003), Poisson models, Poisson models with log-normal random effects and Poisson models with spatially structured random effects are developed. The last use a CAR prior to take account of the spatial correlation in the accident data.

Suppose the number of road accidents at site (either a junction or a road link) i during a fixed period t is Y_{it}. A general form of model, which includes both the spatial effects and unstructured random effects, can be written as:

$$Y_{it} \sim \mathrm{Pois}(\lambda_{it})$$
$$\log \lambda_{it} = \sum_j \beta_j x_{ijt} + \delta_t + v_{it} + \theta_{it}.$$

Conditional on the underlying mean accident frequency λ_{it}, the observed number of accidents at site i is assumed to be independently Poisson distributed. The underlying mean λ_{it} is further modelled based on some relevant explanatory variables (for instance, traffic volume and road length), denoted by x_{ijt}, a trend variable δ_t and other random effects: the ν_{it} captures unstructured random effects and are assumed to be independently normally distributed with mean zero and variance $\sigma^2_{\nu t}$, while θ_{it} captures spatially structured random effects via a CAR prior formulated by

$$\theta_{it} \mid \theta_{jt}, j \neq i \sim N\left(\sum_{j \in N[i]} \frac{w_{ij}}{w_{i+}} \theta_{jt}, \ \frac{\tau_t}{w_{i+}} \right)$$

where $w_{i+} = \sum_{j=1}^{N} w_{ij}$ and τ_t is a scale parameter; $j \in N[i]$ means *j* is a neighbour of *i*. ν_{it} and θ_{it} can be either fixed in time or vary in different years.

For junction accidents and accidents on road links, the structure of neighbours in the CAR prior is determined by the road network. General rules to identify neighbours for a junction and a road link have been explained in the previous section. For the weight matrix W, the simplest choice is $w_{ij} = 1$ if sites *i* and *j* are neighbours, and $w_{ij} = 0$ otherwise. For junction models, W can also be determined by the total road length between two neighbouring junctions -- the shorter the road length, the larger the weight.

6. Acknowledgements

The authors are extremely grateful to Jacqui Bates and Sonal Ahuja of Mott MacDonald Ltd, Anne Patrick of the Ordnance Survey, and Solihull MBC and Coventry City Council for providing access to the West Midlands SPECTRUM database.

7. References

Bivand R, 2004, Spdep: spatial dependence: weighting schemes, statistics and models. R package version 0.2-22. http://cran.r-project.org/src/contrib/Descriptions/spdep.html.

Department for Transport, 2006, Road Casualties Great Britain: 2005. Stationary Office, London.

Gelman A, Carlin J, Stern H and Rubin D, 2003, Bayesian Data Analysis. Chapman & Hall, London.

Lewin-Koh N and Bivand R, 2004, Maptools: tools for reading and handling shapefiles. R package version 0.4-7. http://cran.r-project.org/src/contrib/Descriptions/maptools.html.

Liu Yilin and Jarrett D, 2007, Spatial Bayesian modelling of road accident at the local authority level. In: Mathematics in Transport, proceedings of the Fourth IMA international conference on Mathematics in Transport. Elsevier. (to appear)

Loveday J and Jarrett D, 1992, Spatial modelling of road accident data. In: Mathematics in Transport Planning and Control (J. D. Griffiths, ed.), pp. 433–446. Oxford University Press.

Maher M, 1987, Accident migration: a statistical explanation. *Traffic Engineering and Control*, 28: 480–483.

Miaou S, Song J J and Mallick B K, 2003, Roadway traffic crash mapping: a space-time modelling approach. *Journal of Transportation and Statistics*, 6(1): 33–57. http://www.bts.gov/publications.

Mollié A, 1996, Bayesian mapping of disease. In Markov Chain Monte Carlo in Practice (W R Gilks, S Richardson and D J Spiegelhalter, eds.), pp. 359–379. Chapman and Hall, London.

R Development Core Team, 2007, R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

Schneider R J, Ryznar R M and Khattak A J, 2004, An accident waiting to happen: a spatial approach to proactive pedestrian planning. *Accid. Anal. and Prev*, 36: 193-211.

Spiegelhalter D, Thomas A, Best N and Lunn D, 2003, WinBUGS Version User Manual, MRC Biostatistics Unit. Version 1.4. http://www.mrc-bsu.cam.ac.uk/bugs.

Thomas A, Best N, Arnold R and Spiegelhalter D, 2004, GeoBUGS User Manual, MRC Biostatistics Unit. Version 1.2. http://www.mrc-bsu.cam.ac.uk/bugs.

Upton G and Fingleton B, 1985, Spatial Data Analysis by Example, Vol. I. Wiley, Chichester.