2017-08-28

Analysis of Traffic Collisions in Dublin Using GIS Based System

Jagteshwar Singh
Technological University Dublin, jagteshwar.shan@gmail.com

Niamh O' Reilly
Technological University of Dublin, niamh.oreilly@tudublin.ie

Recommended Citation

Singh, J. & O'Reilly N. (2017). Analysis of traffic collisions in Dublin using GIS based system. *ITRN 2017 conference,* 28th & 29th August 2017, University College Dublin. doi:10.21427/km7e-w124

ANALYSIS OF TRAFFIC COLLISIONS IN DUBLIN USING GIS BASED SYSTEM

Jagteshwar Singh
MSc. Geographic Information Science Candidate
Dublin Institute of Technology

Niamh O Reilly
Programme Chair MSc Geographic Information Science
Dublin Institute of Technology

Abstract

This study aims to analyse traffic collisions in the Greater Dublin Region between the period 2006-2012, using GIS to identify hotspots and examine the relationship between collisions and a range of contributory factors including vehicular speeds, traffic volume, road curvature, road category and distance from intersection that could enable prediction of traffic collisions. To this end, Road Safety Authority (RSA) collision data for Dublin Region geocoded as point events, road profiles, traffic flow characteristics on which these occur are spatially merged using ArcGIS and FME software to establish if a significant relationship exists between collision frequencies on road links and the specific link characteristics and traffic flow characteristics.
The road network has been divided into uniform segments and the collision frequencies on each of these noted. Traffic collisions are rare and random events and often a major proportion of segments would have no instance of collisions, thus following a Negative Binomial distribution. The outputs from GIS exercise are tested through SPSS software using Negative Binomial distribution for modelling the relationship between different variables.

This paper comes at a significant time where efforts are being made to improve the safety of roads within the European Union [1]. Every year, road collisions cause human fatalities together with huge financial loss which can be significantly reduced by improving road safety through the enforcement of traffic laws and road user compliance. By identifying the cause effect relationship and the spatial locations most prone to collisions, prioritized regulatory and safety interventions can be put in place to reduce the collisions on the roads.

1. Introduction

In Ireland, since the introduction of the first Government road safety strategic in 1998 there has been a significant decrease in road fatalities. Since its implementation, road fatalities have decreased by over 60% from 1997-2011 [2]. The Government Road Safety Strategy 2013-2020 aims to reduce the numbers of fatalities and serious injuries on roads to close existing gap between Ireland and the safest countries in the world. This strategy adopts a 'safe systems' approach which recognises that road users are prone to making mistakes but through efficient road design, traffic control and road-user awareness, the adverse collision outcomes could be converted into minor ones. A key input to the safe system is an understanding and evaluation of where and when collisions occur and the risks associated with them. By understanding and predicting such risks, interventions can be implemented to reduce them.

The objective of this study is to identify sections of Dublin road network where collisions tend to cluster and to examine the relationship between occurrence of traffic collisions with road and traffic characteristics including vehicular volume, average road speed, road curvature, road category and distance from intersection. It aims at making inferences based on a large number of traffic volume data and a small number of road collisions. The data regarding traffic characteristics is derived from the National Transport Authority (NTA) Regional Transport Model and is representative of average speeds and volumes over one year.

Dublin city has a low density urban sprawl with a complex web of road network with the M50 motorway circling around the city, several main arteries that connect the busy hubs, together with minor and feeder roads that circulate through residential areas. Table 1 gives an estimate of the length of various road types in Dublin Region.

Road Class	Description	Length in Km
1	National Motorways, Intra-city Roads, Bypass Roads (M50, M1, M4, N11, etc.)	360
2	Major Arterials	787
3	Minor arterials	2194
4	Neighbourhood Roads, Cul-de-sacs, dead-ends	2072

Table 1- Road network in Dublin region

The study area covers Dublin region which consists of Dublin City, South Dublin, Dún Laoghaire–Rathdown and Fingal, an area of 925 km^2. This gives Dublin an overall road density of 5.85 km per Square km, with densities varying considerably from highest in city centre to lowest in suburbs and rural regions. Between the years 2006-2012, a total of 9578 road collisions were recorded by Garda in the Dublin. For analysis purposes, traffic collisions have been categorised into two groups: Severe collisions which resulted in either fatalities or serious injuries, and minor collisions which resulted in minor injuries. The road network in this study has been sub-divided into segments of 100m each. Next, collision counts for each road segment are modelled as binomial distribution using a list of explanatory variables to establish the variables related in a significant way to the collisions.

2. Literature Review

Collisions are point events and tend to cluster around at specific regions called ‘hotspots’. Kernel Density Estimation (KDE) is a very appropriate spatial technique that produces a density function of point events over a region to estimate the clustering phenomenon. KDE along with spatio-temporal clustering has been successfully employed in many studies previously for hotspot identification and traffic collision studies [3]. A hotspot analysis identifies statistically significant spatial clusters of high values(hot-spots) and low values(cold-spots).

The variable to be utilized as a measure of collision, or the “collision index” for a site in various studies has been typically define as one of the following [4].

1- **Crash frequency:** Number of times a crash occurs on a given segment of road. The time period may be annual, seasonal, or monthly.
2- **Crash rate:** is the crash frequency normalized by a variable denoting the exposure of the site to traffic. Generally, traffic volume or Vehicle Kilometers travelled (VKT) is used as an exposure factor. The crash rate takes into account the fact that busier roads will observe more collisions.
3- **Equivalent Property Damage (EPO):** This method puts a price tag on each collision in terms of the economic losses caused by damage to vehicles/infrastructure, loss of life and the pain and bereavement caused to family members, medical expenses, administrative expenses, and loss of productivity.

A study by Montella in 2010 tested the effectiveness of use of various collision index parameters. The Crash frequency was found to be a better parameter than Crash rate in hotspot analysis [5] which raises questions on the widespread use of Crash Rate in favour of frequency. The reasons for this incoherency identified by the author were that the crash rate is biased towards low-volume sites and it incorrectly assumes a linear relationship between crash frequency and traffic volume which is found to be incorrect in many studies.

Another important aspect of the collision studies is the modelling of collisions enabling prediction of crash rate. The collision models generally tend to explain a dependent variable (collision frequency/ collision rate) in terms of explanatory factors. These factors can be broadly classified into the following categories [6,7] :

1- Human factors: Drunk driving, excessive speeding/ rash driver behaviour etc.
2- Traffic characteristics: Traffic speeds, modal split, density, Passenger Car Unit (PCU) volume, peak hour ratio etc.
3- Infrastructure: Road characteristics, functional class, lane width, curvature, junction control, adjoining landuse, socio-economic conditions.
4- Environmental factors: Weather, snow/dry/rain, visibility, etc.
5- Vehicle related factors: Type of vehicle, condition of tyres, brakes, vehicle size, etc.

Traffic collisions are by nature random and rare events that occur on the road network and they can be modelled using mathematical functions that allow for random variations in an observation sample. Collision data generally tends to overdisperse and as such the assumptions of Poisson distribution don't hold true [8]. Therefore, Negative Binomial distribution, which explains easily the overdispersion phenomenon has been successfully employed and is recommended for the study of collision frequency data [9].

Taking into consideration the above recommendations and techniques employed by previous studies, this study uses Negative Binomial Distribution to model crash frequency as a function of a list of explanatory variables extracted from the existing datasets.

3. Data Preparation and Methodology

Traffic collision dataset from the year 2006 to 2012 was obtained from the Road Safety Authority. RSA collision data is preliminary collision information recorded by a Gardai visiting the scene of incident at the time of its occurrence. Data recorded can contain information regarding location of the collision, behaviour of those involved, driver action, contributory action, the vehicle itself and the extent of the casualty, that is, fatality, serious or minor injury. Locational information for each collision is normally recorded by the Garda using a hand-held GPS receiver, the accuracy of which is between 10 to 15 metres. Thus, for the present study, a tolerance of 15m was adopted while identifying the collisions with the roads on which they occur. It should be noted, that an assumption has been made that the geographic location of collisions have been recorded using a handheld GPS receiver, although this may not always be the case; an example being where a collision is reported after the event at the Gardai station and its location is then identified on an Ordinance map or through verbal description, in such cases the accuracy may be outside the above tolerance.

The National Transport Authority of Ireland maintains five regional transport models covering the Republic of Ireland. The model is generated using 4-stage trip generation and assignment model and utilizes inputs from National Household travel survey, CSO census and Traffic count data. The model for Dublin region for the base year 2012 was used in this study. It contains traffic flow information like peak hour traffic volumes, congested and free flow speeds, link capacities etc. having a coverage for most of the road links in Dublin. Along with this, the actual road network with attributes like functional classification, carriageway type, street names etc. was obtained from the Navteq street maps. The mapping is done by the Navteq company driving through all the public roads in Ireland and recording geometrical and physical attributes of the roads.

S.No.	*Dataset description*	*Year*	*Attributes used in study*
1	RSA collision dataset	2006-12	Location, collision severity, time, collision type
2	NTA traffic flow model	2012	Traffic volume, speed
3	Navteq street map	2012	Road geometry, functional class

Table 2- Data sources and attributes used in study

Using FME and ArcGIS 10.4.1 software the RSA Traffic collision dataset and NTA traffic flow model were merged into the Navteq road network. The NTA model represents roads links as straight lines connecting the nodes whereas Navteq street map contains actual road geometry which deviates from assumed straight lines in the model and to merge them an elaborate workflow was adopted. After merging the information into road network, the roads

were divided into uniform segments of 100m length. The collision frequencies on each of these segments for the respective collision category (severe and minor) was found. For each segment, a measure of exposure to traffic 'Vehicle Kilometres Travelled' (VKT) was calculated as the product of link length and the traffic volume on that link. Sinuosity of the segments (ratio of actual road length to the shortest path defined by start and end nodes) was calculated which measures the degree of bendiness of a road. The distance of the segments from road junction was also determined to see if it has any bearing on the collisions.
Some GIS mapping and variable relationships were found using the derived data and is presented in the next section.

4. Analysis

A comparison was made between the collisions occurring on road bends/roundabouts and their time of occurrence (Fig 1). The bendy roads and roundabouts were detected using sinuosity values greater than 1.05 and from the Navteq data attributes.

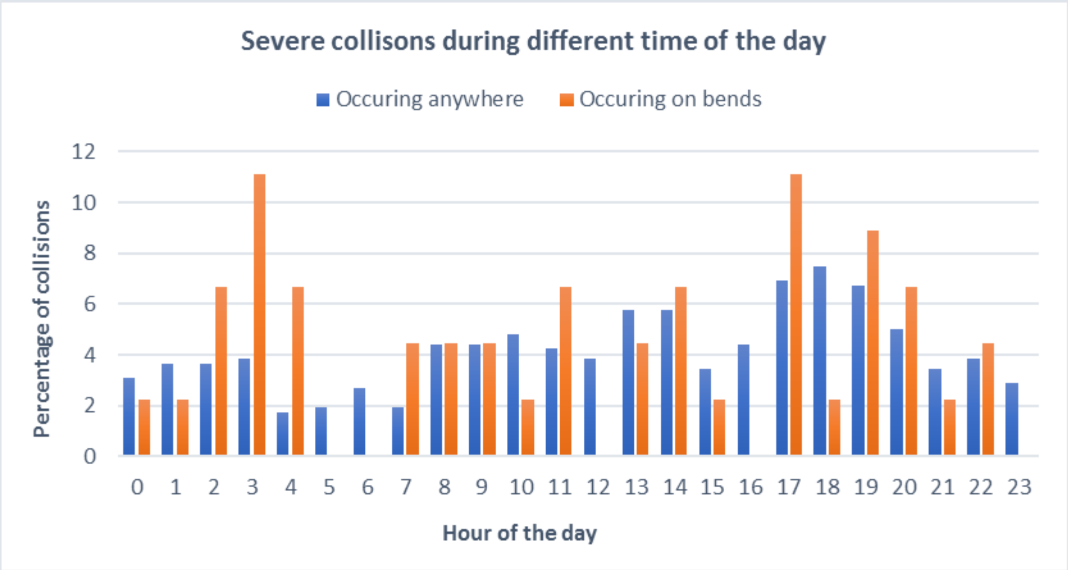

Figure 1- Severe collision hourly distribution all road sections and on bends

The figure reveals that on bends, severe collisions are most common during the night, especially from 1-5 am. About 26.7% of severe collisions occurring on bends occur during the time 1-5 am whereas for all the severe collisions taken collectively, only 12.5% for them occur from 1-5 am. When compared to minor collisions, only about 4% of the minor collisions occur from 1-5 am whether on bends or on straight stretches of road. These figures implicate two things: first, the chances of severe colllision occuring almost triple during the hours 1-5 am as compared to minor collisions. Second, the severe collisions during these hours are twice as more probable on road bends.

Single vehicle collsions (SVC) were noted to become quite higher during th night-time 11pm-7am on bendy road sections. The table below shows the variation of Single Vehicle collisions as a fraction of overall collisions during different times of day and on different road sections.

Time-Window	Severe collisions		Minor collisons	
	Bendy sections (% SVC)	All sections (% SVC)	Bendy sections (% SVC)	All sections (% SVC)
Night time 11pm- 7am	75%	31.96%	44.58%	22.84%
Whole day	26.67%	14.39%	16.22%	9.54%

Table 3- Percentage Single Vehicle collisions for different times of day and road bendiness

The results indicate single vehicle collisions are exacerbated under the critical combination of two conditions: night-time and sharp road curvature. It might suggest the prevalence of drunk driving, driver fatigue or overspeeding during night time leading of loss of control at road bends.

A hotspot analysis was carried out using the severe category of traffic collisions to identify places and times where such events most frequently occur. The time is referred to on a scale from 0 hour to 24 hours to find out if collisions are more likely during certain times of the day (rush hours of morning/evening, night-hours, etc.). A time-window of three hours and search bandwidth of 500m was adopted for finding the clusters i.e. any collisions occurring within 3 hours for a typical day and within 500m of each other were assumed to form clusters. The assumption of 3 hours fits in well with the defined rush hour period of 7-10am and 4-7pm. The speed limit of the road on which the collisions occur was used in the hotspot analysis along with the clustering method defined above. The output of the exercise identified the clustered collisions occurring on high-speed roads defined as the ‘hot-spots’ and on low-speed roads defined as ‘cold-spots’.

Many collisions in Dublin city centre were detected as cold-spot and the clusters on major motorways were reported as hot-spots. The statistically significant hot and cold spots (i.e. collisions which are surrounded by significant number of similar collisions in both space and time) are discussed below:

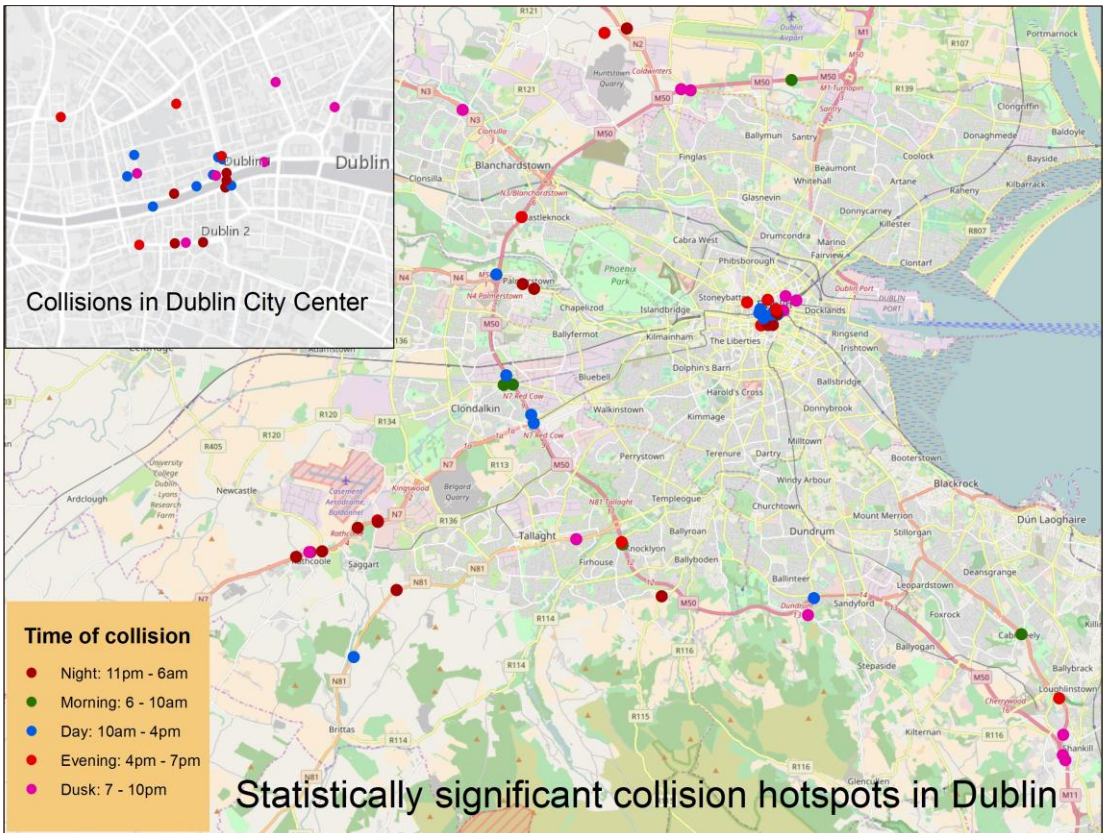

Figure 2- Hotspot analysis- statistically significant clusters in Dublin

A total of 693 severe collisions occurred in Dublin from 2006-12 out of which only 72 were found to show significant amount of clustering in space-time. Of these, the ones occurring on Naas road near Saggart especially where R120 joins N7 motorway occur during the early hours of the morning (1-4 am). Similarly, near Bray at the point of confluence of M50 and M11, collisions occur between 8-11 pm. Additionally, the analysis found significant clustering of collisions near Clondalkin on the M50 motorway between 9-12 am. Within the Dublin city centre, a majority of collisions occur during busy daytime hours (2-7 pm). The when and where of a collision is a key information enabling strategic interventions and planning to make the roads safer.

Speed plays a cticial role in road collisions. A comparison of the severe collision rates (severe collisions per VKT) with the vehicular speeds obtiained from NTA model shows that the collision rates are highest during the day (7am-7pm) in the speed range of 40-50kmph whereas at night (7pm-7am) they are highest at speeds of 55-60kmph and also at 80kmph. It can be observed from the figure that collisions rates and thus the likelihood of collisions duirng night increases drastically at speeds beyond 55kmph which is not observed during the day.

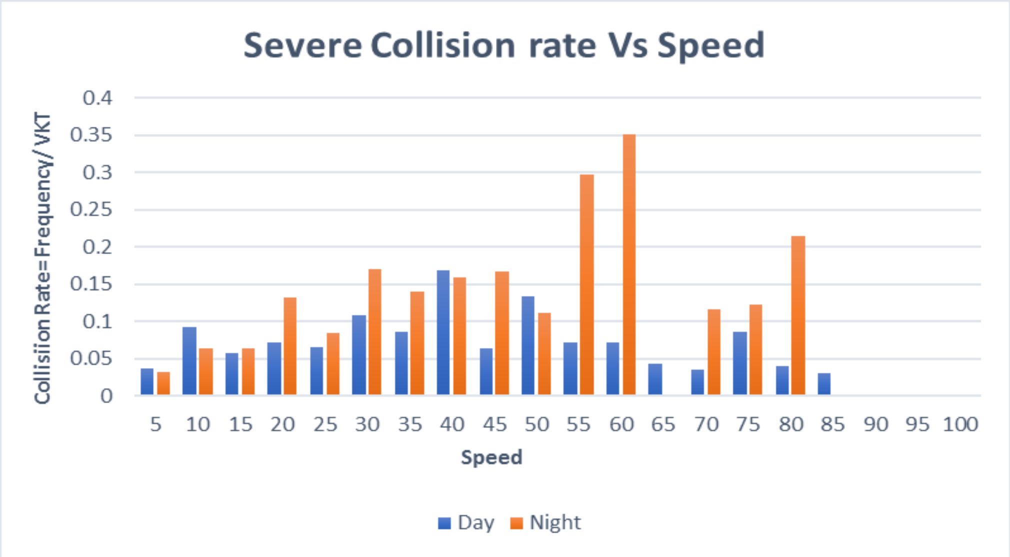

Figure 3- Collision rate vs Speed- Severe collisions

Also, the collisions were analysed with regards to distance from intersections (Table 4). It shows that collisions are predisposed to occur near the intersection, with high values in the range 0-15m. However, single vehicle collisions are more likely to occur away from the intersection in the distance range 50-100 and beyond. The findings suggest that 37.5% of rear-end severe collisions occur at a distance of between 15 and 30m from the intersection: this distance range forms the Stopping Sight Distance (SSD) and is the part where vehicles decelerate and apply brakes to stop or merge with the other traffic and can be more prone to rear-end collisions. The majority of head-on collisions occur on the intersection. The majority of pedestrian collisions occur at the intersection itself as is expected since the pedestrian crossings occur mostly on the intersections and junctions. The figures also suggest that severe pedestrian collisions are more common away from the intersection when compared to minor pedestrian collisions, especially in the distance range of 30-50m.

The minor collisions show a slightly different pattern than severe collisions, the minor ones occur less sparingly away from the junction as opposed to the severe collisions which show an increase in other distance ranges.

Collision Severity	**Collision type**	**Distance from Intersection**						
		0-15m	***15-30m***	***30-50m***	***50-100m***	***100-300m***	***>300m***	***TOTAL***
Severe Collision	Pedestrian	38.7%	16.9%	20.1%	16.5%	7.4%	0.35%	100%
	Single vehicle	21.4%	15.1%	10.3%	28.6%	22.2%	2.38%	100%
	Head-on	34%	12.7%	10.6%	25.5%	12.8%	4.25%	100%
	Rear-end	15.6%	37.5%	12.5%	9.4%	25%	0%	100%
Minor Collision	Pedestrian	40.4%	22.7%	16.1%	15.4%	5.1%	0.24%	100%
	Single vehicle	23.4%	17.8%	16.7%	19.9%	19%	3.18%	100%
	Head-on	41.6%	14.1%	11.7%	15.3%	14.6%	2.68%	100%

	Rear-end	36.2%	17.9%	16.4%	18.2%	10.6%	0.65%	100%

Table 4- Relation between type of collision and distance from intersection

5. Negative Binomial Distribution Model

The severe and minor collision frequency on a segment was modelled as a dependent variable with the traffic volume, speed, road functional class, curvature, distance from intersection and segment length as dependent variables using Negative Binomial (NB) distribution. The relation takes the form:

$$\log(A) = I + (\beta_1 \times DistCat1 + \beta_2 \times DistCat2 + ..) + (\beta_6 \times Rd1 + \beta_7 \times Rd2 + ..) + \beta_n \times SegLen + \beta_m \times Speed + \beta_o \times PCU + \beta_p \times Sinuosity$$

'A' is collision frequency on a segment, DistCat1 is a Boolean variable and is 1 if the segment is within 0-15m from the intersection, otherwise 0. Similarly, Distance categories 2,3,4 are defined by distance range of 15-50m, 50-100m, 100-300m and greater than 300m from the intersection respectively. Rd1, Rd2, Rd3, Rd4 are also Boolean variables to check for the road type. The other variables in the equation are segment length, traffic volume expressed as Passenger Car Unit (PCU) per hour, traffic speed and road sinuosity.
After being tested in SPSS software using NB regression, the following results were obtained

Parameter	Severe collisions			Minor collisions		
	Coeff. β	Std. error	Sig.	Coeff. β	Std. error	Sig.
Intercept I	.437	1.9004	.818	4.252	.6157	<.01
Segment Length	*.005*	*.0006*	*<.01*	*.005*	*.0002*	*<.01*
Speed	-.012	.0024	<.01	-.016	.0009	<.01
PCU per hour	.002	.0017	.163	.007	.0011	<.01
Segment Sinuosity	-3.628	1.8931	.055	-4.487	.6133	<.01
Road Class 1 (base category-bs)	0	-	-	0	-	-
Road Class 2	.563	.1565	<.01	.124	.0614	.044
Road Class 3	-.141	.1399	.312	-.227	.0493	<.01
Road Class 4	-.429	.1485	.004	-.489	.0512	<.01
Junction distance: 0-15m (bs)	0	-	-	0	-	-
Junction distance: 15-50m	-.501	.4618	.278	-.704	.1556	<.01
Junction distance: 50-100m	-.436	.3027	.150	-.553	.1045	<.01
Junction distance: 100-300m	-.392	.1733	.024	-.701	.0675	<.01
Junction distance: >300m	-.896	.3184	.005	-1.205	.1159	<.01

Table 5- NB Distribution Model results

The results of the model test are discussed next with the statistically significant variables found to have an influence on the collision frequency.

6. Results and discussion

The analysis shows that segment length and traffic speed as statistically significant parameters, with an increase in either causing increase in collisions. For severe collisions, the traffic volume doesn't have any significant relation with collision frequency. Whereas for minor collisions, greater the traffic volume, greater the number of crashes. The segment sinuosity also contributes towards the collision frequency marginally, and this contribution is higher in the case of severe collisions. Roads having higher functional class (motorways, national primaries) contribute towards the collision frequencies more strongly than the lower class roads. In terms of distance from intersection, road segments nearer to or on the intersections have higher contribution towards collisions. But this relation is not statistically significant in case of severe collisions. The analysis was repeated using a Poisson lognormal distribution and it achieved a less satisfactory goodness-of-fit. This validates the suitability of NB regression for collision count data.

7. Study limitations and future recommendations

There are a number of limitations with the research that should be noted: vehicle speed used for the research is the modelled vehicle speed and not the actual speed at the time of collision. Hence the relationships established are not for the actual vehicle speed but the average road speed. The traffic volume derived from NTA model is the average value for a year. However, it is important to note that it may vary considerably by time of the day, day of the week or month of the year. For the hotspot analysis, a uniform clustering tolerance was used for the whole Dublin region. However, future research would benefit from adopting different clustering tolerance for urban and rural regions because of the differences in road layouts and collision density. Collisions are much more common in urban regions of Dublin. The same hotspot analysis for minor collisions would also be quite insightful.

For the developed model, some independent variables used can be inter-related. For example, speed and PCU have a well-known parabolic relationship. These aspects along with other variables like modal split of traffic, lane widths, land-use activity surrounding the road, etc. should be studied in future research.

ACKNOWLEDGEMENTS

The author kindly acknowledges the Road Safety Authority for providing the historical collision dataset for use and which upon this research was based. The author is also highly grateful to NTA for providing Navteq street network and regional transport model for Dublin region and to Peter O’ Sullivan from NTA for his constant help and guidance throughout the research work.

REFERENCES

[1] European parliament and council of European Union, 2008. Directive 2008/96/EC on road infrastructure safety management

[2] Dublin City Council, 2016. Road Safety Strategy 2020. Environment & Transportation Department, Wood Quay, Dublin 8.

[3] Erdogan, S., Yilmaz, I., Baybura, T., Gullu, M., 2008. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. Accid. Anal. Prev. 40, 174–181. doi:10.1016/j.aap.2007.05.004

[4] PIARC, World Road Association, Technical Committee on Road Safety C13, 2004. Road Safety Manual

[5] Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42, 571–581. doi:10.1016/j.aap.2009.09.025

[6] 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. Transp. Geogr. 15, 274–285. doi:10.1016/j.jtrangeo.2006.08.005

[7] Vandenbulcke-Plasschaert, G., 2011. Spatial analysis of bicycle use and accident risks for cyclists. Presses univ. de Louvain.

[8] Shankar, V., Mannering, F., Barfield, W. 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention 27 (3), 542–555.

[9] Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part Policy Pract. 44, 291–305. doi:10.1016/j.tra.2010.02.001