

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>



Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data

Mohammed A. Quddus*

Transport Studies Group, Department of Civil and Building Engineering, Loughborough University, Leicestershire LE11 3TU, United Kingdom

ARTICLE INFO

Article history:

Received 30 October 2007

Received in revised form 20 March 2008

Accepted 24 March 2008

Keywords:

Spatial correlation

Uncorrelated heterogeneity

Negative binomial models

Classical spatial models

Bayesian hierarchical models

ABSTRACT

Count models such as negative binomial (NB) regression models are normally employed to establish a relationship between area-wide traffic crashes and the contributing factors. Since crash data are collected with reference to location measured as points in space, spatial dependence exists among the area-level crash observations. Although NB models can take account of the effect of unobserved heterogeneity (due to omitted variables in the model) among neighbourhoods, such models may not account for spatial correlation areas. It is then essential to adopt an econometric model that takes account of both spatial dependence and uncorrelated heterogeneity simultaneously among neighbouring units. In studying the spatial pattern of traffic crashes, two types of spatial models may be employed: (i) classical spatial models for higher levels of spatial aggregation such as states, counties, etc. and (ii) Bayesian hierarchical models for all spatial units, especially for smaller scale area-aggregations. Therefore, the primary objectives of this paper is to develop a series of relationships between area-wide different traffic casualties and the contributing factors associated with ward characteristics using both non-spatial models (such as NB models) and spatial models and to identify the similarities and differences among these relationships. The spatial units of the analysis are the 633 census wards from the Greater London metropolitan area. Ward-level casualty data are disaggregated by severity of the casualty (such as fatalities, serious injuries, and slight injuries) and by severity of the casualty related to various road users.

The analysis implies that different ward-level factors affect traffic casualties differently. The results also suggest that Bayesian hierarchical models are more appropriate in developing a relationship between area-wide traffic crashes and the contributing factors associated with the road infrastructure, socioeconomic and traffic conditions of the area. This is because Bayesian models accurately take account of both spatial dependence and uncorrelated heterogeneity.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The primary objective of this study is to develop a series of relationships (i.e., crash prediction models) between area-level traffic casualties and their contributing factors using both non-spatial (such as negative binomial models) and spatial models (such as traditional spatial models and Bayesian hierarchical models) and to compare the results obtained from these models. The spatial units of this analysis are the 633 census wards from London. Ward-level crash data are disaggregated by severity of the casualty such as fatalities, serious injuries and slight injuries and by severity of the casualty associated with various road users such as motorised transport (MT), non-motorised transport (NMT) and vulnerable road user (VRU). A range of potential contributing factors associated with ward-level road infrastructure, traffic, socioeconomic

characteristics including traffic speed, flow and road curvature are considered in this study.

Crash prediction models to explain observed cross-sectional variations in crash counts using macro-structural covariates at various levels of area-aggregation (e.g., states, counties, other census tracts, etc.) are becoming a fairly routine component in crash research. Researchers usually seek to establish links between the road infrastructure, environmental, traffic, and socioeconomic conditions in spatial units with the counts or rates of traffic crashes observed at various spatial units. To isolate and identify the macro-processes leading to different types of crashes, researchers sometimes estimate crash models with disaggregated crash rates with varying bases for the disaggregation such as by severity of the casualty (such as fatalities, serious injuries, and slight injuries) or by severity of the casualty related to various road users (e.g., motorised transport, vulnerable road users, etc.).

For instance, Levine et al. (1995a) derived a series of statistics that provide explicit measurements of a spatial pattern of crashes and also provide insights into how certain relationships

* Tel.: +44 1509 22 8545; fax: +44 1509 22 3981.

E-mail address: m.a.quddus@lboro.ac.uk.

(for instance, between alcohol consumption and injury severity) have a spatial dimension. Honolulu census tract data have been utilised to explain spatial variation in motor vehicle crashes (Levine et al., 1995b). Kim and Yamashita (2002) conducted an empirical analysis of motor vehicle crashes and land use variables with the aid of geographic information systems (GIS). Hadayeghi et al. (2003) developed a series of macro-level crash prediction models that estimate the number of crashes in planning zones in the city of Toronto as a function of zonal characteristics. Graham and Glaister (2003) analysed ward-level (a census tract) traffic casualty data in England to see how urban scale, density and land use mix affect pedestrian casualties. Noland and Quddus (2004a) also conducted a spatially disaggregated ward-level analysis of England to identify various factors affecting road casualties and were based on cross-sectional traffic crash data associated with different levels of spatial aggregation, most of these above studies were employed a negative binomial (NB) count model. NB models have also been used to develop crash models for cross-sectional time-series data (e.g., Amoros et al., 2003; Noland and Quddus, 2004b¹, and Noland and Oh, 2004²). Integrating NB count models with geographic information systems (GIS), Kim et al. (2006) established the nature and magnitude of relationships between land use, population, economic development and crashes using a uniform 0.1 square mile grid structure from Hawaii. Their study confirmed the finding of Ladron de Guevara et al. (2004) that population-based metrics by spatial units are the most statistically significant predictors of crash occurrences. Kim et al. (2006) recommended the use of a spatial statistical analysis when developing relationships between area-wide land use variables and traffic crashes.

Crash data are collected with reference to location measured as points (*x*- and *y*-coordinates) in space. According to LeSage (1998), two problems arise when sample data has a locational dimension: (1) spatial correlation exists between the observations, and (2) spatial heterogeneity occurs in the relationships that are modelled. Traditional econometrics (including NB models used in crash research) has largely ignored the issue of spatial correlation that violates the traditional Gauss–Markov assumptions used in regression modelling.

An alternative approach is to employ spatial econometric models. Anselin (1988) provides a complete treatment of many aspects of spatial econometrics including the application of Bayesian methods in spatial econometrics. There are generally two methods in spatial econometrics: (1) traditional econometric methods suitable for continuous data, and (2) Bayesian hierarchical methods suitable for non-negative random count data. At higher levels of spatial aggregation (e.g., districts, counties, states), when the number of counts (e.g., crashes) is sufficiently large and non-zero counts are observed in most of the sampled spatial units, the count outcomes may be considered continuous, and traditional spatial analytical methods have been utilised (Messner et al., 1999; Baller et al., 2001).

However, Bhati (2005) indicated that inferences derived from traditional spatial models could be misleading as this does not reflect the true underlying data generating processes. Moreover, as the spatial unit of analysis becomes smaller (such as wards, zip-code, post-code, etc.), the number of count outcomes observed in each sampled unit decreases and the distribution of such counts becomes a highly skewed (to the right) distribution as the number of spatial units with zero counts increases. In order to overcome these issues, researchers used a more flexible Bayesian method in spatial econometrics (Besag et al., 1991; Mollie, 1996; Wolpert and Ickstadt, 1998; Best et al., 2000) and the application of such meth-

ods to crash modelling can be found in Miaou et al. (2003), MacNab (2004), Aguero-Valverde and Jovanis (2006) and Li et al. (2007). Miaou et al. (2003) provides a good overview on the appropriateness of employing a Bayesian hierarchical model in area-wide crash modelling.

Different area-wide characteristics were considered in previous research while developing a crash prediction model using either a non-spatial model (such as an NB model) or a spatial model. These include factors associated with land use (e.g., Graham and Glaister, 2003; Kim et al., 2006), road characteristics such as road length, junctions and roundabouts (e.g., Noland and Quddus, 2004a), environmental conditions such as total precipitation, number of rainy days per year and total snowfall (e.g., Aguero-Valverde and Jovanis, 2006), and various socioeconomic factors such as population, poverty, and employment (e.g., Kim et al., 2006). Some other important area-wide characteristics that affect area-wide traffic crashes are less considered in the literature. These include traffic speed, traffic flow, and road curvature measured at spatial-level.

The rest of the paper is structured as follows. In the next section, a brief discussion of the data used in the analysis is presented. This is followed by a description of the models considered in this research. Then, the results obtained from the developed models are presented with the similarities and differences among them. Finally, conclusions are drawn and further research suggestions are discussed.

2. Data

The spatial units of the analysis are the census wards from the Greater London metropolitan area. According to the UK Census 2001, there are 633 wards in London and each ward consists, on average, of about 11,350 resident population. The electronic ward boundary data were obtained from UK Ordnance Survey (OS) data via EDINA services. Data on traffic casualties were extracted from the STATS19 UK National Road Crash Database that has information on the recorded location of each crash. Ward-level casualty data were extracted from the STATS19 data using a GIS technique. Since previous research suggests that factors affecting traffic casualties vary by severity of the casualty (e.g., Noland and Quddus, 2004a; Aguero-Valverde and Jovanis, 2006), ward-level traffic casualty data were disaggregated into fatalities, serious injuries and slight injuries. Ward-level casualty data were also disaggregated by severity of the casualty associated with motorised transport (MT³), non-motorised transport (NMT⁴), and vulnerable road users (VRU⁵) to identify any differences in influential factors. Since a large number of wards have a zero fatality count, fatalities are combined with serious injuries resulting in a killed or serious injury (KSI) category. Casualty data were aggregated for 3 years of data, 2000–2002. Fig. 1 shows the spatial distribution of total serious casualties (for years 2000–2002). It is noticeable from this figure that traffic casualties are spatially correlated among neighbouring wards.

There are three major categories of explanatory variables: (1) traffic characteristics, (2) road characteristics and (3) socio-demographic factors. Environmental factors are not considered in this study as weather conditions such as snowfall and rainfall tend to be similar across different wards in London.

¹ They used a random effects NB model.

² They used a fixed effects NB model.

³ Cars, taxi, bus, goods vehicles, and other motor vehicles (motorcycles are not included).

⁴ Pedestrians, cyclists, horse riders.

⁵ Motorcyclists, pedestrians, cyclists, horse riders.

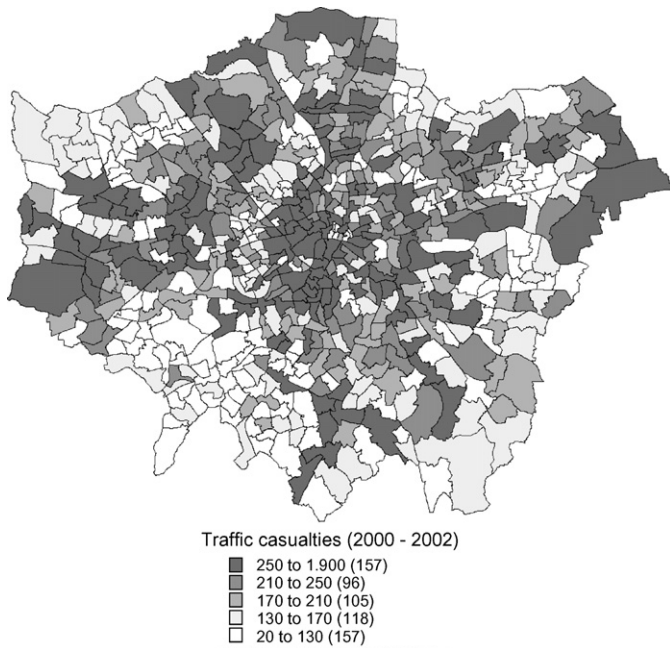


Fig. 1. Spatial distribution ward-level total traffic casualties in London (2000–2002).

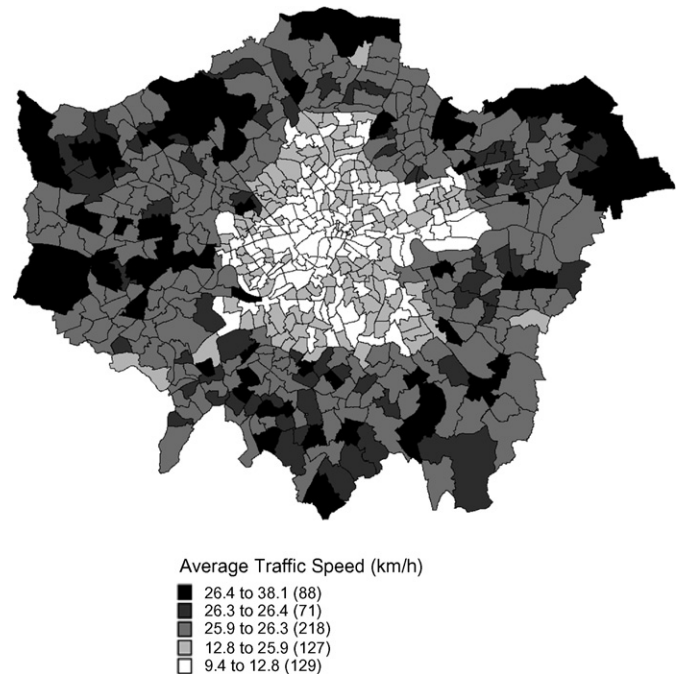


Fig. 2. Ward-level average traffic speed (km/h) in London.

2.1. Traffic characteristics

Traffic characteristics such as traffic speed and flow can be considered as the most important factors influencing traffic crash occurrences. Most previous studies conducting on area-wide spatial analysis of traffic crashes do not consider them as such data are not normally available at a spatial-level. Glaister and Graham (2006) in their study on the spatial implications of transport pricing estimated ward-level traffic speed (km/h) and traffic flow (PCU⁶-km/h) for England using detailed spatially disaggregated road-level traffic flow and speed relationship data (for the year 2000) supplied by the UK Department for Transport (DfT). With the authors' permission, this data are used in this study. Fig. 2 shows the spatial distribution of ward-level traffic speed (km/h) within London. As expected, outer London wards have higher average speeds than inner London wards. The other traffic variable—number of registered cars was obtained from the UK Census 2001.

In addition to these above factors, it is also important to include a ward-level *exposure to risk* variable in a crash prediction model. The ideal exposure variable would be annual vehicle miles travelled (VMT) in each ward. Unfortunately, such data are not available at the ward-level. Therefore, a good measure of an alternative exposure variable was sought. Census 2001 data include the total number of registered cars in each ward. This could be used as an exposure variable. However, the problem is that cars registered to the neighbourhood wards might also have travelled to the ward of interest. To account for such cars from the surrounding areas, the following gravity model is employed to measure the exposure to risk of a ward.

$$TC_i = C_i + \sum_{j=1}^{n-1} \frac{C_j}{d_{ij}} \quad i \neq j \quad (1)$$

⁶ Traffic flows are expressed not in vehicles but in generic units called passenger car units (PCU) which are an indicator of the space that different vehicles occupy relative to that of a passenger car.

where, TC_i represents the amount of total cars that might have travelled within ward i , C_i is the cars registered at ward i , C_j is the cars registered at ward j , d_{ij} is the centre to centre distance between wards i and j , and n is the total number of wards. The largest distance from the centre of Greater London to the centre of the outermost ward is about 30 km (in the E–W direction). When calculating TC_i using Eq. (1), wards which fall within the distance of 40 km from the centre of Greater London are considered. This is to take into account cars that might have entered to the outer London wards from outside of London. The resulting figure approximates the number of cars driving around the ward.

2.2. Road characteristics

Various factors associated with the road infrastructure are thought to be responsible for crash occurrences. The roadway factors considered in this study are junctions, roundabouts and length of various types of roads such as motorways, A roads, B roads, and minor roads. Data on these features of the road network were obtained from the UK Ordnance Survey. These variables were measured at ward-level using a GIS technique. There is a large variation among the wards in terms of their geographic areas with the smallest ward has an area of 0.13 km² and the largest ward has an area of 28.9 km². A ward with a relatively high value of roadway length may indicate a dense road network that may result more occurrences of injuries. Wards within the city centre normally fall in this category. Junctions are normally known as crash hot-spots. Therefore, wards containing more junctions may have more traffic casualties. Lord et al. (2007) stated that roundabouts have the potential to significantly reduce the crashes involving angle movements and the severity of crashes by slowing the operating speed of all vehicles, while maintaining a high capacity for moving traffic through a junction. Therefore, the presence of more roundabouts within a ward may also affect traffic casualties. However, 296 wards in London have no roundabouts suggesting that this variable may not pick up the real effect of roundabout on traffic casualties.

Previous research suggests that areas (local authority districts of England) with straighter roads have more crashes than areas with many road bends (Haynes et al., 2007). To investigate this further for the case of wards (relatively much smaller than districts) within London, the average road curvature of a ward was estimated from a high accuracy digital road map of London (map scale 1:2500). The following equation developed by Haynes et al. (2007) was employed to calculate ward-level road curvature known as bend density (BD):

$$BD_i = \frac{SP_i - n_i}{L_i} \quad \text{where } L_i \neq 0 \quad (2)$$

where SP_i represents the total shape points (i.e., vertices) of all roads within a ward i , n_i is the total number of nodes (junctions and end points of roads) of ward i , and L_i is the total length of all roads in ward i . BD of a ward with $L = 0$ is assumed to be zero.⁷ It is envisaged that wards with a high value of road curvature may lead to fewer serious injury crashes as it is thought to be a protector factor (Haynes et al., 2007).

2.3. Socio-demographic factors

The literature suggests that various socio-demographic variables such as population, employment, poverty, economic activity affect traffic casualties (e.g., Graham and Glaister, 2003; Noland and Quddus, 2004a; Aguero-Valverde and Jovanis, 2006; Kim et al., 2006). It is expected that different age groups affect traffic casualties differently. Therefore, ward's population (obtained from the UK Census 2001) is divided into two age groups such as resident population under aged 60 years and population aged 60 years or over. Ward-level total employment was also obtained from the UK Census 2001. Previous research also suggests that area deprivation has an impact on traffic crashes (Graham and Glaister, 2003; Noland and Quddus, 2004a; Aguero-Valverde and Jovanis, 2006). Unfortunately, we do not have data on ward-level income deprivation. The percentage of households with no cars within a ward is used as a proxy to measure the level of poverty of that ward.

Table 1 shows summary statistics of both dependent and explanatory variables used in the analysis. This includes mean, standard deviation (S.D.), minimum and maximum values of the variables.

3. Methods

As discussed above, cross-sectional count data may be modelled in two ways: (1) using non-spatial models, and (2) using spatial models. Both of these modelling techniques are briefly presented below.

3.1. Non-spatial models

Crashes are discrete and non-negative, as well as possibly random and sporadic. Therefore, it may be natural to model crash outcomes using a Poisson or a NB regression model (e.g., Jovanis and Chang, 1986; Shankar et al., 1995; Miaou, 1994). Regardless of whether the model adopted is a Poisson model or a NB model, there is an assumption that total crash counts for an entity (i.e., a ward) i during a particular period (i.e., Y_i) are independent from the occurrence of crashes in any neighbouring wards j with a shared border (i.e., the assumption of independence across geographic space). Previous research suggests that crash data normally follow a NB distribution (i.e., a Poisson-gamma mixture) which

can address the issue of extra-Poisson variation (i.e., overdispersion) normally present in crash data (e.g., Miaou, 1994; Shankar et al., 1995; Abdel-Aty and Radwan, 2000; Lord, 2000; Ivan et al., 2000). Overdispersion may result from unobserved heterogeneity that is inadequately captured by the explanatory variables (\mathbf{X}) in the conditional mean function (Cameron and Trivedi, 1998). Using a Bayesian modelling approach, Mitra and Washington (2007) reported that extra-variation is a function of explanatory variables when the mean function (expected crash count) is poorly defined and suffers from omitted variables.

A commonly used log-linear relationship between the expected numbers of crashes in ward i i.e., $E[Y_i|X_i] = \theta_i$ and the covariates \mathbf{X} for the case of Poisson distribution⁸ is:

$$\ln \theta_i = \ln(EV_i) + (\beta_0 + \beta X_i) \quad (3)$$

where β is a vector of estimable coefficients representing the effects of the covariates, EV_i is the exposure variable and β_0 is the intercept term. If heterogeneity is present in the observations, overdispersion may occur as discussed above. Although the source of overdispersion in event count data cannot be distinguished, its presence can be adjusted by introducing a stochastic component (i.e., ε_i) into the relationship so that Eq. (3) can be rewritten as

$$\ln \theta_i = \ln(EV_i) + (\beta_0 + \beta X_i + \varepsilon_i) \quad (4)$$

where ε is a random error that is assumed to be uncorrelated with \mathbf{X} . One can think of ε either as the combined effects of unobserved variables that have been omitted from the model (Gourieroux et al., 1984) or as another source of pure randomness (Hausman et al., 1984). Eq. (4) is known as a NB regression model given that ε is assumed to be independent of \mathbf{X} and has a two-parameter gamma distribution.

The NB model presented in Eq. (4) is able to control for unobserved heterogeneity caused by omitted variables. However, this model may not be able to take account of any spatial dependence (i.e., autocorrelation effects) among spatial units (e.g., wards). Therefore, NB regression models estimated by maximum likelihood (ML) largely ignore the issue of spatial correlation that violates the traditional Gauss-Markov assumptions used in traditional regression modelling. The generalised estimation equation (GEE) with the NB as the link function has been employed to describe the temporally or spatially correlated crash frequencies at the intersection level (e.g., Lord and Persaud, 2000; Wang and Abdel-Aty, 2006). Such a procedure may not be able to capture spatial dependence of traffic crashes at spatial units. Spatial models controlled for the impact of neighbourhood area are able to properly account for both spatial correlation and unobserved heterogeneity (Anselin, 1988).

3.2. Spatial models

As discussed, there are generally two techniques in spatial econometrics: (1) traditional econometric methods suitable for cross-sectional continuous spatial data, and (2) Bayesian hierarchical methods suitable for non-negative random count spatial data.

3.2.1. Traditional econometric methods

Two specifications of spatial econometric models are available for continuous data: (i) adding an explanatory variable in the form of a spatially lagged dependent variable known as spatial autoregressive (SAR) model, or (ii) adding spatially lagged error structure known as spatial error model (SEM) (Anselin, 1988). These models are presented below:

⁷ None of the London wards has a value of total road length equal to zero (0).

⁸ $P(Y_i) = \frac{e^{-\theta_i} (\theta_i^{Y_i})}{Y_i!}$.

Table 1
Summary statistics of ward-level factors

Variables	Observations	Mean	S.D.	Minimum	Maximum
Response variables					
Total fatalities	633	1.379	1.465	0	10
Total serious injuries	633	26.758	18.185	1	236
Total slight injuries	633	179.814	122.425	25	1650
MT killed or seriously injured (KSI)	633	11.258	8.487	0	82
MT slight injuries	633	102.724	69.718	3	907
NMT killed or seriously injured (KSI)	633	9.463	9.069	0	146
NMT slight injuries	633	39.272	39.682	0	710
VRU killed or seriously injured (KSI)	633	14.981	12.955	0	196
VRU slight injuries	633	66.281	64.346	0	1080
Traffic characteristics					
Traffic flow (PCU km/h)	633	1.1E+07	5.4E+06	2.9E+06	6.5E+07
Average speed (km/h)	633	21.324	6.761	9.413	38.080
Road characteristics					
# of junctions	633	85.596	30.613	15	367
# of roundabouts	633	1.090	1.616	0	16
Length of motorway (km)	633	0.103	0.710	0.000	10.551
Length of A-road (km)	633	2.745	1.978	0.000	18.630
Length of B-road (km)	633	0.789	1.032	0.000	11.166
Length of minor road (km)	633	16.399	6.690	0.710	53.513
Road curvature (1/m)	633	0.029	0.006	0.013	0.053
# of bus stops	633	27.439	13.005	4.000	92.000
Socio-demographic factors					
Resident population aged less than 60	633	9931.9	2054.2	100	14951
Resident population aged 60 or over	633	1413.7	465.2	3	3087
# of employees	633	6095.2	12656.5	523	173127
# of households with no cars	633	1774.8	815.2	33	3976
Total cars (estimated using Eq. (1))	633	4359.1	1442.1	139.7745	8311.519

The SAR model is given below:

$$Z_i = \rho WZ_i + \beta X_i + \varepsilon_i \tag{5}$$

where Z is a $N \times 1$ vector of cross-sectional dependent variable, WZ is a spatially lagged dependent variable for spatial weights matrix W , ρ is the scalar for spatial lag coefficient, β is the vector of parameters to be estimated, X is the matrix of exogenous explanatory variables, ε is an $N \times 1$ vector of normally distributed error terms with zero mean and variance σ^2 . The spatial lag WZ can be considered as a spatially weighted average of the dependent variable at neighbouring spatial units. This spatial lag term is assumed to be correlated with error terms, even though the later are independent and identically distributed (Anselin, 1988).

The SEM model is given below:

$$\begin{aligned} Z_i &= \beta X_i + u_i \\ u_i &= \lambda W u_i + \varepsilon_i \end{aligned} \quad \text{where } \varepsilon_i \sim N(0, \sigma^2 I_n) \tag{6}$$

where, u is the error term expressing spatial dependence and λ is the spatial autoregressive coefficient, and all other terms are as previously defined.

A possible third model is the combination of Eqs. (5) and (6) in which both spatial lag and spatial error are simultaneously considered.

The procedure to model a count dataset using Eqs. (5) or (6) is to convert the cross-sectional count dependent variable into an approximately continuous variable and then apply a SAR model or a SEM model. The most commonly used transformation is the log transformation. In addition to this, the dependent variable, Z , needs to be converted into a rate variable to control for the exposure variable (EV). This allows us to compare the results (i.e., the values of the coefficients) between NB models and SAR or SEM models. Therefore, Eqs. (5) and (6) can be expressed as

$$\ln \left(\frac{Y_i}{EV_i} \right) = \rho W \ln \left(\frac{Y_i}{EV_i} \right) + \beta X_i + \varepsilon_i \quad \text{where } \varepsilon_i \sim N(0, \sigma^2 I_n) \tag{7}$$

$$\begin{aligned} \ln \left(\frac{Y_i}{EV_i} \right) &= \beta X_i + u_i \\ u_i &= \lambda W u_i + \varepsilon_i \end{aligned} \quad \text{where } \varepsilon_i \sim N(0, \sigma^2 I_n) \tag{8}$$

These models can be estimated using a maximum likelihood (ML) technique (Anselin, 1988).

3.2.2. Bayesian hierarchical model

The Bayesian hierarchical model proposed by Besag et al. (1991) and subsequently used by Aguero-Valverde and Jovanis (2006) and Li et al. (2007) is presented below:

$$\begin{aligned} Y_i &\sim \text{Poisson}(\theta_i) \\ \ln \theta_i &= \ln(EV_i) + (\beta_0 + \beta X_i) + SC_i + UH_i \end{aligned} \tag{9}$$

where SC_i are the spatial random effects (i.e., spatial correlation), UH_i are the unobserved (i.e., uncorrelated) heterogeneity, and all other terms are as previously defined. It is noticeable from Eq. (9) that both SC_i and UH_i are included as multiplicative terms in the mean conditional function so that they provide a non-negative mean of crashes. This model allows the data to decide how much of the residual crash risk is due to spatially structured variation (SC_i), and how much is unstructured over-dispersion (UH_i).

In this Bayesian hierarchical model the posterior distribution of all unknown parameters is proportional to the product of the likelihood and the prior distributions. A uniform prior distribution is assumed for β_0 , a highly non-informative Normal prior is assumed for all β 's with zero mean and variance 10,000, UH_i are assigned as an exchangeable normal prior with $N(0, \tau_{UH}^2)$ where τ_{UH}^2 is the precision (i.e., 1/variance) with a prior gamma distribution $Ga(0.5, 0.0005)$ as suggested by Wakefield et al. (2000) and used by Aguero-Valverde and Jovanis (2006). The effects of spatial correlation, SC_i , are assigned as a conditional auto-regressive (CAR)

Table 2
Variables included in different models

Variables	Total			MT		NMT		VRU	
	Fatalities	Serious injuries	Slight injuries	KSI	Slight injuries	KSI	Slight injuries	KSI	Slight injuries
Traffic characteristics									
ln(traffic flow (PCU/h))	✓	✓	✓	✓	✓	✓	✓	✓	✓
Average speed (km/h)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Road characteristics									
Length of motorway (km)	✓	✓	✓	×	✓	×	×	✓	✓
Length of A-road (km)	×	✓	✓	✓	✓	✓	✓	✓	✓
Length of B-road (km)	×	✓	✓	×	✓	✓	✓	✓	✓
Length of minor road (km)	×	✓	✓	✓	✓	✓	✓	✓	✓
ln(road curvature)	×	×	×	✓	✓	×	×	✓	✓
ln(# of bus stops)	×	×	×	×	×	✓	✓	✓	✓
Socio-demographic factors									
ln(resident population aged less than 60)	✓	✓	✓	✓	✓	✓	✓	✓	✓
ln(resident population aged 60 or over)	✓	✓	✓	✓	✓	✓	✓	✓	✓
ln(# of employees)	✓	✓	✓	✓	✓	✓	✓	✓	✓
ln(# of households with no cars)	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ means the variable is included in the model, × means the variable do not included in the model.

prior with $N(\bar{S}_i, \tau_i^2)$ as suggested by Besag (1974) where:

$$\bar{S}_i = \frac{\sum_j SC_j w_{ij}}{\sum_j w_{ij}} \quad \text{and} \quad \tau_i^2 = \frac{\tau_{SC}^2}{\sum_j w_{ij}} \quad \text{for } i \neq j \quad (10)$$

in which $w_{ij} = 1$ if i and j are adjacent (i.e., a shared border) or 0 otherwise⁹, τ_{SC}^2 is also assumed as a gamma prior with $Ga(0.5, 0.0005)$.

The Bayesian hierarchical model can be estimated using the Markov chain Monte Carlo (MCMC) method (Gelman et al., 1995; Carlin and Louis, 1996; Brooks, 1998).

4. Estimation results

The relationship between ward-level traffic casualties and different contributing factors has been developed using both a non-spatial model such as a NB model and spatial models including a spatial autoregressive (SAR) or a spatial error model (SEM) and a Bayesian hierarchical model. Since ward-level casualty data are disaggregated by severity of the casualty (such as fatalities, serious injuries, and slight injuries) and by severity of the casualty related to road user (i.e., MT, NMT, and VRU), the relationship has been developed for categories of casualty. This has led to the development of a total of nine models such as fatalities, serious injuries, slight injuries, KSI related to MT, slight injuries related to MT, KSI related to NMT, slight injuries related to NMT, KSI related to VRU and slight injuries related to VRU. The explanatory variables for each of the models have been chosen based on the principle that the variables are theoretically important to explain the variation in area-wide crash counts and there has been no multicollinearity among them. For instance, the variable—length of motorway has not been included in the models related to NMT casualties as this category of road users does not use motorways. Any explanatory variable found to be statistically insignificant in all models has been dropped from the models. For example, the variable—number of junctions and roundabouts within a ward was found to be statistically insignificant in all models and therefore, omitted from the set

of explanatory variables. The final sets of explanatory variables for each of the nine models are shown in Table 2. The second column of this table represents the following relationship between the counts of total fatalities and their contributing factors.

Counts of total fatalities

$$= f(\text{traffic flow, traffic speed, length of motorway, residents under 60, residents 60 or over, number of employees, number of households with no cars})$$

In such a way, the relationships for the other response variables can be shown. It is noticeable that traffic characteristics such as traffic flow and average speed are important in all models. Socio-demographic factors such as resident population¹⁰, employment and number of households with no cars are also included in all models. Some of the explanatory variables with non-zero observations are transformed into the logarithmic scale, including traffic flow, curvature, population, employment and households with no cars. The log transformation generally reduces the variances among the variables (within and between) and minimises the heteroscedasticity among the variables (Gujarati, 2003). Another advantage of using the log transformation is the easy interpretation of a model parameter, which is equal to a constant elasticity rather than a slope coefficient.

The primary hypothesis was that ward-level factors affecting different casualty would be different and the results from the Bayesian hierarchical model would be more coherent compared with the results of NB model and SAR or SEM models. The results from each of the models with the similarities and differences among them are presented below.

4.1. Non-spatial models

A series of NB models were developed at different levels of casualties such as total fatalities, serious injuries and slight injuries. The

⁹ There are normally two ways to generate w_{ij} to be employed in Eq. (10): (1) the distance among spatial units and (2) the contiguity among spatial units. Since crashes are clustered locally in various parts of London, it is assumed that the rook contiguity based weights matrix would be more appropriate.

¹⁰ More disaggregated age cohorts including population under 18, population 18–24, population 25–59 and population 60 or over were also considered. However, population under 18 and population 18–24 were consistently found to be statistically insignificant. Therefore, two age cohorts are used in the analysis.

Table 3
Estimation results for NB models by severity of the casualty

NB models	Total traffic casualties					
	Fatal		Serious injuries		Slight injuries	
	Coef	t-stat	Coef	t-stat	Coef	t-stat
Traffic characteristics						
ln(Traffic flow (PCU km/h))	1.3024	8.14	0.3694	1.91	0.3455	1.79
Average speed (km/h)	−0.0370	−3.81	−0.0375	−5.79	−0.0334	−5.27
Road characteristics						
Length of motorway (km)	0.1760	3.71	0.1350	4.07	0.1366	4.08
Length of A-road (km)	–	–	0.0566	2.14	0.0590	2.23
Length of B-road (km)	–	–	0.0546	2.12	0.0332	1.33
Length of minor road (km)	–	–	0.0389	5.20	0.0424	5.97
Socio-demographic factors						
ln(Resident population aged less than 60)	−0.0114	−0.05	0.2427	1.65	0.4978	3.71
ln(Resident population aged 60 or over)	−0.3116	−1.94	−0.3425	−3.64	−0.5095	−5.42
ln(# of employees)	−0.0235	−0.37	0.0898	2.39	0.1044	2.85
ln(# of households with no cars)	0.3994	2.65	0.2008	2.21	0.0852	1.04
Constant	−28.8	−12.66	−13.0	−4.50	−11.2	−3.90
Overdispersion parameter	0.2591	4.23	0.3025	14.68	0.3330	17.90
Observations	633.0		633		633	
Log-likelihood ratio test of overdispersion	Reject H0: of no overdispersion		Reject H0: of no overdispersion		Reject H0: of no overdispersion	

results are presented in Table 3. The overdispersion parameter in all three models was found to be statistically and significantly different from zero suggesting that the use of NB models is more appropriate than the use of Poisson models for the data analysed. In addition to this, the log-likelihood ratio test for overdispersion rejects the null hypothesis of no overdispersion in all three models (see Table 3). It is well known that count models (i.e., a NB model) need a mechanism to deal with the fact that counts (i.e., total casualties) can be made over different observation scenarios. For instance, we have ward-level traffic casualty counts for 633 different wards in London. The amount of total cars estimated by Eq. (1) was taken as the exposure variable in all NB models.

The majority of explanatory variables in the fatality model were found to be statistically significant with the expected signs. Both traffic characteristics variables were statistically significant. All road characteristics variables were also significant except the variable—length of B-road. Resident population were divided into two categories: (1) population aged under 60 and (2) population aged 60 and over. The results suggest that wards with more old people are associated with fewer traffic fatalities. This may be an expected result given that mobility of the older people is relatively less compared with the younger people who tend to take more risks while travelling either as a driver/rider or a pedestrian/cyclist. The variable—households with no cars which was used as a proxy for the poverty of the ward was positively associated with fatalities. This is consistent with previous research (Graham and Glaister, 2003). One unexpected result in the fatality model is that average speed shows a negative association with fatalities suggesting that fatalities within a ward decrease with the increase of ward-level average traffic speed. This is not consistent with the hypothesis that fatalities increase with the increase in average traffic speed. This issue will be discussed further in the subsequent sections. As shown in Table 3, a separate set of variables were selected for the serious injury model. All variables were found to be statistically significant. Resident population aged under 60 were positively associated and population aged 60 or over were negatively associated with serious injuries. Unlike the fatality model, length of B-road and total employment became statistically significant in the serious injury model. The results of the slight injury model were found to be almost similar to the results of the serious injury model except the variable—household with no cars is insignificant. The signs of the coefficients of the variables are consistent with previous studies

with the exception of the impact of average speed on both serious and slight injuries.

Another series of NB models were estimated for casualties related to different road users including MT, NMT and VRU. Two types of casualty such as KSI and slight injuries were considered for each category of road users such as KSI and slight injuries. Six NB models (3×2) were then estimated. The results are shown in Table 4. Again, the log likelihood ratio test rejects the null hypothesis of overdispersion suggesting that NB models are superior to Poisson models. Table 4 shows clear differences between these and the earlier models suggesting that ward-level factors affecting casualties related to road user type are different. Traffic flow was statistically insignificant in all models. This is may be due to the fact that the effect of this variable was picked up by other exposure type variables such as total employment and resident population. The statistically significant variables across all six models were length of A-road, length of minor road, and resident population aged 60 or over with the expected signs. Area-wide traffic speed was found to be statistically significant with a wrong negative coefficient in all models except in the MT KSI model. Both total employment and households with no cars were found to be positively associated with NMT and VRU casualties. The variable—road curvature which is a proxy for the amount of curved roads within a ward was consistently found to be insignificant in all models. Surprisingly, the variable—number of bus stops within a ward was also found to be statistically insignificant in all models associated with NMT and VRU casualties.

Residuals from all nine NB models were analysed to investigate whether they were spatially correlated using the global Moran's I test. The results suggest that residuals from all models were spatially correlated (except the NB model for MT KSI in which average speed was found to be statistically insignificant). This is not surprising as crashes are spatially clustered in different parts of London (see Fig. 1). It is known that models with spatially correlated residuals may result biased estimates of the parameters which may lead to erroneous interpretation and wrong conclusions (Anselin, 1988; LeSage, 1998). This could be one of the reasons why NB models showed the wrong sign in the coefficient of average speed. Another possible reason might be that data for the best exposure variable, VMT, are not available at ward-level to be included in the models. It is interesting to see whether the spatial models to be presented in the following sections can provide the correct estimates for the impact of average speed on traffic casualties.

Table 4
 Estimation results for NB models by severity of casualties by road user type

Non-spatial NB models	MT casualties				NMT casualties				VRU casualties			
	KSI		Slight injuries		KSI		Slight injuries		KSI		Slight injuries	
	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Traffic characteristics												
ln(Traffic flow (PCU km/h))	0.1804	0.75	0.2386	1.02	-0.0867	-0.39	-0.2751	-1.29	0.1838	0.81	-0.1091	-0.45
Average speed (km/h)	-0.0120	-1.5	-0.0167	-2.19	-0.0172	-2.24	-0.0241	-3.31	-0.0386	-5.04	-0.0406	-5.06
Road characteristics												
Length of motorway (km)	-	-	0.0879	1.97	-	-	-	-	0.1142	2.60	0.0473	0.99
Length of A-road (km)	0.1015	3.37	0.0820	2.55	0.0979	3.37	0.0876	3.29	0.0663	2.10	0.0910	2.71
Length of B-road (km)	-	-	0.0172	0.59	0.0683	2.17	0.0813	2.61	0.0814	2.67	0.0930	2.90
Length of minor road (km)	0.0645	7.32	0.0502	5.79	0.0201	2.28	0.0379	4.62	0.0175	1.87	0.0370	3.89
ln(Road curvature)	-3.3205	-0.48	0.2761	0.05	-	-	-	-	-7.7215	-1.23	-8.1574	-1.25
ln(# of bus stops)	-	-	-	-	0.0311	0.39	0.0369	0.48	0.1027	1.35	0.0633	0.80
Socio-demographic factors												
ln(Resident population aged less than 60)	0.2279	1.2	0.6585	4.10	0.1445	0.77	0.2753	1.59	0.1781	1.01	0.2885	1.64
ln(Resident population aged 60 or over)	-0.3169	-2.53	-0.5004	-4.53	-0.4675	-3.97	-0.5413	-4.68	-0.4006	-3.63	-0.4962	-4.22
ln(# of employees)	0.0214	0.43	0.0243	0.55	0.1636	3.46	0.2022	4.41	0.1143	2.55	0.1730	3.68
ln(# of households with no cars)	0.0679	0.58	-0.0752	-0.76	0.5602	4.78	0.3789	3.51	0.3621	3.29	0.2704	2.49
Constant	-10.3	-2.87	-10.29	-2.92	-8.6	-2.55	-3.9	-1.20	-10.81	-3.18	-4.9077	-1.36
Overdispersion parameter	0.4995	14.55	0.4619	18.21	0.4172	13.16	0.4765	16.89	0.3912	14.07	0.5066	17.46
Observations	633		633		633		633		633		633	
Log-likelihood ratio test of overdispersion	Reject H0: of no overdispersion		Reject H0: of no overdispersion		Reject H0: of no overdispersion		Reject H0: of no overdispersion		Reject H0: of no overdispersion		Reject H0: of no overdispersion	

Table 5
Estimation results for SEM models

Spatial error models (SEM)	Total serious		Total slight		Motorised slight	
	Coef	t-stat	Coef	t-stat	Coef	t-stat
Traffic characteristics						
ln(Traffic flow (PCU km/h))	0.5713	3.08	0.5882	3.57	0.3499	1.56
Average speed (km/h)	-0.0215	-2.77	-0.0163	-1.34	-0.0101	-1.12
Road characteristics						
Length of motorway (km)	0.1025	2.89	0.1228	3.89	0.0730	1.70
Length of A-road (km)	0.0507	2.07	0.0358	1.64	0.0873	2.93
Length of B-road (km)	0.0091	0.40	-0.0071	-0.35	0.0174	0.64
Length of minor road (km)	0.0191	2.57	0.0166	2.51	0.0207	2.29
ln(Road curvature)	-	-	-	-	-1.9405	-0.29
Socio-demographic factors						
ln(Resident population aged less than 60)	0.1209	0.83	0.0562	0.43	0.1656	0.92
ln(Resident population aged 60 or over)	-0.2588	-2.67	-0.2295	-2.64	-0.2493	-2.14
ln(# of employees)	0.1204	3.25	0.1269	3.85	0.0382	0.85
ln(# of households with no cars)	0.3160	3.26	0.3344	3.86	0.1843	1.56
Constant	-16.95	-6.10	-15.14	-6.12	-11.21	-3.33
Spatial autoregressive coef (λ)	0.8246	16.02	0.8478	17.71	0.7778	13.23
Observations	633		633		633	
Tests for $\lambda = 0$						
Wald test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	
Likelihood ratio test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	
Lagrange multiplier test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	

4.2. Spatial models

As discussed in the methodology section, two types of spatial models are employed: (1) traditional spatial econometric models and (2) Bayesian hierarchical model.

4.2.1. Traditional spatial econometric model

Since ward-level casualty data for most casualty categories include wards with zero casualty count, Eqs. (7) or (8) could only be employed for three categories of casualties namely total serious injuries, total slight injuries and MT slight injuries (see Table 1). Therefore, traditional spatial models have only been applied to these three types of casualties.

First of all, the spatial association of ward-level casualty data (i.e., serious injuries, slight injuries, and MT slight injuries) was tested using a global *Moran's I*, which measures similarities and dissimilarities in observations across space (Anselin, 1988). The

test rejects the null hypothesis of no spatial correlation at the 95% confidence level (p -value < 0.005) for all three cases and *Moran's I* statistic was found to be a positive and significant suggesting that there is a positive *spatial correlation* among ward-level traffic crashes. A SAR model (also known as a spatial lag model) was applied to each of these three categories of casualties. However, the autoregressive coefficient (ρ) was found to be statistically insignificant in all cases suggesting that a SAR specification is not appropriate for the data. A SEM model was then applied and the autoregressive coefficient (λ) was found to be statistically significant at the 100% confidence level in all three cases. In addition to this, a number of statistical tests namely the Wald test, the likelihood ratio test and the Lagrange multiplier (LM) test all reject the null hypothesis of no spatial dependence in the error term. The results from the SEM model for total serious injuries, slight injuries and MT slight injuries are presented in Table 5.

Table 6
Estimation results for Bayesian hierarchical models by severity of the casualty

Bayesian hierarchical models	Total traffic casualties					
	Fatal casualties, posterior estimate		Serious injuries, posterior estimate		Slight injuries, posterior estimate	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Traffic characteristics						
ln(Traffic flow (PCU km/h))	1.036**	0.1192	0.0200**	0.0327	0.0361**	0.0368
Average speed (km/h)	-0.0298	0.0122	-0.0163	0.0066	-0.0186	0.0061
Road characteristics						
Length of motorway (km)	0.1778**	0.0499	0.1244**	0.0315	0.1425**	0.0313
Length of A-road (km)	-	-	0.1016**	0.0150	0.0923**	0.0186
Length of B-road (km)	-	-	0.0148	0.0209	-0.0060	0.0200
Length of minor road (km)	-	-	0.0336**	0.0062	0.0298**	0.0075
Socio-demographic factors						
ln(Resident population aged less than 60)	-0.055	0.1539	-0.1317	0.0729	0.0599**	0.0196
ln(Resident population aged 60 or over)	-0.0979	0.1368	-0.0722	0.0667	-0.0388	0.0393
ln(# of employees)	0.0745	0.0605	0.1477**	0.0368	0.1254**	0.0609
ln(# of households with no cars)	0.1747**	0.1556	0.2718**	0.1169	0.0919	0.0579
Constant	-24.9300	2.1110	-6.8460	0.5955	-5.0640	1.0690
S.D. of SC	0.533**	0.1378	1.017**	0.0466	0.9882**	0.0998
S.D. of UH	0.0467**	0.0284	0.0623**	0.0464	0.1598**	0.0876

** Statistically significantly different from zero (95% credible sets show the same sign).

The set of explanatory variables considered in the total serious injury, total slight injury and MT slight injury models are shown in Table 2. Most of the variables were found to be statistically significant in all models. The results of the SEM models are similar to the results of the NB models for the case of the total serious injury model. Area-wide *average speed* was negatively associated with serious injuries but was statistically insignificant in both slight injury models. It was expected that the SEM model which able to take into account the effect of spatial dependence might provide the correct positive sign for variables such as *average speed*. This is not the case for the serious injury model. One explanation would be that ward-level counts for serious injuries are low compared with ward-level counts for slight injuries and therefore, the transformation from a discrete count variable to a continuous variable may not appropriate for the case of serious injuries. This suggests that the preserving of integer structure of the count data is important if the mean of the counts is relatively low.

4.2.2. Bayesian hierarchical model (BHM)

As discussed in the methodology section, count data associated with spatial units can be modelled with a Bayesian hierarchical model that takes account both spatial correlation and unobserved heterogeneity. A series of Bayesian hierarchical models (as shown in Eq. (9)) were estimated for different severity of the casualty (i.e., fatalities, serious injuries, and slight injuries). We are interested in the posterior distributions for the parameters (such as β 's) and the random effects parameters (SC and UH). The posterior estimates (means and standard deviations) of all parameters were obtained using Bayesian inference using Gibb's sampling (known as *winBUGS*) developed by Lunn et al. (2000). The results are presented in Table 6 (for total casualties) and Table 7 (for MT, NMT, VRU casualties). To check the convergence of the MCMC simulations, three chains for the initial values of the parameters were employed. The first 10,000 updates in each chain were removed as burn-ins and a further 50,000 updates were run for each model to obtain a final set of posterior estimates. The variable—*total cars* estimated by Eq. (1) was taken as the exposure variable (EV_i) in all models.

As can be seen in Tables 6 and 7, the posterior means for the standard deviations of spatial correlation (SC) for all types of casualties range from 0.53 to 1.12 and statistically significant in all cases suggesting that observed traffic crashes are spatially correlated among neighbouring wards. The spatial correlation is high in the serious and slight injury models relative to the fatality models. This is consistent with the finding of Agüero-Valverde and Jovanis (2006) who applied a Bayesian hierarchical model to the panel county-level casualty data in Pennsylvania. Similar to the SC, the posterior means for the standard deviations of uncorrelated heterogeneity (UH) are also statistically significant.

Compared to the non-spatial NB models and linear spatial error models presented in Tables 3–5, respectively, almost same sets of explanatory variables were found to be statistically significant in the spatial Bayesian hierarchical models (see Tables 6 and 7). The values of the coefficients are also close to the values found in the NB models. However, the exception is the impact of *average speed* and *resident population aged 60 or over*. Table 6 shows that ward-level *average speed* is now statistically insignificant in all models. Given that traffic speed within London does not vary notably, this is an expected result. Most of the effects due to speed might have captured by other explanatory variables such as *traffic flow* which was found to be statistically significant in most models. The coefficient for *resident population aged 60 or over* was found to be statistically and significantly different from zero in all NB models. However, this variable became statistically insignificant in all Bayesian mod-

Table 7
Estimation results for Bayesian hierarchical models by severity of casualties by road user type

Bayesian hierarchical models	MT casualties			NMT casualties			VRU casualties		
	Slight injuries, posterior estimate		S.D.	Slight injuries, posterior estimate		S.D.	Slight injuries, posterior estimate		S.D.
	Mean	S.D.		Mean	S.D.		Mean	S.D.	
Traffic characteristics									
ln(Traffic flow (PCU km/h))	0.01076	0.0450	0.0214	0.0028	0.0563	0.0242	-0.1152	0.0477	0.0105
Average speed (km/h)	-0.00227	0.0099	0.0070	-0.01383	0.0089	0.0087	-0.01813	0.0075	0.0087
Road characteristics									
Length of motorway (km)	-	-	0.0375	0.0934**	0.0208	0.0188	0.07879	0.0468	0.0496
Length of A-road (km)	0.1301**	0.0200	0.0143	0.0928**	0.0208	0.0158	0.111**	0.0197	0.0179
Length of B-road (km)	-	0.01617	0.0280	0.0708**	0.0328	0.0308	0.05342	0.0283	0.0317
Length of minor road (km)	0.0338**	0.0079	0.0068	0.003825	0.0086	0.0069	0.0134**	0.0080	0.0072
ln(Road curvature)	-0.2663	8.3540	10.5300	-	-	-	-7.912	8.1760	5.7240
ln(# of bus stops)	-	-	-	0.0495**	0.0928	0.1038	0.1146	0.0809	0.0707
Socio-demographic factors									
ln(Resident population aged less than 60)	0.2082**	0.0903	0.0561	-0.05581	0.1061	0.0501	-0.02239	0.0880	0.0568
ln(Resident population aged 60 or over)	-0.1044**	0.0992	0.0459	-0.2362	0.1342	0.1265	-0.2339	0.0744	0.0775
ln(# of employees)	0.1389**	0.0413	0.0216	0.1374**	0.0587	0.0447	0.119**	0.0425	0.0353
ln(# of households with no cars)	0.03138	0.0749	0.0542	0.4934**	0.2174	0.1101	0.3722	0.1040	0.0562
Constant	-7.382	0.9080	0.8566	-9.166	0.7972	8.1	-6.152	0.7976	0.6271
S.D. of SC	1.124**	0.1935	0.1053	0.8328**	0.2446	0.1319	0.9206**	0.1453	0.1236
S.D. of UH	0.3398**	0.1650	0.0442	0.4721**	0.1561	0.0408	0.4205**	0.0955	0.0391

** Statistically significantly different from zero (95% credible sets show the same sign).

els except the MT KSI model in which the variable was found to be negatively associated with the MT KSI. This is consistent with that of the NB model. Therefore, we cannot conclude that the results for this variable are significantly different between NB and Bayesian models.

As in the NB models, the variable—*households with no cars* which is used a proxy for poverty was found to be statistically significant in all models except the models associated with MT casualties. Both the NB and Bayesian models yield the expected positive sign for the coefficient of *number of bus stops* within a ward but this was only found to be statistically significant in the Bayesian models (NMT and VRU casualties). This is an expected result in the sense that bus stops are thought to be crash hot-spots for these groups of road users. Road curvature which is used a proxy for the amount of curved roads was found to be negatively associated with MT serious injuries. In line with the non-spatial models presented in Tables 3 and 4, average ward-level road curvature was found to be statistically insignificant in all models suggesting that wards with many road bends do not suffer more crashes relative to wards with many straighter roads in London. This maybe due to the fact that there is not enough variation in *bend density* within the wards of London metropolitan area as Haynes et al. (2007) indicated that road network within a metropolitan area is straighter than that of a non-metropolitan area.

In summary, the results from the Bayesian hierarchical models are quite similar to the non-spatial NB models in many cases. The most significant difference between NB models and Bayesian hierarchical models for the same datasets was the result for *average traffic speed*. While employing NB models, this variable was found to be statistically significant and negatively associated with all casualties (except MT KSI). In the Bayesian hierarchical models, however, this variable was consistently found to be statistically insignificant in all models. It is believed that NB models provided the wrong sign in the coefficient of average speed. As discussed, this could be due to the spatial dependence in the crash data among neighbouring wards or the lack of a proper proxy variable such as ward-level VMT. The suspicion may fall to the former, as average speed was statistically insignificant in the NB model for MT KSI in which the residuals for the estimated model were not spatially correlated. Moreover, some of the variables which were found to be statistically insignificant in the NB models became significant in the Bayesian models with the same signs. Intuitively, it can be said that the results from the Bayesian models are more consistent with the literature and also more logical compared with the results from NB models or SEM models (in the case for MT KSI). This may be due to the fact that Bayesian models are able to accurately take into account the spatial correlation and uncorrelated heterogeneity normally present in aggregated area-level crash count data.

5. Conclusions

This paper has developed a series of relationships between different ward-level traffic casualties and ward characteristics using both non-spatial NB models and classical spatial models and Bayesian hierarchical models. The statistical test—global Moran's *I* indicates that ward-level traffic occurrences exhibit spatial autocorrelation (i.e., spatial dependence) and this has subsequently been confirmed by the residuals (which have found to be spatially correlated) from the NB models and the posterior means for the standard deviations of spatial correlation in the Bayesian hierarchical models. This implies that non-spatial models may not be appropriate for such data. However, the results from the NB models and the Bayesian hierarchical models are quite similar in many cases.

The classical spatial models also provided results that are consistent with the Bayesian models except for the case of serious injuries where the mean of the counts is relatively low. The results from the Bayesian hierarchical models are more consistent with the literature and more coherent in all cases. It has been confirmed from this analysis that different area-level factors affect various categories and severity of the casualty differently. This finding is in-line with the findings by other researchers (e.g., Kim et al., 2006; Agüero-Valverde and Jovanis, 2006). Traffic flow has a positive association with casualties. The variables households with no cars and total employment are statistically significant in almost all models. It is surprising that ward-level average traffic speed has no impact on fatalities and serious injuries. A possible reason is that free-flow maximum speeds in London generally do not exceed about 40 mph (65 km/h) and the highest average speed in the data is about 38 km/h.

From the results found in paper, it can be suggested that the parameter estimates of crash prediction models associated with spatial units using NB models would be appropriate if the residuals from such models are not spatially correlated. It is envisaged that spatial dependence among spatial units would decrease at higher levels of spatial aggregation such as regions, counties and states. At a lower levels of spatial aggregation such as wards, enumeration districts postcode sectors, and super output areas (SOA), the presence of spatial correlation may be more common. This is certainly an active area for further research.

In line with the findings of Miaou et al. (2003) and Agüero-Valverde and Jovanis (2006), Bayesian hierarchical models have found to be an appropriate model to analysis area-wide traffic crash occurrences. However, these types of fully parametric models need to invoke some form of strong distributional assumption *a priori*. In non-experimental settings, such as the occurrences of traffic crashes, it is difficult to assume *a priori*. Hence, model estimates and inferences derived from them can be sensitive to distributional assumptions. Bhati (2005) suggests the use of an alternative semi-parametric method – the cross-entropy (CE) method – that avoids parametric distributional assumptions. It would be interesting to see how the results might change when such a semi-parametric model is applied to the data. There are a number of other ways to extend the spatial analysis used in this study. One way is to conduct such a spatial analysis for both larger (i.e., boroughs of London) and smaller (i.e., super output areas of London) scale area-aggregations and their effects on the analysis. Another is to extend this spatial analysis for the whole of Great Britain. This may address some of the confounding results of this study.

There are, however, some limitations in the data used in our analysis. For instance, STATS19 national road crash data recorded by the British Police may suffer from under-reporting, especially for the case of slight injury crashes. The actual location (*x* and *y* coordinates) of a crash reported in STATS19 database may also contain measurement errors. This may result an error in area-wide traffic crash counts. In addition, ward-level traffic data such as traffic speed and flow used in this analysis were estimated by Glaister and Graham (2006) from road-level disaggregated data rather than directly measured them at ward-level. This might had an impact on the results.

References

- Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. *Accid. Anal. Prev.* 32 (5), 633–642.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.
- Agüero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accid. Anal. Prev.* 38 (3), 618–625.

- Amoros, E., Martin, J.L., Laumon, B., 2003. Comparison of road crashes incidence and severity between some French counties. *Accid. Anal. Prev.* 35 (4), 537–547.
- Baller, R.D., Anselin, L., Messner, S.F., Deane, G., Hawkins, D., 2001. Structural covariates of U.S. County homicide rates: Incorporating spatial effects. *Criminology* 39 (3), 561–590.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc., Ser. B* 36 (2), 192–236.
- Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* 43, 1–59.
- Best, N.G., Ickstadt, K., Wolpert, R.L., 2000. Spatial Poisson regression for health and exposure data measured at disparate resolutions. *J. Am. Stat. Assoc.* 95, 1076–1088.
- Bhati, A.S., 2005. Modeling count outcomes with spatial structures: An information-theoretic approach, Discussion Paper, Justice Policy Center, The Urban Institute, Washington DC.
- Brooks, S.P., 1998. Markov chain Monte Carlo method and its application. *The Statistician* 47, 69–100.
- Cameron, A.C., Trivedi, P.K., 1998. *Regression Analysis of Count Data*. Cambridge University Press.
- Carlin, B.P., Louis, T.A., 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London, UK.
- Gelman, A., Carlin, J.C., Stern, H., Rubin, D.B., 1995. *Bayesian Data Analysis*. Chapman and Hall, New York.
- Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica* 52, 701–720.
- Glaister, S., Graham, D.J., 2006. Spatial implication of transport pricing. *J. Trans. Econ. Policy* 40 (2), 173–201.
- Graham, D.J., Glaister, S., 2003. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. *Urban Stud.* 40 (8), 1591–1607.
- Gujarati, D.N., 2003. *Basic Econometrics*, 4th edition. McGraw Hill, Boston.
- Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2003. Microlevel accident prediction models for evaluating safety of urban transportation systems. *Transport. Res. Rec.* 1840, 87–95.
- Hausman, J.C., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents–R&D relationship. *Econometrica* 52 (4), 909–938.
- Haynes, R., Jones, A., Kennedy, V., Harvey, I., Jewell, T., 2007. District variations in road curvature in England and Wales and their association with road-traffic crashes. *Environ. Plan. A* 39 (5), 1222–1237.
- Kim, K., Brunner, I.M., Yamashita, E.Y., 2006. The influence of land use, population, employment and economic activity on Accidents. *Transport. Res. Rec.: J. Transport. Res. Board* 1953, 56–64, TRB, National Research Council, Washington DC.
- Kim, K., Yamashita, E.Y., 2002. Motor vehicle crashes and land use: empirical analysis from Hawaii. *Transport. Res. Rec.: J. Transport. Res. Board* 1784, 73–79, TRB, National Research Council, Washington DC.
- Ivan, J.N., Wang, C., Bernardo, N.R., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. *Acc. Anal. Prev.* 32 (6), 787–795.
- Jovanis, P., Chang, H.L., 1986. Modeling the relationship of Accidents to miles traveled. *Transport. Res. Rec.* 1068, 42–51.
- Ladron de Guevara, F., Washington, S., Oh, J., 2004. Forecasting Crashes at Planning Level Simultaneous Negative Binomial Crash Models Applied in Tucson, Arizona. *Transport. Res. Rec.: J. Transport. Res. Board* 1897, 191–199, TRB, National Research Council, Washington DC.
- LeSage, J.P., 1998. Spatial econometric, working paper, Department of Economics, University of Toledo.
- Levine, N., Kim, K., Nitz, L., 1995a. Spatial analysis of Honolulu motor vehicle crashes. I. Spatial patterns. *Acc. Anal. Prev.* 27 (5), 663–674.
- Levine, N., Kim, K., Nitz, L., 1995b. Spatial analysis of Honolulu motor vehicle crashes. II. Zonal generators. *Acc. Anal. Prev.* 27 (5), 675–685.
- Li, L., Zhu, L., Sui, D.Z., 2007. A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *J. Transport Geogr.* 15 (4), 274–285.
- Lord, D., 2000. The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models. Ph.D. Dissertation, Department of Civil Engineering, University of Toronto, Toronto.
- Lord, D., van Schalkwyk, I., Chrysler, S., Staplin, L., 2007. A strategy to reduce older driver injuries at intersections using more accommodating roundabout design practices. *Acc. Anal. Prev.* 39 (3), 427–432.
- Lord, D., Persaud, B., 2000. Crash prediction models with and without trend: application of the generalized estimating equations (GEE) procedure. *Transport. Res. Rec.: J. Transport. Res. Board* 1717, 102–108, TRB, National Research Council, Washington DC.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* 10, 325–337.
- MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area accident and injury analysis. *Acc. Anal. Prev.* 36 (6), 1028–1091.
- Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Acc. Anal. Prev.* 26 (4), 471–482.
- Miaou, S., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. *J. Transport. Stat.* 6 (1), 33–57.
- Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. *Acc. Anal. Prev.* 39 (3), 468–469.
- Mollie, A., 1996. Bayesian mapping of disease. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, pp. 359–379.
- Messner, S.F., Anselin, L., Baller, R.D., Hawkins, D., Deane, G., Tolnay, S., 1999. The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *J. Quant. Criminol.* 15 (4), 423–450.
- Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. *Acc. Anal. Prev.* 36 (4), 525–532.
- Noland, R.B., Quddus, M.A., 2004a. A spatially disaggregate analysis of road casualties in England. *Acc. Anal. Prev.* 36 (6), 973–984.
- Noland, R.B., Quddus, M.A., 2004b. Improvements in medical care and technology and reductions in traffic-related fatalities in Great Britain. *Acc. Anal. Prev.* 36 (1), 103–113.
- Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Acc. Anal. Prev.* 27 (3), 371–389.
- Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Acc. Anal. Prev.* 38 (6), 1137–1150.
- Wakefield, J.C., Best, N.G., Waller, L., 2000. *Bayesian Approaches to Disease Mapping, on Spatial Epidemiology: Methods and Applications*. Oxford University Press.
- Wolpert, R.L., Ickstadt, K., 1998. Poisson/Gamma random field models for spatial statistics. *Biometrika* 85, 251–267.