Exploring crash-risk factors using Bayes' theorem and an optimization routine

Dr Maria-Ioanna M. Imprialou
Lecturer
School of Civil and Building Engineering
Loughborough University
Loughborough LE11 3TU
United Kingdom
Tel: +44(0)1509 228545
E-mail : M.Imprialou@lboro.ac.uk

Professor Mike Maher
Honorary Professor
Dept of Civil, Environmental & Geomatic Engineering
University College London
London WC1E 6BT
United Kingdom
E-mail: m.maher@ucl.ac.uk

Professor Mohammed Quddus*
Professor of Intelligent Transport Systems
School of Civil and Building Engineering
Loughborough University
Loughborough LE11 3TU
United Kingdom
Tel: +44(0)1509 228545
E-mail : M.A.Quddus@lboro.ac.uk

* *Corresponding author*

Word Count: 5,544 + 5 Figures*250+ 1 Table*250 = 7,044
Paper submitted for presentation at the 95th Annual Meeting of the Transportation Research Board (TRB) in Washington D.C., USA

Submission date: *31 July 2015*

1
2

ABSTRACT

Regression models used to analyse crash counts are associated with some kinds of data aggregation (either spatial, or temporal or both) that may result in inconsistent or incorrect outcomes. This paper introduces a new non-regression approach for analysing risk factors affecting crash counts without aggregating crashes. The method is an application of the Bayes' Theorem that enables to compare the distribution of the prevailing traffic conditions on a road network (i.e. *a priori*) with the distribution of traffic conditions just before crashes (i.e. *a posteriori*). By making use of Bayes' Theorem, the probability densities of continuous explanatory variables are estimated using kernel density estimation and a posterior log likelihood is maximised by an optimisation routine (Maximum Likelihood Estimation). The method then estimates the parameters that define the crash risk that is associated with each of the examined crash contributory factors. Both simulated and real-world data were employed to demonstrate and validate the developed theory in which, for example, two explanatory traffic variables speed and volume were employed. Posterior kernel densities of speed and volume at the location and time of crashes have found to be different that prior kernel densities of the same variables. The findings are logical as higher traffic volumes increase the risk of all crashes independently of collision type, severity and time of occurrence. Higher speeds were found to decrease the risk of multiple-vehicle crashes at peak-times and not to affect significantly multiple-vehicle crash occurrences during off-peak times. However, the risk of single vehicle crashes always increases while speed increases.

Keywords: Bayes' theorem, maximum likelihood estimation, speed, volume, crashes.

INTRODUCTION

The development of successful road safety measures relies on the in-depth understanding of the factors and the mechanisms that are related to crash occurrences. There are numerous factors associated with driving attitudes, traffic or external conditions and often combinations of these factors that lead to crashes *(1, 2)*. Since the first systematic crash analyses that emerged approximately seven decades ago, there have been constant and significant advances on our understanding of crash occurrences *(3)*. The randomness and the complexity that characterise road crashes have not yet permitted a full explanation of these phenomena.

A large proportion of the current knowledge on crash contributory factors comes from crash count models that attempt to identify and explain crash precursors. Although the statistical approaches of count models are increasingly more accurate and integrated *(4)*, crash models often provide inconsistent results even for the most significant traffic variables such as speed and traffic volume. The assumption that a relationship that applies at a group level (link-level) necessarily applies to an individual level (crash-level), termed as aggregation bias, is likely to be a source of inconsistent and erroneous results in crash analyses *(5, 6)*. Count models incorporate by definition this characteristic: crashes are aggregated (typically with spatial criteria) and are represented in models by characteristic values of the examined variables that usually reflect the most typical situation on the roadway. In this way, the actual circumstances of when crashes occurred, which are likely to be unusual, are not necessarily represented in the models *(7, 8)*.

To define accurately the role of potential crash contributory factors it is important to express the actual traffic conditions that are directly related with the crashes in the models. This paper aims to set out an approach to the modelling of collision data where the objective is to estimate the effect of traffic conditions at the time of the crash without aggregating crashes. The data then consist of the traffic conditions applying at or around the time and location of each crash. These conditions are therefore a biased sample of the conditions that generally prevail on the network, and this statistical conditionality needs to be accounted for in any modelling of the crash dataset. Hence the method to be employed is an application of Bayes' theorem for the case of a continuous variable aided by the maximum likelihood estimation algorithm. The prior is the distribution of a variable of interest (e.g. speed) on the network and the likelihood is the crash risk associated with the variable. The posterior density function describes the distribution of a variable of interest just before crashes. The approach is applicable for examination of the impact of one or more variables. This study focuses on the effect of speed and traffic volume on peak and off-peak by considering individual motorway crashes by severity and collision type.

LITERATURE REVIEW

Speed is considered as an important crash risk factor and is related with a large proportion of crashes *(9–12)*. This is explainable considering that speed is a potential crash contributory factor that is always present on the network in contrast to many others that have random (e.g. rainfall) or periodic character (e.g. darkness) *(13)*. It is justified that crashes that occur under higher speed conditions are more likely to have serious impact *(9, 14–16)*. Higher travel speeds are also associated with many crash triggering factors such as lower reaction times, longer decisions, braking and stopping distances, reduction of manoeuvrability and increased possibilities of manipulation error, loss of control and exceeding the critical speed on a curve *(9, 17–19)*. On the other hand, higher speeds are also related with more uniform distribution of speeds (i.e. lower speed variance) that is considered to be beneficial for road safety *(19–22)*.

The studies that examine the role of speed on crashes using statistical modelling have provided different crash-speed relationships. Most of the studies that applied linear regression models found driving speeds to be linearly or exponentially related with crashes *(23–26)*. However, a few studies contradicted the common belief proposing that speed is inversely related to crash frequency *(20)* or that the crash-speed relationship is statistically insignificant *(22)*. Some more recent papers that explored the impact of speed on crashes using advanced crash models have also reported a statistically insignificant relationship between speed and crashes *(27–29)* but others suggested that this relationship is negative *(30)*. However, in models that represent aggregated pre-crash conditions, speed was found to trigger crashes *(7)*. Pei et al. *(16)* attempted to explain the results' inconsistencies suggesting that the crash-speed relationship strongly depends on the selected measure of exposure; the

relationship was shown to be negative when the exposure variable was vehicle miles travelled but positive when the exposure was expressed as vehicle hours travelled.

The relationship of speed with crashes cannot be accurately defined without controlling for the simultaneous effect of other traffic and road characteristics *(9)*. Traffic flow is one of the most important and well-studied crash precursors in the literature. Naturally, the number of vehicles on the roadway is directly proportional with the number of vehicle interactions that can potentially lead to collisions *(19)*. AADT has been found by a large number of researchers to have an increasing relationship with crash frequency *(31–35)*, that in other words means that the busier a road is, the more crashes are expected to occur on it. However, not all crash types are equally related with high traffic; single vehicle crashes have been found to be related with lower traffic than multiple vehicle crashes *(36–40)*.

Inconsistencies or errors in the results might be related with a variety of methodological and data limitations such as the use of less appropriate statistical models, omission bias and data aggregation bias *(4)*. The latter is considered to have significant impact on the validity of the models' results *(5, 6)*, however it is relatively less studied and addressed. Count models that examine crash frequency encompass some kind of crash data aggregation. The dominant crash aggregation method is based on topological and temporal criteria. Crash counts that occurred on pre-defined road links (i.e. link-based models) or areas (i.e. area-wide models) during a certain time period are aggregated and modelled against selected explanatory variables under the assumption that the crash frequency on a particular location can be explained by its most frequently observed, or average, conditions. Research shows however that crashes are related with sudden and unexpected circumstances that may be not possible to be represented by characteristic values such as annual averages *(8)*.

The recently proposed condition-based crash data aggregation approach is an attempt to address aggregation bias issues by developing crash datasets that represent the actual pre-crash traffic and geometric conditions *(7)*. In condition-based modelling crashes are allocated to one out of a number of equally likely pre-crash scenarios that resembled more accurately the conditions on the roadway just before the crash occurrence. This approach enabled an improved representation of the conditions that are related with crashes, but it still involves a level of aggregation which might have impact on the results. This paper proposes an alternative method for defining the role of traffic related potential risk factors that do not require any form of crash aggregation. The method will be demonstrated firstly using synthetic data and then it will be applied to real-world motorway crashes in England with the aim of developing crash-speed-volume relationships

THEORETICAL AND EMPIRICAL DERIVATIONS

The purpose is to develop a non-regression based approach for the modelling of crash data that does not require crash aggregation with the aim of estimating the effect of explanatory variables at the time of the crash. This is achieved by making use of Bayes' Theorem for the case of a continuous variable to derive the conditional probability density of the explanatory variables (e.g. the speed and flow values), given that a crash has occurred. We consider the occurrence of a crash in a location / time interval combination, which will be referred to here as *segments*. First, the method is formulated by considering a single variable (i.e. speed). Following, the method is extended to the case of multiple explanatory variables related to crashes.

One explanatory variable

Consider having available *N* traffic measurements on all the segments of a network over a year and n traffic measurements that represent the traffic conditions at or around the time and location of each individual crash that occurred on the network during the study period. The probability of a crash occurring in any of such traffic measurements is very small (i.e. $N \gg n$). This probability is assumed to be dependent upon the conditions that apply at that point and time (e.g. traffic flow, speed). For simplicity of presentation it will be firstly assumed that this probability is a function only of speed (v) and so is denoted by $g(c|v)$ which might typically be of the form $g(c|v) = bv^{\beta}$ in which b is a constant and β is the speed parameter.

Assuming that $f(v)$ denotes the prior probability density function of speed values over all segments, and $g(c|v)$ is the risk of a crash in a segment with speed v, then Bayes' Theorem for the

case of a continuous variable can be applied to derive the posterior density function $p(v|c)$ describing the distribution of speeds applying amongst the crash set:

$$p(v|c) = \frac{f(v)g(c|v)}{\int f(v)g(c|v)dv} = \frac{f(v)\,v^{\beta}}{\int f(v)\,v^{\beta}dv} \quad (1)$$

Note that the denominator of equation (1) is (apart from the constant b) the probability of a collision on any randomly chosen segment, and is a function of the parameter β. Then, if we have a set of n crashes with associated speed values v_1 ... v_n the log likelihood function can be written as follows:

$$logL(\beta) = \sum_{j=1}^{n} \log\left(p\left(v_j|c_j\right)\right) \quad (2)$$

Equation (2) is maximised with respect to β using the *maxLik* function in *R (41)*.
To illustrate this approach, we generate a large sample of segments' speeds (say N_h) from a normal distribution with mean 100 km/h and standard deviation 13 km/h i.e. **N**(100,13^2). For each one, we generate a random uniform variate to determine whether or not a crash will occur (with probability bv^{β}) using $\beta = 2$, and the values of speed samples (N_h) and b such that about 1,600 crashes are generated.

Based on the value that was selected for β it is assumed that higher speeds are more represented than lower speeds amongst the crash speeds (e.g. the risk of a collision at 120 km/h is nine times greater than at 40 km/h, because β = 2). Figure 1 presents the plots of the density functions of a) the prior and b) the posterior distributions of speeds. It can be seen that the latter is slightly shifted to the right, as higher speeds are relatively more represented in the crash set than in the segments as a whole.

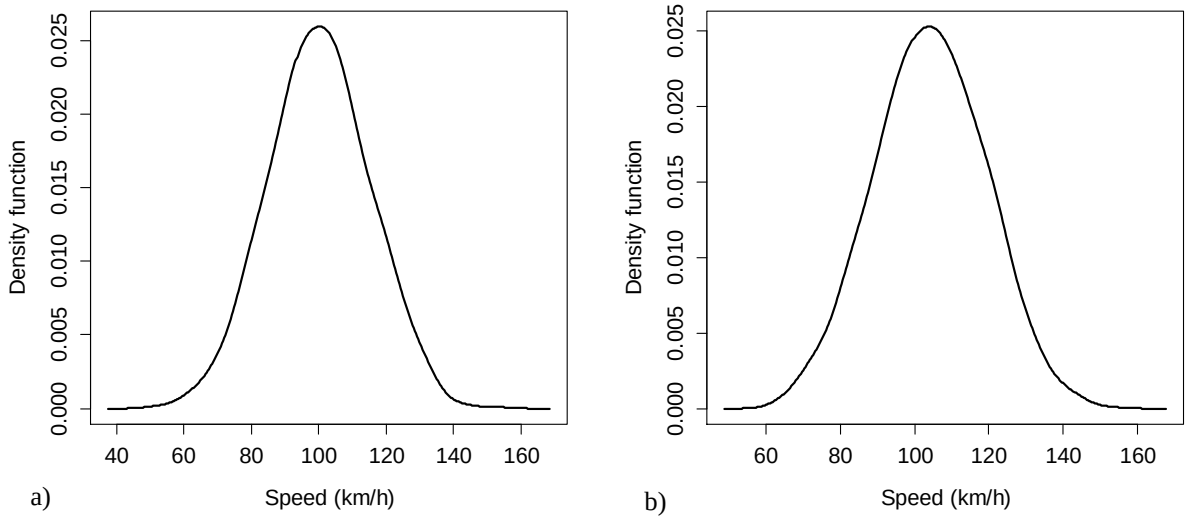

Figure 1: a) Prior and b) posterior speed distributions

In applying equations (1) and (2), the prior distribution $f(v)$ must be given – or an estimate of it should be employed. Since the distribution of $f(v)$ is unknown, the kernel density estimate of a sample of all speeds is used. The integration in the denominator of Equation (1) is obtained by summing over the points used in the kernel density estimation. Using code in R that generates a sample of artificial data as described above and then maximises the log likelihood, the estimate of $\hat{\beta}$ is found to be 1.92 with a standard error of 0.19, confirming the good behaviour of the method.

Extension to more than one explanatory variable

Crash risk cannot be determined solely by the speed levels on a segment. The model outlined above can be extended to include two or more explanatory variables. Assuming that the risk $g(c|q,v)$ of a collision on a segment is now a function of traffic volume (q) and speed (v) so that $g(c|q,v) = bq^{\alpha}v^{\beta}$, it is possible to estimate the parameters α and β from the observed values of the traffic volume and speed that prevailed at the time and location of each of a set of collisions. As before, in addition to the observations on the crash set, the values of volume and speed for a randomly chosen sample of segments are required, so as to provide an estimate of the prior density function $f(q,v)$. Then Bayes' Theorem (see equation (3)) can again be applied to derive the posterior density function $p(q,v|c)$ describing the distribution of flows and speeds applying amongst the crash set:

$$p(q,v|c) = \frac{f(q,v)g(c|q,v)}{\int\int f(q,v)g(c|q,v)dqdv} = \frac{f(q,v)\,q^{\alpha}v^{\beta}}{\int\int f(q,v)\,q^{\alpha}v^{\beta}dqdv} \quad (3)$$

The log likelihood is then:

$$logL(\alpha,\beta) = \sum_{j=1}^{n} \log\left(p(q_j, v_j|c_j)\right) \quad (4)$$

The log likelihood was again maximised with respect to the parameters α and β using the *maxLik* function in R *(41)*.

For a similar illustration as for the univariate model, artificial data are generated with $\alpha = 0.85$ and $\beta = 2$, and with some partial relationship between the flow values and the speed values (with mean speed reducing as flow increases), but with random (normally distributed) error terms added to the mean speed value. The size (n_h) of the collision set is around 1,600, as is that of the previous example. The estimates obtained from one such dataset are: $\hat{\alpha}$ = 0.93 (standard error = 0.07), and $\hat{\beta}$ = 2.04 (standard error = 0.22), which are reasonably close to the true values of 0.85 and 2.0 respectively. The contour plots of kernel density estimates of the prior (left) and posterior (right) are shown in Figure 2. The lower 25%, 50% and 75% contours of the density regions are indicated. The underlying linear relationship between speed (vertical axis) and flow (horizontal axis) can be seen, as can be seen the shift towards relatively higher flow and higher speed as we move from the prior to the posterior.

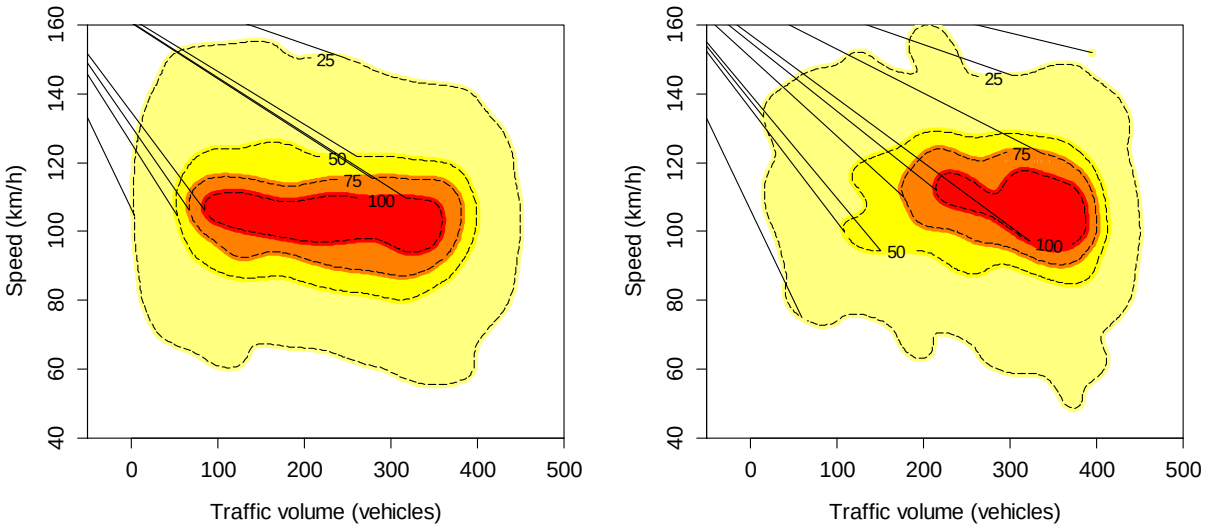

Figure 2: Contour plot of the a) prior and the b) posterior kernel density functions for speed and volume.

DATA DESCRIPTION

As shown in the previous section, the developed models have provided acceptable parameter estimates for the synthetic datasets. It is therefore our expectation that the model would exhibit the same behaviour for the case of real world data. As has been outlined in the previous section, in order to apply the model shown in equation (3) the distributions of the traffic conditions on the road network and just before crashes must be known.

Network traffic data of the year 2012 were available for the entire Strategic Road Network of England (SRN). The SRN that consists of all the motorways and the major A-roads, is the busiest and the most significant network of the country *(41)*. Traffic data were extracted from the UK Highways Agency Journey Time Database (JTDB) *(42)* which includes link-level traffic measurements obtained by inductive loop detectors which are installed all over the road network. The measurement interval is 15 minutes and the total number of (junction to junction) links is 2,505, resulting in a dataset of approximately 88 million observations. The variables used for this analysis are: 15-minute average speed (km/h) and 15-minute volume (vehicles).

Crash data were obtained from the National Crash Database, the so-called STATS19 reports *(43)*. During 2012 there were overall 10,520 reported crashes of which 1.9% were fatal, 11.7% serious and 86.4% slight. Along with a unique reference code there are a number of elements of information that are reported for each crash such as:

- Crash date
- Time of the crash
- Crash location (coordinates)
- Road name
- Road type
- Crash Severity
- Number of involved vehicles
- Vehicle(s)' direction

The traffic conditions on the roadway that are related with crashes are not directly available from the STATS19 reports. To obtain this information, attributes of the network and the crash datasets were merged. The objective of this merger was to identify as precisely as possible the prevailing traffic pre-crash conditions. The reported crash locations were employed to indicate the road links where each of the crashes occurred. Crash locations in the form of coordinates are likely to contain errors. To improve the accuracy of the matching between links and crashes, crash locations were refined using a Fuzzy-logic based crash mapping algorithm that provides almost 99% crash allocation *(44)*. In this way, after identifying a road link for each crash, the reported crash date and time were used to identify the traffic conditions prior the collision. Initially, the set of measurements on the date of the crash that includes the crash time was identified (e.g. for a crash "A" that occurred at 07:03 the set of traffic measurements is the 07:00-07:15). The measurement interval of network data is quite long (15 minutes) thus for crashes that occurred on the beginning of the interval, such as the one in the example, these measurements might be more representative of the traffic conditions after the collision rather than before. To avoid this undesirable effect that may lead in misrepresentation of the traffic conditions, the measurement set prior to the time of the crash was also considered (e.g. for crash "A" this is the 06:45-07:00 traffic measurements set). The final traffic conditions for each crash were estimated using a weighted average of the values of these two measurement sets with the weights determining the proportion of the 15-minute interval before the crash that corresponds to each set. (e.g. for crash "A" the weight for the interval 06:45-07:00 is $\frac{12}{15} = 0.8$ and for the interval 07:00-07:15 is $\frac{3}{15} = 0.2$).

RESULTS

Analysing all crashes together might not be the most appropriate way to evaluate crash precursors because crashes of different severity levels and collision types might be related with dissimilar contributory factors (36, 39). Therefore, main carriageway crashes of the database were divided into five categories:

- All
- Fatal and serious injury
- Slight injury
- Single vehicle
- Multiple vehicle

In addition, to investigate the potential differences between the triggering factors of crashes that occur at peak and off-peak hours the aforementioned five crash categories were examined separately for peak and off-peak hours. The peak and off-peak time intervals were defined from the hourly annual average traffic volume measurements that are presented in Figure 3. Observing the volume levels it can be seen that the volume is higher from 7am to 7pm and so this time interval was set to define "peak times". The rest of the day was considered to be "off-peak"

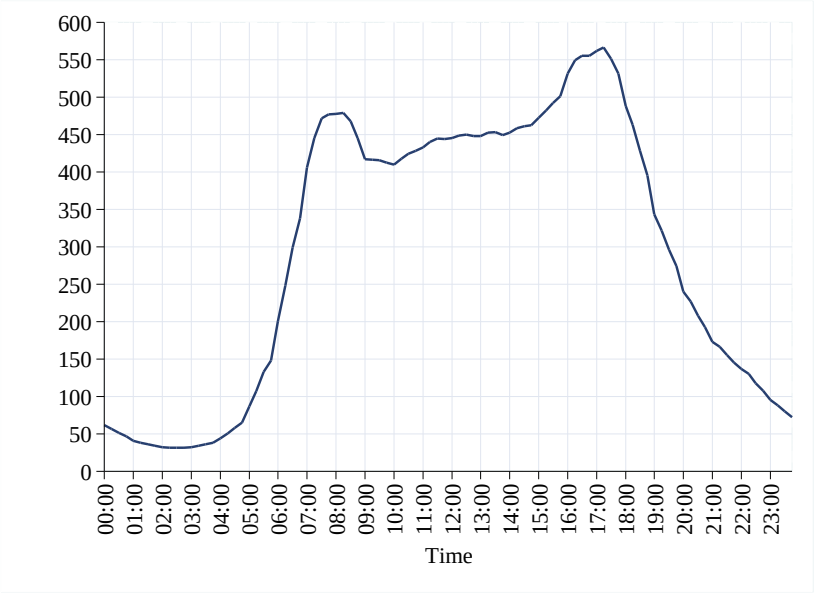

Figure 3: Annual average traffic volume over a day.

For each crash category the impact of speed and volume was examined first separately (Equations (1), (2)) and then combined (Equations (3), (4)). The speed only models will be henceforth referred as SO and the volume only as VO. Models that take into account the effect of speed and volume simultaneously will be abbreviated SV.

From Table 1 it can be seen that the results between crashes that occurred at peak times are significantly different than those for off-peak crashes. The risk of motorway crashes, apart from the single vehicle ones, during peak times was found to increase at lower speeds and higher volumes. This finding implies that crash frequency at peak times is more related with congestion than speed. This is consistent with studies that examined the impact of congestion on crash frequency and severity *(45, 46)*. Multiple vehicle crashes were found to have the strongest relationship with traffic volume a finding similar to those of other researchers *(36, 46)*. The exponent of the risk function for traffic volume was estimated to be 1.1 which indicates, for instance, that the risk of crashes when the traffic volume is 400 is 2.14 (i.e. 21.1) times higher than when it is 200, *ceteris paribus*. The risk of single vehicle crashes that occurred at peak times is proportional to both the traffic speed and volume, which is also in line with literature that suggests that single vehicle crashes are one of the most speed-related crash types *(47)*. Figure 4 shows that the posterior density is shifted to the right compared to the prior

density. This implies that high speeds are more represented in the crash dataset than the network dataset.

Table 1 summarises the estimated parameters of each model along with their standard errors. The estimates represent the exponents (α and β) of the risk functions (i.e. $g(c|v)$ for the SO model, $g(c|q)$ for the VO model and $g(c|v,q)$ for the SV model). Negative exponents show that as the value of the examined variable decreases the probability of a crash increases and positive exponents imply the opposite. Figures 4 and 5 show the contour plots of the prior and the posterior speed-volume distributions for all peak and off-peak crashes respectively.

From Table 1 it can be seen that the results between crashes that occurred at peak times are significantly different than those for off-peak crashes. The risk of motorway crashes, apart from the single vehicle ones, during peak times was found to increase at lower speeds and higher volumes. This finding implies that crash frequency at peak times is more related with congestion than speed. This is consistent with studies that examined the impact of congestion on crash frequency and severity *(45, 46)*. Multiple vehicle crashes were found to have the strongest relationship with traffic volume a finding similar to those of other researchers *(36, 46)*. The exponent of the risk function for traffic volume was estimated to be 1.1 which indicates, for instance, that the risk of crashes when the traffic volume is 400 is 2.14 (i.e. $2^{1.1}$) times higher than when it is 200, *ceteris paribus*. The risk of single vehicle crashes that occurred at peak times is proportional to both the traffic speed and volume, which is also in line with literature that suggests that single vehicle crashes are one of the most speed-related crash types *(47)*. Figure 4 shows that the posterior density is shifted to the right compared to the prior density. This implies that high speeds are more represented in the crash dataset than the network dataset.

Table 1: Models' parameter estimations for peak and off-peak crashes.

Peak Times										
Variable	All crashes		Fatal & serious-injury crashes		Slight-injury crashes		Single vehicle crashes		Multiple vehicle crashes	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speed (SO)	**-0.655**	0.034	**-0.329**	0.114	**-0.691**	0.035	**0.565**	0.108	**-0.865**	0.034
Volume (VO)	**0.774**	0.019	**0.338**	0.050	**0.836**	0.020	**0.139**	0.033	**1.004**	0.023
Speed (SV)	**-0.989**	0.030	**-0.545**	0.109	**-1.035**	0.031	**0.461**	0.107	**-1.213**	0.031
Volume (SV)	**0.859**	0.019	**0.387**	0.051	**0.924**	0.020	**0.104**	0.033	**1.102**	0.022
Off-Peak Times										
Variable	All crashes		Fatal & serious-injury crashes		Slight-injury crashes		Single vehicle crashes		Multiple vehicle crashes	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speed (SO)	**0.761**	0.119	**0.650**	0.288	**0.786**	0.141	**0.848**	0.183	**0.682**	0.175
Volume (VO)	**0.525**	0.022	**0.266**	0.048	**0.587**	0.025	**0.216**	0.029	**0.858**	0.033
Speed (SV)	0.092	0.120	0.148	0.279	0.077	0.137	**0.448**	0.191	-0.235	0.158
Volume (SV)	**0.472**	0.023	**0.241**	0.046	**0.532**	0.026	**0.190**	0.027	**0.815**	0.034

**Statistically significant estimations are marked with bold font.*

Off-peak crashes were found to be more related with high speed conditions, a result rather expected as driving with inappropriate speed tends to be more frequent at night time *(48)*. When speed was examined individually (SO models), the exponent of the risk function was found to be positive for all crash types and severities. However, when the models controlled for traffic volume the effect of speed on the risk function was eliminated. In SV models the parameter of speed for all crashes excluding the single vehicle ones, was found to be statistically insignificant. For single vehicle crashes the exponent for speed is 0.448 which means that the risk of single vehicle crash at off-peak times when speed is 120 km/h is approximately 1.2 (i.e. $1.5^{0.448}$) times higher than when speed is 80 km/h. The effect of traffic volume on off-peak crash frequency is positive for all off-peak crashes but

relatively lower compared to those of peak time crashes. From Figure 5 it can be seen that the posterior density has a small rightwards shift towards the axis of volume. There is also a slight shift up in the lower section of the contours, which indicates that low speeds were less observed in the off-peak crash dataset than in the network dataset.

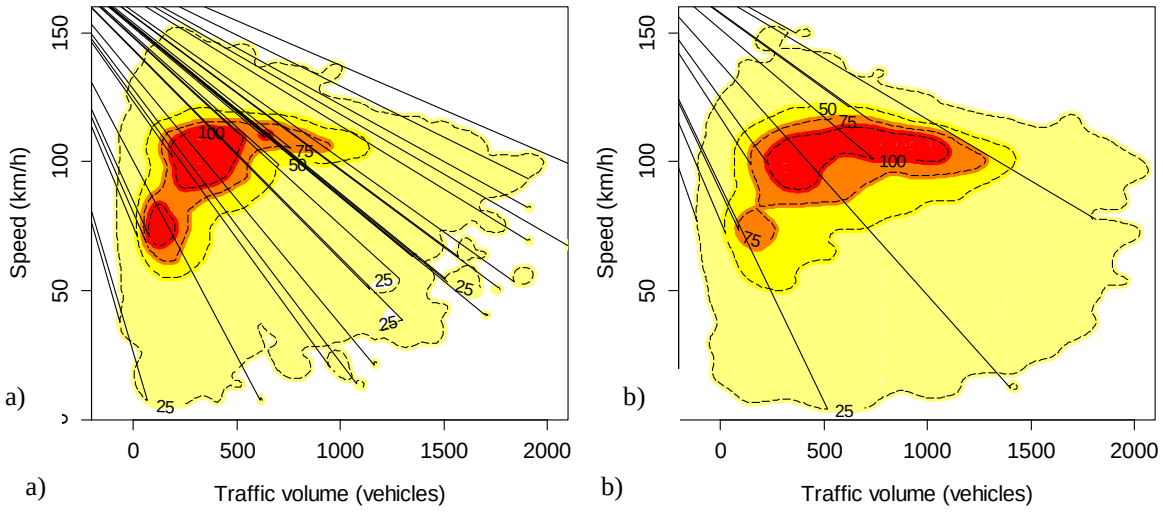

Figure 4: Contour plots of the a) prior and b) posterior kernel density of speed and volume for all crashes at peak times

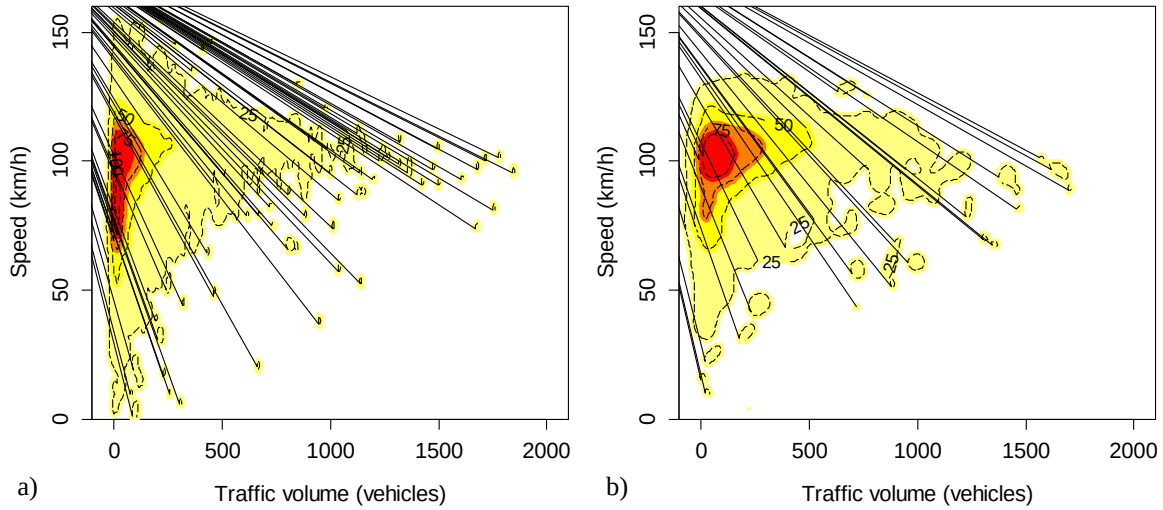

Figure 5: Contour plots of the a) prior and b) posterior kernel density of speed and volume for all crashes at off-peak times

CONCLUSIONS

The purpose of this paper was to demonstrate and apply a new method for modelling crash data in order to define the impact of the levels of traffic speed and volume on motorway crashes. In

contrast to count regression models, that are typically employed for examining crash frequency, the proposed method does not require crash data aggregation but deals explicitly with the traffic conditions prevailing at the time and location of each recorded crash. In this way, aggregation bias problems that are expected to decrease the reliability of modelling outcomes are avoided. As it can be expected that traffic conditions influence the likelihood of crashes, it follows that the distribution of traffic conditions amongst the crash set (the posterior distribution) will be different from that prevailing in the population of traffic conditions generally (the prior distribution). The method is theoretically founded on the estimation of conditional probability as expressed by the Bayes' Theorem and, by effectively comparing the prior and posterior distributions of traffic conditions, it aims to estimate the parameters that define the risk that is associated with each of the examined crash contributory factors. The application of the proposed model on synthetic datasets validated its capability of providing reliable estimates before being applied to the real-world datasets.

The results provide useful insight on the relationship of traffic conditions with crashes. In summary, peak-time crashes, apart from the single vehicle ones but including those that had serious outcomes, were found to be mainly related with congested traffic; they increase while speed drops and volume rises. The relationship of off-peak crashes with speed was found to be positive when it was examined individually but in models that control for the effect of traffic volume, its effect was negligible. However, the risk of single vehicle crashes independently of the time of their occurrence was found to be proportional with speed even when volume was included in the model. From the modelling outcomes it is confirmed that high traffic volumes are linked with more crashes of all types and severities although the maximum effect is on multiple vehicle collisions. The relationship of speed with crashes varies according to the general traffic conditions and therefore it is not straightforward to reply to the question of whether speed has negative impact of road safety. However, considering that speed was found to contribute to crashes that are related with low density conditions (i.e. single vehicle, off-peak) when drivers tend to be free to choose their own speeds, speeding can be regarded as crash triggering action.

In order to better understand crash risk, more factors should be taken into consideration. Traffic density, speed variance and v/c ratio are some of them. The absence of these factors in the developed models may have affected the estimations. Another limitation of this study is that the measurement interval of the traffic measurements (i.e. 15 minutes) was quite long to provide precise pre-crash conditions estimations. In the future, similar analyses should include a more complete specification of risk that will lead to a broader view of the impact of traffic characteristics on crashes and also higher resolution measurements (e.g. 15 seconds). An interesting extension of the model would also include the development risk functions that will incorporate spatial characteristics that are possibly related with crash occurrences such as road geometry.

REFERENCES

1. Brown ID (1982) Exposure and experience are a confounded nuisance in research on driver behaviour. Accid Anal Prev 14:345–352.
2. Montella A (2011) Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid Anal Prev 43:1451–63.
3. Hagenzieker MP, Commandeur JJF, Bijleveld FD (2014) The history of road safety research: A quantitative approach. Transp Res Part F Traffic Psychol Behav 25:150–162.
4. Lord D, Mannering F (2010) The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp Res Part A Policy Pract 44:291–305.
5. Davis GA (2004) Possible aggregation biases in road safety research and a mechanism approach to accident modeling. Accid Anal Prev 36:1119–1127.
6. Clark W, Avery K (1976) The effects of data aggregation in statistical analysis. Geogr Anal 8:428–438.
7. Imprialou M-IM, Quddus M, Pitfield DE, Lord D (2016) Re-visiting crash–speed relationships: A new perspective in crash modelling. Accid Anal Prev 86:173–185.

8. Hossain M, Muromachi Y (2013) Understanding crash mechanism on urban expressways using high-resolution traffic data. Accid Anal Prev 57:17–29.
9. Aarts L, Van Schagen I (2006) Driving speed and the risk of road crashes: A review. Accid Anal Prev Prev 38:215–224.
10. Clarke DD, Ward P, Bartle C, Truman W (2010) Killer crashes: fatal road traffic accidents in the UK. Accid Anal Prev 42:764–70.
11. Shibata A, Fukuda K (1994) Risk factors of fatality in motor vehicle traffic accidents. Accid Anal Prev 26:391–397.
12. Assum T (1997) Attitudes and road accident risk. Accid Anal Prev 29:153–159.
13. Elvik R, Christensen P, Amundsen A (2004) Speed and road accidents: An evaluation of the Power Model.
14. Joksch HC (1993) Velocity change and fatality risk in a crash—A rule of thumb. Accid Anal Prev 25:103–104.
15. Kloeden CN, Mclean AJ, Moore VM, Ponte G (1997) Travelling Speed and the Risk of Crash Involvement Volume 1 - Findings. South Australia
16. Pei X, Wong SC, Sze NN (2012) The roles of exposure and speed in road safety analysis. Accid Anal Prev 48:464–71.
17. Fildes B, Lee S (1993) The Speed review: Road Environment, Behaviour, Speed Limits, Enforcement and Crashes. Victoria
18. Hale AE (1990) Safety and speed. A systems view of determinants and control measures. IATSS Res. 14:
19. Navon D (2003) The paradox of driving speed: two adverse effects on highway accident rate. Accid Anal Prev 35:361–7.
20. Lave C (1985) Speeding , Coordination , and the 55 MPH Limit. Am Econ Assoc 75:1159–1164.
21. Graves P, Lee D, Sexton R (1993) Speed variance, eforcement and the optimal speed limit. Econ Lett 42:237–243.
22. Garber NJ, Gadiraju R (1989) Factors Affecting Speed Variance and Its Influence on Accidents. Transp Res Rec J Transp Res Board 1213:64–71.
23. Fildes BN, Rumbold G, Leening A (1991) Speed Behaviour and Drivers' Attitude to Speeding. Victoria,Austalia
24. Baruya A, Finch DJ (1994) Investigation of traffic speeds and accidents on urban roads. Traffic Manag. Road Safety. Proc. Semin. J held PTRC Eur. Transp. Forum,September 12-16, Vol. P381
25. Quimbly A, Maycock G, Palmer C, Buttress S (1999) The factors that influence a driver ' s choice of speed — a questionnaire study. Transport Research Laboratory, Crowthorne, England
26. Taylor MC, Lynam DA, Baruya A (2000) The effects of drivers' speed on the frequency of road accidents. Transport Research Laboratory, Crowthorne, England
27. Kweon Y-J, Kockelman K (2005) Safety effects of speed limit changes: Use of panel models, including speed, use, and design variables. Transp Res Rec J Transp Res Board 148–158.
28. Quddus M (2013) Exploring the Relationship Between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS. J Transp Saf Secur 5:27–45.
29. Kockelman KM, Ma J (2007) Freeway Speeds and Speed Variations Preceding Crashes , Within and Across Lanes. Transp Res Forum 46:43–61.
30. Baruya A (1998) Speed-accident relationships on European roads. 9th Int. Conf. Road Saf. Eur. 1:
31. Milton J, Mannering F (1998) The relationship among highway geometrics , traffic-related elements and motor-vehicle accident frequencies. Transportation (Amst) 25:395–413.
32. Abdel-Aty M, Radwan AE (2000) Modeling traffic accident occurrence and involvement. Accid Anal Prev 32:633–42.
33. Miaou SP, Lum H (1993) Modeling vehicle accidents and highway geometric design relationships. Accid Anal Prev 25:689–709.
34. Anastasopoulos PC, Mannering FL (2009) A note on modeling vehicle accident frequencies with random-parameters count models. Accid Anal Prev 41:153–9.

35. Chang L-Y (2005) Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Saf Sci 43:541–557.
36. Qin X, Ivan JN, Ravishanker N (2004) Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid Anal Prev 36:183–191.
37. Martin JL (2002) Relationship between crash rate and hourly traffic flow on interurban motorways. Accid Anal Prev 34:619–629.
38. Lord D, Manar A, Vizioli A (2005) Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. Accid Anal Prev 37:185–99.
39. Kim DG, Washington S, Oh J (2006) Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections. J Transp Eng 132:282–292.
40. Bham GH, Javvadi BS, Manepalli URR (2012) Multinomial Logistic Regression Model for Single-Vehicle and Multivehicle Collisions on Urban U.S. Highways in Arkansas. J Transp Eng 138:786–797.
41. Henningsen A, Toomet O (2011) maxLik: A package for maximum likelihood estimation in R. Comput Stat 26:443–458.
42. Department for Transport (2011) Road Network Policy Consultation. London, England
43. Highways Agency (2011) HATRIS JTDB Reference Manual.
44. Department for Transport (2011) STATS19 road accident injury statistics – report form. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/ 230590/stats19 .pdf.
45. Imprialou M-IM, Quddus M, Pitfield DE (2014) High accuracy crash mapping using fuzzy logic. Transp Res Part C Emerg Technol 42:107–120.
46. Wang C, Quddus M a., Ison SG (2011) Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accid Anal Prev 43:1979–1990.
47. Yan X, Radwan E, Abdel-Aty M (2005) Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. Accid Anal Prev 37:983–95.
48. Lang SW, Waller PF, Shope JT (1996) Adolescent driving: Characteristics associated with single-vehicle and injury crashes. J Safety Res 27:241–257.
49. Ivan JN, Pasupathy RK, Ossenbruggen PJ (1999) Differences in causality factors for single and multi-vehicle crashes on two-lane roads. Accid Anal Prev 31:695–704.