Impacts of speed variations on freeway crashes by severity and vehicle type

Pushpa Choudhary
Research student
Transportation systems engineering,
Department of Civil Engineering,
Indian Institute of Technology (IIT) Bombay
Powai, Mumbai-400 076, India
Tel: +91-22-2576-4348;
E-mail: pushpa11choudhary@gmail.com

Dr Marianna Imprialou
Lecturer (Assistant Professor) in Transport Engineering
School of Civil & Building Engineering
Loughborough University
Loughborough LE11 3TU
United Kingdom
Tel: +44 (0) 1509223422
Email: M.Imprialou@lboro.ac.uk

Dr Nagendra R Velaga (Corresponding Author)
Associate Professor
Transportation Systems Engineering,
Department of Civil Engineering,
Indian Institute of Technology (IIT) Bombay
Powai, Mumbai-400 076, India
Tel: +91-22-2576-7341;
E-mail: n.r.velaga@iitb.ac.in velaga@iitb.ac.in

Dr Alok Choudhary
Reader in Supply Chain Management
School of Business and Economics,
Loughborough University
Loughborough LE11 3TU
United Kingdom
Tel: + 44 (0)1509 228302;
E-mail: A.Choudhary@lboro.ac.uk

Abstract

Speed variations are identified as potentially important predictors of freeway crash rates; however, their impacts on crashes are not entirely known. Existing findings tend to be inconsistent possibly because of the different definitions for speed variations, different crash type consideration or different modelling and data aggregation approaches. This study explores the relationships of speed variations with crashes on a freeway section in the UK. Crashes split by vehicle type (heavy and light vehicles) and by severity mode (killed/serious injury and slight injury crashes) are aggregated based on the similarities of the conditions just before their occurrence (condition-based approach) and modelled using Multivariate Poisson lognormal regression. The models control for speed variations along with other traffic and weather variables as well as their interactions. Speed variations are expressed as two separate variables namely the standard deviations of speed within the same lane and between-lanes over a five-minute interval. The results, similar for all crash types (by coefficient significance and sign), suggest that crash rates increase as the within lane speed variations raise, especially at higher traffic volumes. Higher speeds coupled with greater volume and high between-lanes speed variation also increase crash likelihood. Overall, the results suggest that specific combinations of traffic characteristics increase the likelihood of crash occurrences rather than their individual effects. Identification of these specific crash prone conditions could improve our understanding of crash risk and would support the development of more efficient safety countermeasures.

Keywords: accidents; speed variation; road safety; crash severity; heavy goods vehicles; multivariate count modelling.

1. Introduction

Speed and speed variations are considered to be among the most important crash contributory factors. Several ITS applications such as Variable Speed Limits (VSL) or cooperative systems are designed to provide speed harmonization anticipating that this will lead to lower crash rates (Farah and Koutsopoulos, 2014; Strömgren and Lind, 2016). However, studies considering speed variations as a contributory factor are relatively low in number and their results are varying (Kockelman and Ma, 2007; Quddus, 2013; Shi et al., 2016). Some of the studies find speed variations to be positively associated with crashes (Quddus, 2013; Tanishita and Wee, 2016; Wang et al., 2018) while others find non-significant relationships between speed variations and crash risk (Kockelman and Ma, 2007). Others also report changes in the effects of speed after including speed variance in models (Garber and Gadiraju, 1989).

The often-conflicting results of the existing studies may be related to the multiple definitions used to express speed variations, the differences in modelling approaches or data quality and pre-processing methods. All these suggest that further exploration of this contributory factor is needed. Current advances in crash modelling can be proved useful in the examination of the impact of speed variations on crashes. Recently, crash data aggregation has been found to be highly influential on the estimated coefficients of time-varying variables such as speed and traffic flow (Imprialou et al., 2016a; Imprialou et al., 2016b; Xu et al., 2018; Yu et al., 2018). When crashes are aggregated according to the similarities of the traffic conditions just before their occurrence, modelling results appear to be more reliable than in traditional location-based approaches (Imprialou et al., 2016b). Additionally, research has shown that independent variables in crash modelling have unique effects on different crash types and these are more accurately estimated when the correlations between the examined crash types are taken into consideration (i.e. multivariate count models) (Huang et al., 2017; Lord and Mannering, 2010).

Although there are many multivariate crash prediction models that examine crashes by severity, there is a very limited number of studies that divides crashes by the involved vehicle types and none of them focuses on heavy goods vehicles. This paper analyses the effects of speed variations along with other traffic and weather variables on different types of crashes and specifically by vehicle types (heavy and light vehicles) and by severity type (killed/serious injuries and slight injuries; Property Damage Only (PDO) crash data were not available and therefore this crash type was excluded from the analysis). Multivariate Poisson lognormal regression models are used to develop the relationships that are applied on a dataset aggregated with the condition-based approach.

2. Literature Review

The impact of traffic characteristics on crash frequency and severity has been widely studied in the literature and has offered useful insight into the development of effective mitigation measures. Speed has received a lot of research attention, but the findings regarding its relationship with crash rates are inconsistent (Aarts and Schagen, 2006). It is clear that higher speed is associated with higher crash severity, but the impact of speed on crash frequency is not clearly defined yet. Some studies suggest a positive relationship between speed and crash frequency (Imprialou et al., 2016a; Imprialou et al., 2016b; Kloeden et al., 2002; Taylor et al., 2000); however, others have shown a negative or an insignificant relationship (Kockelman and Ma, 2007; Quddus, 2013; Stuster, 2004). There is also a common belief that speed does not necessarily lead to more crashes as long as there are no extreme speed differences between vehicles on a roadway section. These differences that are typically referred to as speed variations and have been identified as a potentially significant contributory factor; however, their exact effect on crashes remains inconclusive (Aarts and Schagen, 2006; Kockelman and Ma, 2007; Quddus, 2013; Solomon, 1964). There have been significantly fewer studies focusing on speed variations than on speed and other traffic, geometric or environmental variables (Quddus, 2013). This is may be partially because speed variations are not directly measurable and may be hard to be computed unless the available traffic data are highly spatially and/or temporally aggregated.

The effects of speed and its variations were initially studied by Solomon (1964) in a case-control study that suggested that vehicles moving much faster or slower than the modus speed were exposed to higher crash risks introducing the theory "Variance kills". Some subsequent studies reported that speed variation is so highly influential for triggering crashes that it makes the effect of mean speed negligible, suggesting that "Variance kills, not speed" (Garber and Gadiraju, 1989). This was in line with the findings by Quddus (2013) who found that speed variation is associated positively with the crash rates but, the average speed is not. However, it contradicts the outcomes of other studies that find both speed and speed variance to be significant factors for predicting crash frequency (Levy and Asch, 1989; Tanishita and Wee, 2016). Studies on real-time crash prediction have shown negative associations of average speed with crashes, while a positive relationship between speed variation and crashes (Abdel-Aty et al., 2012; Wang et al., 2016; Wang et al., 2015a; Xu et al., 2016; Yu and Abdel-Aty, 2014). Moreover, the effects of speed and speed variations seemed to be related to other traffic variables such as flow (Abdel-aty and Pemmanaboina, 2006; Xu et al., 2016). For instance, Abdel-aty and Pemmanaboina (2006) mentioned that high-speed variation coupled with high occupancy and low variation in volume leads to higher likelihood of a crash, while, Xu et al. (2016) showed that, high-speed variance in high-density traffic flow leads to higher crash risk.

The inconsistencies among the results may be related to the differences between analytical methods and also with the definition of speed variations. Speed variation has been represented by multiple different measures such as differences in speed at individual vehicle level (Kloeden et al., 2002; Solomon, 1964), differences at section level traffic characteristics (Quddus, 2013), the difference between the 90th to the 50th percentile of speeds in each lane (Golob et al., 2004), speed differences between and across lanes (Kockelman and Ma, 2007) and others.

The differences in results could also be related to different crash types. For instance, Kweon and Kockelman, (1996) showed that the effects of speed variation were dependent on crash severity and that specifically slight-injury crashes were associated with high-speed variance. Current crash prediction modelling suggests that separate models for different crash types are not adequate; and therefore, multivariate modelling approaches came into application (e.g. Huang et al., 2017; Imprialou et al., 2016b; Lord and Mannering, 2010; Martensen and Dupont, 2013). Though there are various studies on crash contributory factors by severity levels, there are very few studies focusing on crashes by vehicle type and these are mostly focused on urban environments without making a distinction between heavy and light vehicles (Huang et al., 2017). Whereas, it has been known that due to their unique characteristics (weight, size, stopping distances etc.) heavy vehicles' crash contributory factors should be investigated separately (Wei et al., 2017). Moreover, as per authors' best knowledge, there is no study on investigating the effects of speed variation on heavy vehicle crashes.

Other than speed, traffic volume is one of the most studied factors in crash rate predictions (Aarts and Schagen, 2006; Garber and Ehrhart, 2000). Weather conditions could also affect crash risk (e.g. Abdel-aty and Pemmanaboina, 2006; Wang et al., 2015b; Xu et al., 2016). Typically rainy weather is found to be associated with higher crash rates in most of the previous studies (Abdel-aty and Pemmanaboina, 2006; Lee et al., 2003), possibly because, the wetness of pavement reduces friction, making stopping distances longer (Abdel-aty and Pemmanaboina, 2006).

This study explores further the relationships of traffic characteristics with crash rates with a special focus on the impact of speed variations (defined as speed differences within and between-lanes). Freeway crashes are split by vehicle types (heavy and light vehicle crashes) and severity (killed or serious and slight injury) and are fitted using multivariate count models. In order to achieve a more accurate representation of the conditions just before crashes, data are aggregated following a condition-based approach (Imprialou et al., 2016b).

3. Data Collection and Preparation

To analyse the impact of speed variations on crashes, traffic and weather data have been employed. The study area was decided to be a section of the South-North motorway M1 (Junctions 1-24 (Figure1), located between London and East Midlands Airport) that is one of the most important and busy motorways in England that links London with the North of the country. The length of the study area is 175km per direction and most of its links include three running lanes in each direction. The crash data for three years (from 2013 to 2015) was obtained from the National Road Accident Database of the United Kingdom (STATS 19) (Department for Transport, 2011). Among others, the data included information on severity, involved vehicle types, time, date and location of the crashes. During the study period, there were 1,075 fatal and injury crashes in total, of which 11.25% resulted in killed or seriously injured casualties (henceforth: KS crashes) and 88.75% in slight injuries (henceforth: SL crashes). As the study area belongs to the Strategic Road Network of England, that carries almost two-thirds of England’s freight, 15.90% of all crashes had at least one commercial vehicle with weight over 3.5 tones involved i.e. heavy vehicles (henceforth these crashes will be referred to as HV-crashes). The rest of the crashes (84.10%) were between mainly passenger vehicles or vans with weight 3.5 tonnes or less i.e. light vehicles (henceforth: LV-crashes)[1].

Figure 1: M1 motorway Junctions 1-24, UK (source: Google Maps (2017))

Traffic data were obtained from the Motorway Incident Detection and Automatic Signalling database (MIDAS) of Highways England (Highways England, 2017). The data was

---

[1] In the present study, a crash is defined as HV-crash if the crash includes at least one heavy goods vehicle. Whereas, LV-crashes are the crashes that involve at least one light vehicle but excluding the crashes which include heavy goods vehicle. Therefore, the crashes which include both heavy goods vehicle and light vehicle are classified as HV-crashes. This definition of crashes by vehicle type has been employed in a number of other studies such as: Chen and Chen, (2011); Lemp et al., (2011) and Zou et al., (2017).

collected through 689 inductive loop detectors installed in the study area and provided one-minute-level traffic data disaggregated by running lane. The traffic variables that were used for this analysis were traffic volume and mean speed (km/h) by lane. To develop the final dataset for the analysis, the data were aggregated to the five-minute level and through this aggregation process the following variables have been developed:

- *Total volume*: The total volume was estimated by the summation of the number of vehicles present on a road section between two subsequent loop detectors in each of the running lanes during a 5 min interval.

$$Total\ volume = \sum_{1}^{T}(\sum_{1}^{L} Volume_{t,l}) \tag{1}$$

where l: lane index (1 to 3) and t: number of minutes (1 to 5).

- *Average speed*: For each one-minute interval, mean speeds across the lanes were calculated and then, the average speed for 5 minutes was considered as the average speed.

$$Average\ Speed = \frac{1}{T}\sum_{1}^{T}(\frac{1}{L}\sum_{1}^{L} Speed_{l,t}) \tag{2}$$

where T: total number of minutes (here T=5) and L: the total number of lanes of the road section.

- *Between-lanes speed variation*: For each one-minute interval, the standard deviation of speeds between the lanes was calculated and then, the average of these standard deviations for 5 minutes was considered as the between-lanes speed variation.

$$Between\ lanes\ speed\ variation\ = \frac{1}{T}\sum_{1}^{T}\left(\sqrt{\frac{\sum_{1}^{l}(Speed_{l,t}-\overline{Speed_t})^2}{L}}\right) \tag{3}$$

where $\overline{Speed_t}$: average speed for all lanes for minute t.

- *Within lane speed variation*: For each lane, the standard deviation of speeds for a 5-minute interval was calculated and then the average of these standard deviations for all the three lanes was considered as within lane speed variation.

$$Within\ lane\ speed\ variation = \frac{1}{L}\sum_{1}^{L}\left(\sqrt{\frac{\sum_1^T \left(Speed_{l,t} - \overline{Speed_l}\right)^2}{T}}\right) \tag{4}$$

where $\overline{Speed_l}$: average speed for 5 minutes within lane l.

- *Vehicle hours travelled*: Estimated by multiplying the average travel time on each section (based on average speed) and the corresponding total volume in each 5-minute interval.

Weather conditions were extracted from the open database of MetOffice, the United Kingdom's national weather service (MetOffice, 2016). The weather data was collected on hourly basis from eight weather stations which were found adjacent to the study site based on their geographic locations. Each of the loop detectors in the study area was assigned with one of these eight stations based on the proximity of the station with the loop detector. For the sake of simplicity, weather conditions in this analysis were split into two categories indicating presence or absence of rain. Further, based on the time of the observation of the traffic data, it was matched with the hourly weather data, to provide the weather information for the same 5-min interval.

3.1 Condition-Based Dataset

Data aggregation in crash modelling has been found to influence the results of the analysis significantly (Imprialou et al., 2016a; Imprialou et al., 2016b). Traditionally, crash count models are applied onto location-based datasets, where the number of crashes per location unit (e.g. road link, section or intersection) is modelled against averages of the examined independent variables (e.g. the annual average of speed, AADT, number of lanes). This approach may be effective for the examination of permanent road characteristics such as road geometry. However, it can be less suitable for understanding the impact of time-varying traffic characteristics on crashes. For instance, in an analysis that employs a location-based dataset, speed variation can be only expressed by the annual variance of speed on the study area which might be not representative of the traffic conditions that are related to crashes.

To address this aggregation bias, an alternative condition-based aggregation approach has been proposed, as it indicates the prevailing traffic conditions just prior to the crashes, which can eventually help in identifying the extreme traffic characteristics which might have contributed to crashes. A condition-based model aggregates the crashes based on the similarity of the traffic conditions prior to their occurrence rather than the adjacency of their locations. Therefore, a condition-based dataset includes a number of scenarios that cover all the possible traffic conditions in the study area and each of these scenarios is matched with the respective number of crashes that happened under these conditions (Imprialou et al., 2016b). Consequently, to develop a condition-based dataset, the traffic conditions before each of the examined crashes need to be identified. Pre-crash conditions were defined as the traffic and weather conditions at the closest upstream loop detector to the crash location, five minutes prior to the reported crash time. Some crashes (N=140) were removed from the dataset as they had missing values for the traffic or weather parameters. The final dataset consists of 153 HV crashes (16.37%), 782 LV crashes (83.63%). In terms of severity, 130 crashes (13.9%) were identified as KS crashes, and 805 were identified SL crashes (86.1%).

To prepare the pre-crash scenarios, the traffic variables were grouped into equal frequency categories. The reason behind the formation of different scenarios of the traffic and weather data is to represent all possible conditions which could be present in the study area just before the crashes. In order for the scenarios to have equal frequency and to be mutually exclusive, the traffic characteristics were divided into quantiles. The formation of the database is visualised in Figure 2.

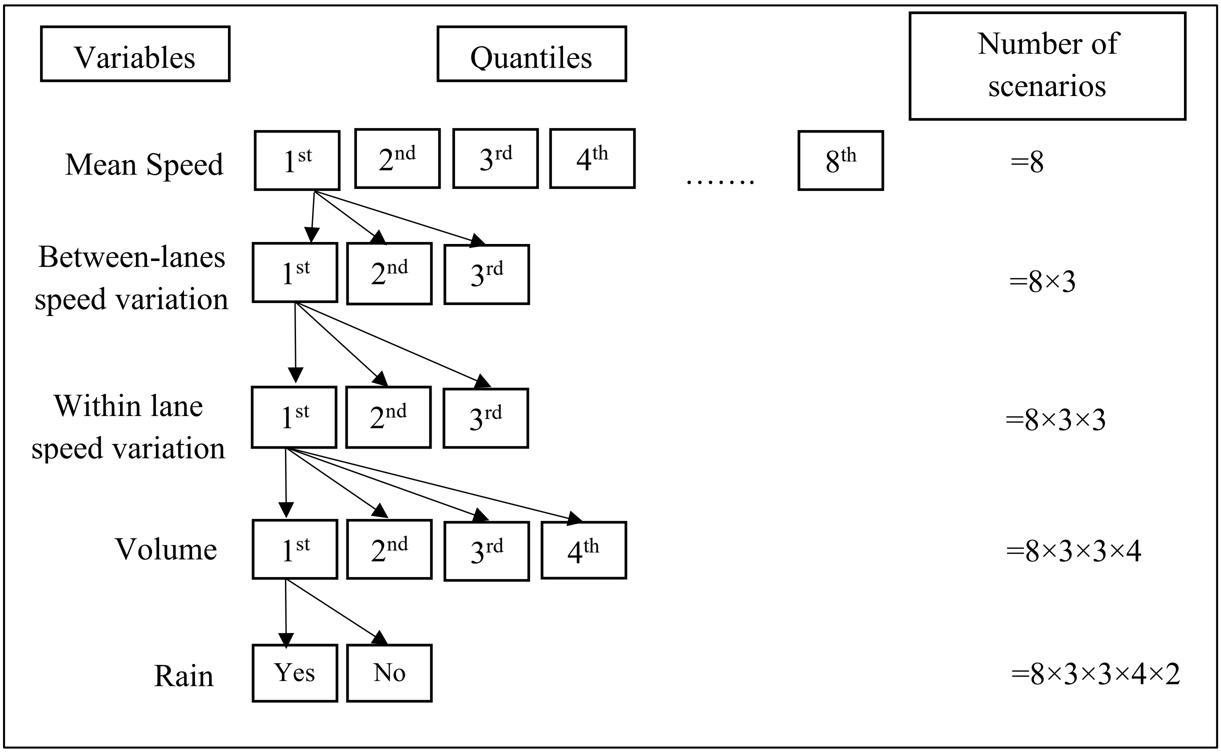


Figure 2 Flow diagram representing the sequence of scenario creation.

The number of scenarios of the condition-based dataset was empirically defined. The order of dividing the variables was followed as per the study aim. As the main aim of the study was to identify the effects of speed and speed variance on the crash frequency, firstly the speed was split then each speed quantile was divided into different quantiles for between-lanes speed variations; and similarly, the within lane speed variance was split under each quantile of for between-lanes speed variations. Further the sequence was followed by splitting the volume and rain variables respectively. The number of scenarios was determined in order to develop a dataset with relatively small number of observations so as to avoid generating too many zeros that might be problematic for the modelling estimations (see (Imprialou, 2015) for a detailed explanation). During the analysis several other scenario aggregations have been tested but the estimated coefficients did not change from those of the model that will be presented in Section 5.

As shown in Figure 2, the best scenario combination was achieved by dividing the average speed into eight quantiles (octiles), and further, dividing the between-lanes speed variation for each separate speed quantile into three quantiles (tertiles). Similarly, the within lane speed variation was divided into three quantiles for each quantile of between-lanes speed variation. The volume was divided into four equal frequency groups (quartiles) for each within lane speed variation category. Finally, the grouping was done for weather conditions (rain/no rain). This grouping led to 576 scenarios (8×3×3×4×2) which included all possible combinations of variables and each observation represented a distinct traffic and weather scenario. The current study developed and compared the outcomes of two datasets that expressed traffic and weather conditions at two different time intervals prior of each crash in the dataset: (1) 0-5 minute interval, (2) 5-10 minute interval.

The traffic characteristics were represented by the median of each quantile. Each crash was then matched with one of the 576 scenarios. The crash frequency for each scenario was presented by vehicle types (HV and LV) and by severity levels (KS and SL). Table 1 shows the descriptive statistics of the study dataset. The exposure on a condition scenario is dependent on the number of vehicles and duration of their movement under these conditions (Imprialou et al., 2016a). Therefore, the total vehicle hours travelled per scenario was selected as the exposure variable for the models.

Table 1 Descriptive statistics for the study dataset

Variable	Mean	SD	Min	Max
Crash variables				
By transport mode				
HV crashes	0.36	0.98	0.00	8.00
LV crashes	1.36	3.00	0.00	28.00
By Severity levels				
KS crashes	0.23	0.64	0.00	5.00
SL crashes	1.40	3.14	0.00	29.00
Traffic variables				
Speed (km/h)	105.35	11.31	41.07	120.90
Between-lanes speed variation (km/h)	14.12	4.45	3.82	49.81
Within lane speed variation (km/h)	5.56	2.08	2.43	12.83
Volume (vehicles in 5 min interval)	177.33	113.88	27.00	399.00
Speed*Volume (km/h*vehicles)	18556.63	11886.22	2423.25	38519.67
Speed*Between-lanes speed variation (km/h*km/h)	1483.02	428.99	226.50	4014.96
Speed*Within lane speed variation (km/h*km/h)	582.82	219.39	195.94	1407.42
Volume* Between-lanes speed variation (vehicles*km/h)	2473.99	1746.29	152.74	10958.94
Volume* Within lane speed variation (vehicles*km/h)	939.93	700.66	85.79	4030.05
Between-lanes speed variation *Within lane speed variation (km/h*km/h)	78.86	40.07	9.33	279.47
Weather variables				
Rain	0.50	0.50	0.00	1.00

4. Methodology

Different crash types sourcing from the same dataset may be potentially correlated. Omission of these correlations from the modelling process, by developing separate count models for each crash type, is likely to lead to erroneous standard errors (Park and Lord, 2007). Multivariate Poisson Lognormal (MVPLN) regression can control for over-dispersion as well as for the correlations between dependent variables; and it has been applied in a number of studies

(Huang et al., 2017; Park and Lord, 2007). This study explores the relationships of speed variations with crash rates by developing two MVPLN models: one that examines the aforementioned traffic and weather variables by vehicle types (HV and LV crashes) and another by severity level (KS and SL crashes).

In an MVPLN, the number of crashes by type (vehicle type or severity) for a dataset with n observations (i.e. condition scenarios) follows a Poisson distribution:

$$y_{ik} \sim Poisson(\lambda_{ik}), \quad i = 1, 2, 3, \dots \dots n; k = 1, 2, \dots . K \tag{5}$$

where i: index of observation, k: index of crash type, y_{ik}:observed number of crashes for k crash type for i^{th} observation and λ_{ik} expected mean for k type crashes for i^{th} observation. Following is the link function for λ_{ik} :

$$ln(\lambda_{ik}) = \beta_{k0} + \sum_{m=1}^{m} \beta_{km} X_{ikm} + ln(e_i) + \varepsilon_{ik} \tag{6}$$

where β_{k0} intercept of k crash type; β_{km}: coefficient of m^{th} explanatory variable for k crash type, X_{ikm}: value of m^{th} explanatory variable for i^{th} observation for k crash type. ε_{ik}: unobserved heterogeneity for i^{th} observation for k crash type.

ε_i is assumed to follow multivariate normal (MVN) distribution and controls for the correlations within the unobserved heterogeneity:

$$\varepsilon_{ik} \sim MVN(0, \Sigma), \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \cdots & \sigma_{11} \\ \sigma_{21} & \sigma_{22} \cdots & \sigma_{2k} \\ \vdots & \vdots & \vdots \\ \sigma_{k1} & \sigma_{k2} \cdots & \sigma_{kk} \end{pmatrix} \tag{7}$$

where Σ is the variance–covariance matrix of the unobserved heterogeneity.

The model's parameters estimation was done using Markov chain Monte Carlo (MCMC) in a Bayesian framework because the direct computation of the marginal distribution of accident counts is not possible to be obtained directly (for more information see: Ma, 2006;

Park and Lord, 2007; Imprialou et al., 2016b; Wang et al., 2015a). The prior distribution for β is multivariate normally distributed:

$$\beta \sim MVN\left(\beta_0, R_{\beta_0}\right) \tag{8}$$

The conjugate prior distribution of the inverse of the variance-covariance matrix for the heterogeneity and the spatial correlation follows a Wishart distribution (Huang et al., 2017; Park and Lord, 2007):

$$\sum\nolimits^{-1} \sim Wishart(R, d) \tag{9}$$

where β_0, R_{β_0} and R are known non-informative hyper parameters and d is equal to the degrees of freedom (number of the examined crash types, in this case $d = k = 2$).

5. Results and Discussion

The models were fitted using WinBUGS software which incorporates full Bayes model estimation approach using the Markov Chains Monte Carlo (MCMC) method (Spiegelhalter et al., 2003). Each model was developed with 200,000 iterations of two Markov chains and the initial 50,000 iterations were discarded from the final model estimates. The actual functional forms of the relationships between traffic variables and crashes are not known and potential interactions between traffic variables cannot be ruled out. Therefore, the present study examined the effects of speed variations using several interaction-term combinations in addition to the individual traffic variables.

All the traffic variables along with all their multiplicative interaction combinations[2] and rain were taken as explanatory variables in both multivariate models in various combinations. The final models that are presented here, were chosen based on the lowest DIC (Deviance Information Criterion) value.

The best-fitting models for vehicle type and severity type are presented in terms of posterior means, standard deviations (SD), MC Error and the 95% credible intervals of the estimated coefficients in Tables 2 and 3, respectively. The correlations between the crash types in each model were also calculated and it was found that both the models showed very high correlations (0.981 and 0.980 for the crash types by vehicle type and by severity levels, respectively). This suggests that the different crash types are related to each other and should be modelled using multivariate models. For both the models, the best fitted variable combination included all traffic and weather variables plus the following interactions: a) volume and speed, b) volume and within lane speed variation and c) Speed and between-lanes speed variation.

Table 2 Multivariate model results for crash rates by vehicle type (HV and LV crashes)

HV crashes						
Variables	Mean	SD	MC Error	2.50%	Median	97.50 %

[2] Possible multiplicative interaction combinations: (i) Volume* Speed, (ii) Volume* Between-lanes speed variation, (iii) Volume* Within lane speed variation (iv) Speed* Between-lanes speed variation (v) Speed* Within lane speed variation (vi) Between-lanes speed variation* Within lane speed variation

Speed	**-0.1292**	0.016	0.001	-0.161	-0.129	-0.098
Volume	**-0.03544**	0.008	0.000	-0.051	-0.035	-0.021
Within lane speed variation	**-0.4776**	0.091	0.002	-0.654	-0.477	-0.300
Between-lanes speed variation	**-0.2538**	0.087	0.003	-0.420	-0.256	-0.079
Rain	**6.537**	0.673	0.016	5.357	6.485	7.993
Volume*Speed	**0.000183**	0.000	0.000	0.000	0.000	0.000
Volume*Within lane speed variation	**0.002204**	0.000	0.000	0.001	0.002	0.003
Speed*Between-lanes speed variation	**0.004118**	0.001	0.000	0.002	0.004	0.006
Intercept	-1.11	1.391	0.043	-3.788	-1.147	1.648
Ln(exposure)	1	Vehicle hours travelled				
LV crashes						
Variables	Mean	SD	MC Error	2.50%	Median	97.50 %
Speed	**-0.1226**	0.015	0.001	-0.152	-0.122	-0.094
Volume	**-0.04516**	0.007	0.000	-0.061	-0.045	-0.032
Within lane speed variation	**-0.4173**	0.069	0.002	-0.552	-0.418	-0.282
Between-lanes speed variation	**-0.241**	0.079	0.003	-0.404	-0.242	-0.085
Rain	**7.994**	0.595	0.019	6.937	7.958	9.292
Volume*Speed	**0.000269**	0.000	0.000	0.000	0.000	0.000
Volume*Within lane speed variation	**0.002449**	0.002	0.000	0.000	0.002	0.003
Speed*Between-lanes speed variation	**0.003549**	0.004	0.001	0.000	0.002	0.005
Intercept	-1.293	1.323	0.046	-3.902	-1.267	1.319
Ln(exposure)	1	Vehicle hours travelled				
Model performance parameters						
$\bar{D}$	1353.49					
p_D	198.213					
DIC	1551.7					

Note: Boldface indicates statistically significant coefficients at the 95% credible interval.

Table 3 Multivariate model results for crash rates by severity levels (KS and SL crashes)

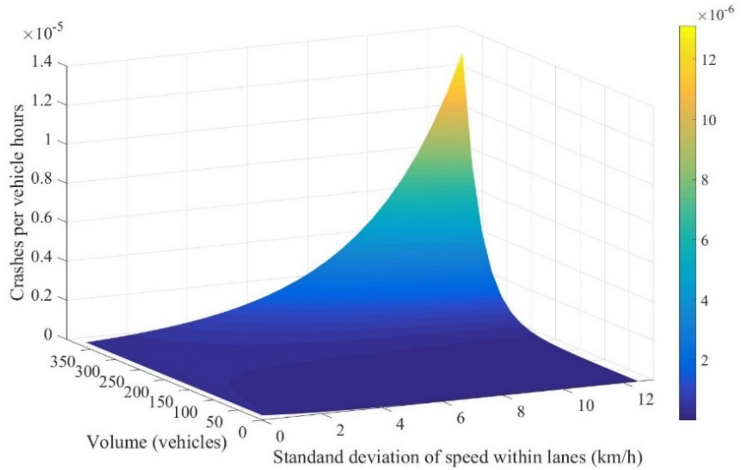
KS crashes						
Variables	Mean	SD	MC Error	2.50%	Median	97.50%
Speed	**-0.1332**	0.020	0.001	-0.167	-0.133	-0.096
Volume	**-0.04594**	0.009	0.000	-0.061	-0.046	-0.028
Within lane speed variation	**-0.3616**	0.097	0.002	-0.521	-0.361	-0.174
Between-lanes speed variation	**-0.3801**	0.136	0.005	-0.619	-0.368	-0.137
Rain	**6.546**	0.716	0.017	5.456	6.491	8.088
Volume*Speed	**0.000257**	0.000	0.000	0.000	0.000	0.000
Volume*Within lane speed variation	**0.002648**	0.001	0.000	0.002	0.003	0.004
Speed*Between-lanes speed variation	**0.005087**	0.001	0.000	0.002	0.005	0.008
Intercept	-1.008	1.624	0.05152	-3.594	-1.059	2.283
Ln(exposure)	1	Vehicle hours travelled				
SL crashes						
Variables	Mean	SD	MC Error	2.50%	Median	97.50%
Speed	**-0.1506**	0.017	0.001	-0.178	-0.152	-0.116
Volume	**-0.03954**	0.007	0.000	-0.051	-0.039	-0.026
Within lane speed variation	**-0.4756**	0.073	0.002	-0.593	-0.477	-0.331
Between-lanes speed variation	**-0.4706**	0.100	0.004	-0.632	-0.472	-0.284
Rain	**8.334**	0.692	0.023	7.273	8.280	9.800

Volume*Speed	**0.000214**	0.000	0.000	0.000	0.000	0.000
Volume*Within lane speed variation	**0.002501**	0.000	0.000	0.002	0.003	0.003
Speed*Between-lanes speed variation	**0.006268**	0.001	0.000	0.004	0.006	0.008
Intercept	0.7937	1.502	0.053	-1.785	0.886	3.433
Ln(exposure)	1	Vehicle hours travelled				
Model performance parameters						
$\bar{D}$	1334.65					
p_D	198.067					
DIC	1532.71					

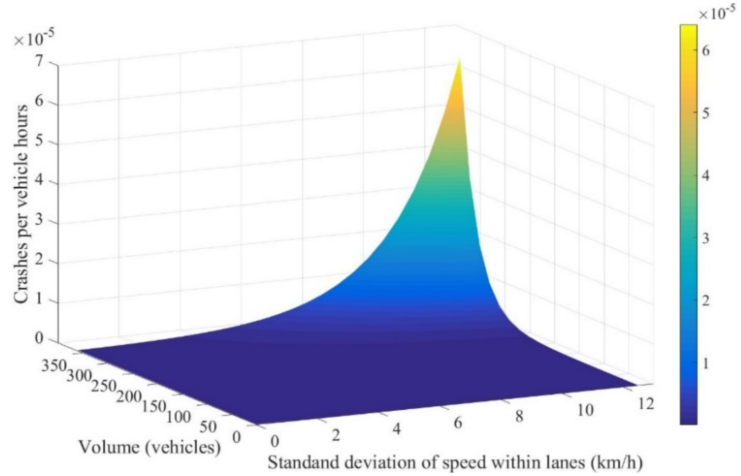
Note: Boldface indicates statistically significant coefficients at the 95% credible interval.

As the main aim of the study is to examine the relationships of speed variations with crashes, the discussion focuses on these effects. Both the variations have negative coefficients but, as both are also present in interaction terms, direct interpretation of the individual coefficients is not possible. To facilitate the interpretation of the interaction of volume and within lane speed variation, the crash rates are plotted against the entire range of within lane speed variations and volume in Figures 3A, 3B, 3C and 3D for HV, LV, KS and SL crashes respectively. The effects of other variables are kept constant (at their mean) while estimating crash rates. For example, the equation used for developing the graph for the HV crash model (Figure 3A) is:

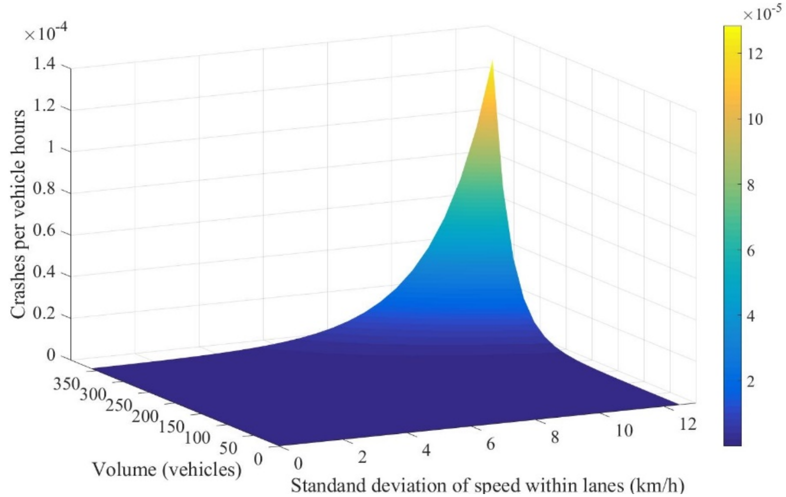
$$\frac{HV\ Crashes}{Veh\ hours} = \exp\Big(-0.1292 \cdot \overline{Speed} - 0.03544 \cdot \overline{Volume} \\ - 0.4776 \cdot \text{Within lane speed variation} \\ - 0.2538 \cdot \overline{Between\ lanes\ speed\ variation} + 6.537 \cdot rain \\ + 0.000183 \cdot \big(\overline{Volume} \cdot \overline{Speed}\big) \\ + 0.002204 \cdot \text{Volume} \cdot \text{Within lane speed variation} \\ + 0.004118 \cdot (\overline{Speed} \cdot \overline{Between\ lanes\ speed\ variation}) - 1.11\Big) \tag{10}$$



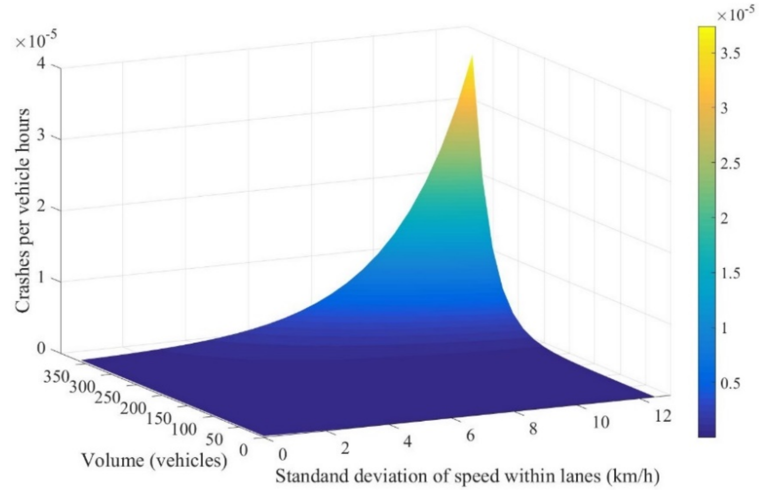
(3A) HV crashes



(3B) LV crashes



(3C) KS crashes



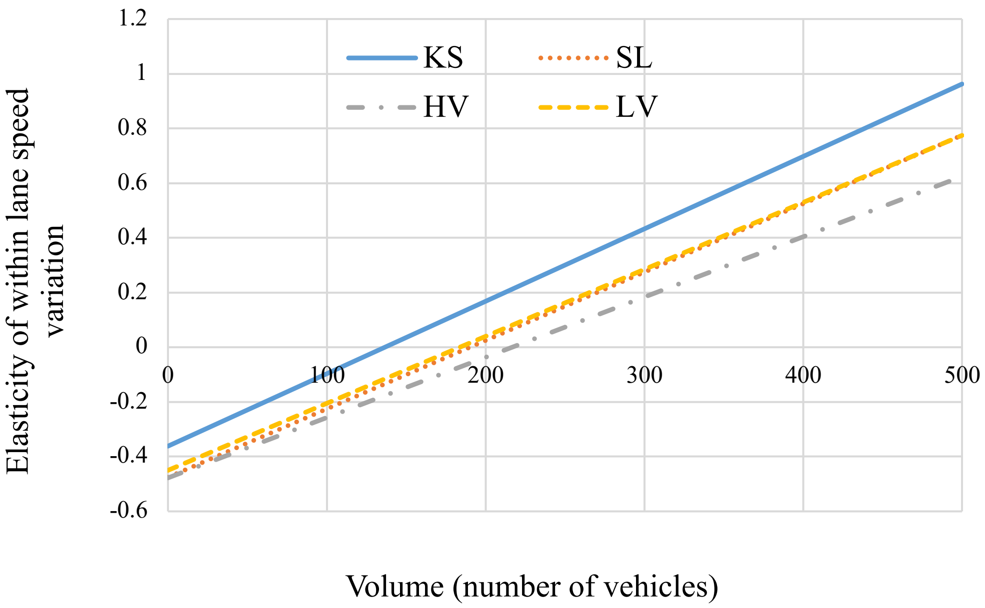
(3D) SL crashes

Figure 3 3d Contour graphs of crashes per vehicle hours as a function of within lane speed variation and volume

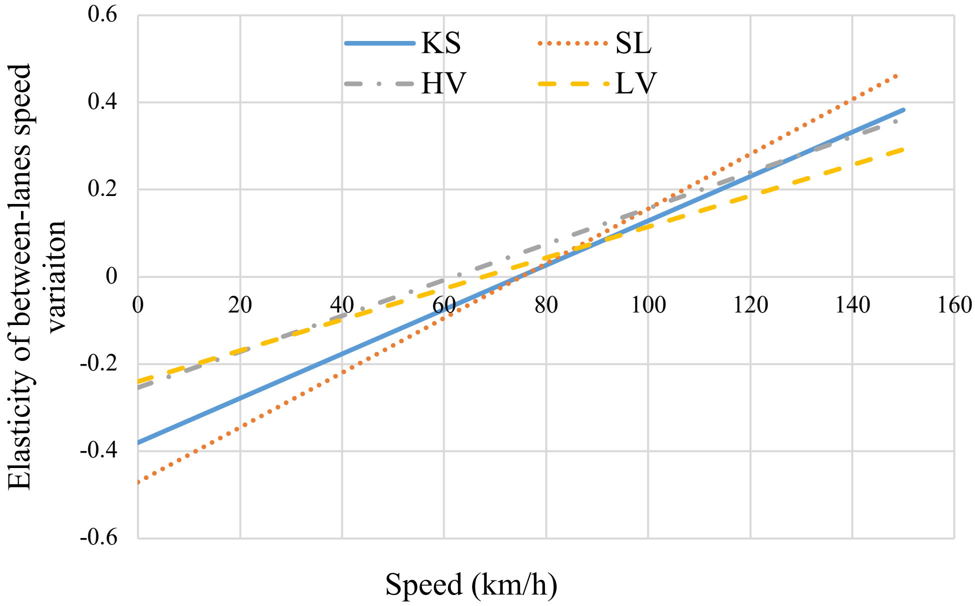
The curves show that the relationship of crash rates with the *within lane speed variation* varies according to the volume conditions on the road. More specifically, all crash types seem to be triggered by within lane speed variation at higher volumes. This is more clearly demonstrated by Figure 4A that shows the elasticity of within lane speed variation as a function of traffic volume. The threshold values of volume where the elasticities become positive are 216, 183, 132, and 187 for HV, LV, KS, and SL crashes respectively. This means that when traffic volume is higher than these values, increases in within lanes speed variation are likely to lead to more crashes. The present study results are in line with the previous study Garber and Erhart (2000) who showed that high variation in speed results into higher crash rates in the presence of high flow per lane, whereas the lower volume may not affect the crash rate significantly. Further, the present results show that the KS crashes have higher elasticities compared to the SL ones (Figure 4A). One of the possible reasons behind this could be that the route analysed in the study is a freeway, as the literature shows that the crashes on the roads characterised with high speed limits are more prone to severe crashes (Zhu and Srinivasan, 2011). In high speed conditions, increase in the within lane speed variance can further worsen the situations in terms of severity.

Traffic conditions with high volume and high speed variation within the same lane represent conditions with lower levels of service and therefore unstable flow. These conditions can create higher crash risk because of the limited spacing between vehicles and therefore lower time to react to sudden changes in nearby vehicle speeds (Li et al., 2018; Xu et al., 2016). So, as expected, under these conditions, more coordinated traffic would be safer. On the other hand, the results for low-volume conditions (i.e. lower crash rates during higher within lane speed variation) are less straightforward to explain. Typically, lower volume conditions are mainly associated with free flow conditions with low demand, however, it can also be observed at slow moving conditions due to congestion during the peak periods. A possible explanation for the first lower volume condition may be that in these conditions, drivers have more freedom to select their comfortable speeds and maintain sufficient spacing from other vehicles. Therefore, even if the differences in speeds within the same lane are high, these do not lead to frequent crash-prone interactions. In the second scenario, slow moving conditions due to congestion during the peak periods, restricts the freedom of the drivers to vary the speed, therefore, it cannot be accounted for the high-speed variance conditions. The results regarding flow and within lane speed variation are consistent with some previous studies which found that crashes happen more in the presence of high-speed variation during congested flow conditions (Golob et al., 2004). The within lane elasticity curve shown in Figure 4A exhibits that an increase in the within lane speed variation and volume will lead to a more sharp increase for KS crashes than the SL crashes. As the higher within lane speed variation relates with the situations of more extreme speeds on the roadway (too slow and too fast), this could be the possible reason for the sharp increase in the crash rate for higher speed variations.

Interaction effects of between-lanes speed variation and speed on the HV, LV, KS and SL crash rates are shown in Figure 5A, 5B, 5C and 5D respectively. The shape of the curves (for all crash types) show that the effect of between-lanes speed variation on the crash risk changes in the presence of different average speeds. Figure 4B shows the elasticities of between-lanes speed variations with respect to average speeds. It indicates that the crash risk increases when both the average speed and the between-lanes speed variations are increasing. Specifically, when average speeds are higher than 61, 67, 75 and 75 km/h for HV, LV, KS and SL crashes respectively crash risk is constantly positively associated with increased between-lanes speed variations. In fact, traffic conditions with speeds lower than these thresholds are particularly rare in the study area as in more than 97% of the time the speed is higher than 70km/h. Comparing the elasticities, it is observed that between-lanes speed variation cause higher crash risk for HV crashes than the LV crashes; and surprisingly, the SL elasticities are higher than the KS crashes under these circumstances.



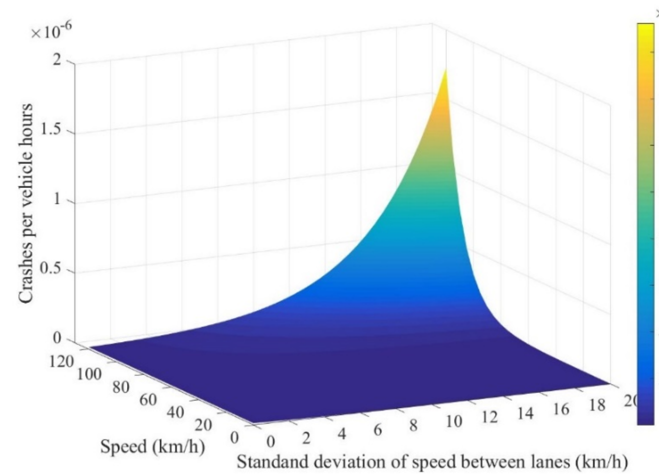
4(A) Elasticity of within lane speed variation across different types of crash rates for a range of volume values



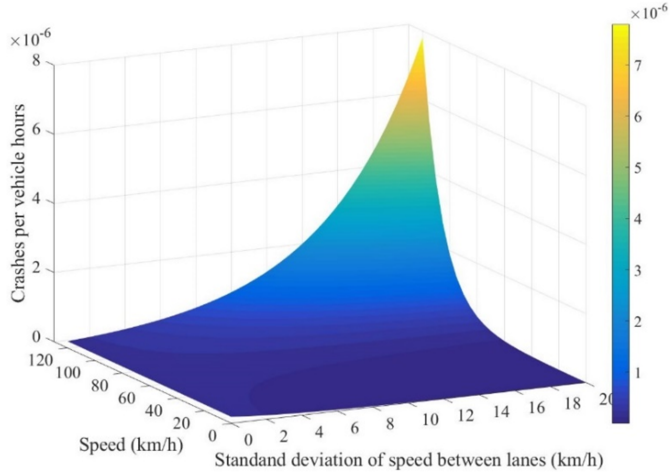
4(B) Elasticity of between-lanes speed variation across different types of crash rates for a range of speed values

Figure 4 Elasticity plots of within lane (4(A)) and between-lanes (4(B)) speed variations

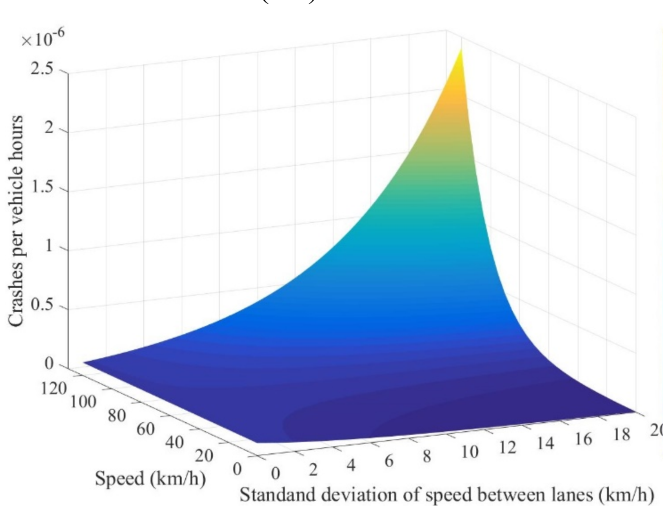
The positive relationship between crash rates and the between-lanes speed variation practically at all speed conditions in the study area possibly indicates crashes related to lane changing or overtaking manoeuvres (Ma et al., 2017; Potts et al., 2007; Wang et al., 2017). Overtaking manoeuvres tend to be more frequent under high speed conditions and if manoeuvres are combined with higher speed differences between the lanes, may trigger more side impacts. Higher between-lanes speed variation may be caused by the presence of heavy goods vehicles on the road, which tends to be slower than the rest of the traffic, especially at free-flow conditions. The LV are more likely to change lanes than HV to increase speed because it's easier to manoeuvre for LV, this increases the instances of encounter of LV with the HV. Subsequently the crashes involving multivehicle (HV and LV) increase because of the higher between-lanes speed variation. But as the present the study terms multivehicle crashes as HV crashes if at least one HV is involved in the crash, this can explain the fact that the elasticity for HV crashes is higher than for LV.



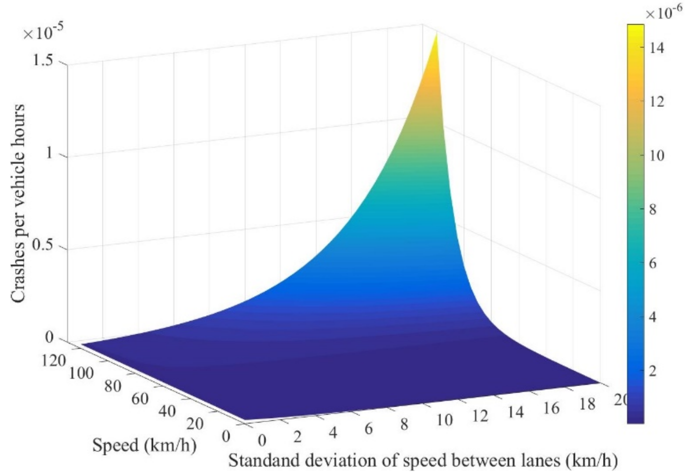
(5A) HV crashes



(5B) LV crashes



(5C) KS crashes



(5D) SL crashes

Figure 5 3d Contour graphs of crashes per vehicle hours as a function of between-lanes speed variation and speed

The model results show that the impact of speed on crashes is associated with volume and between-lanes speed variations, which complicates its interpretation. The interactions suggest that under high volume and low between-lanes speed variation, higher speeds are associated with lower crash rates. As the between-lanes speed variation increases though, higher speeds will lead to more crashes. These results extend some previous findings (Garber and Ehrhart, 2000; Kloeden et al., 2002; Tanishita and Wee, 2016) which observed that higher crash rates are observed if higher speeds are coupled with high variation in speed.

Aligning with the previous studies, it was shown that the presence of the rain increases crash risk (Abdel-aty and Pemmanaboina, 2006; Naik et al., 2016). The coefficient of the rain variable shows that the effect of rain is higher for the LV crashes when compared to the HV crashes. A possible reason behind higher crash risk for the LV during the rain could be related to the better training of heavy goods vehicle drivers in driving under rainy conditions. Surprisingly the results suggest that rain has higher effects on SL crashes than the KS crashes, which is different from the previous findings but could be explained by the lower speed during driving in rainy weather that might result in less serious crashes.

Both the datasets (0-5 minute prior of crashes and 5-10 minute prior of crashes) resulted in similar models in terms of main effects of traffic variables and therefore, for brevity only the first model was presented in this section. The main difference was observed in the weather variable “Rain”. More specifically, presence of rain was found to be negatively associated with the probability of a crash occurring in the following 5-10 minutes. This difference in results might imply more careful driving behaviour during rainy period but it can also be attributed to inaccuracies in weather data as some of the weather stations were situated quite far away from some crash locations.

6. Conclusions

This study focused on modelling the effects of speed variations on freeway crash rates by vehicle type (HV and LV) and for different levels of severity (KS and SL). Crash data were aggregated following a condition-based data aggregation approach in order to achieve better representation of time-varying variables. The crash frequencies of a three-year period on a segment of M1 motorway were modelled using Multivariate Poisson lognormal regression. The traffic variables along with their interactions and weather variables were investigated for their possible influence on crash risk. All the examined variables were found to have a statistically significant impact on crash rates and the signs of the estimated coefficients were identical for all the four examined crash types. Following are the main contributory findings of the study:

a) The study results showed that the crash rate increases with increase in the within lane speed variances at higher volume conditions.
b) The crash rate also increases with increase in the between-lane speed variances at high speed conditions.
c) The within lane speed variance is identified as a higher risk for LV crashes than the HV crashes and the chances of KS crashes are higher than the SL crashes.
d) Whereas, the between-lane speed variance is related with higher crash risk for HV crashes than the LV crashes.

Overall, the results suggest that the speed and its variations are not solely responsible for the higher crash rates, but the combination of specific traffic conditions play an important role in crash occurrences. Additionally, the results show that the speed variation should be considered in two different dimensions (between-lanes and within lane) to better interpret the crash triggering situations and to develop better and more precise safety measures.

These results could be helpful in understanding crash risk at different traffic conditions and to that end in the development of more efficient countermeasures for traffic management agencies and the road freight industry. The outcomes of this study could also contribute to the design of in-vehicle crash warning systems applicable to both commercial and private vehicles.

As this analysis focused on a busy freeway section that does not include extreme geometry, in order to generalise the outcomes of the models it could be beneficial to consider a larger and more diverse road network and to incorporate geometric data in the models. Additionally, the present study did not analyse the PDO crashes, therefore, further research should include PDO crashes, so that the results can be generalised for crashes of all severity types. The current study also does not examine differences in single and multi-vehicle crashes separately owing to the limited number of single vehicle crashes in the study area. Therefore, a future study is required to obtain more insights into the impacts of speed variations on different collision types.

Acknowledgement

We would like to thank Highways England and Mott Mac Donald for providing us access to Motorway Incident Detection and Automatic Signalling database (MIDAS) for conducting this research project.

This research is partially supported by the European Union Europe Aid-funded Project "EU-India Research & Innovation Partnership for Efficient and Sustainable Freight Transportation (REINVEST)," Contract Number: R/141842. The contents of this publication are the sole responsibility of the authors of this paper and can in no way be taken to reflect the views of the European Union.

References

Aarts, L. and Schagen, I. Van (2006) 'Driving speed and the risk of road crashes : A review', *Accid. Anal. Prev.*, 38, 215–224.

Abdel-Aty, M. A., Hassan, H. M. and Ahmed, M. (2012) 'Real-Time Analysis of Visibility Related Crashes: Can Loop Detector and AVI Data Predict Them Equally?', in *Transp. Res. Board 91st Annu. Meet.*

Abdel-aty, M. A. and Pemmanaboina, R. (2006) 'Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and ITS Traffic Data', *IEEE Trans. Intell. Transp. Syst.*, 7(2), 167–174.

Chen, F. and Chen, S., 2011. 'Reliability-based assessment of vehicle safety in adverse driving conditions.' Transp. Res. Part C,, 19(1), 156-168

Department for Transport (2011) *STATS19 road accident injury statistics*.

Farah, H. and Koutsopoulos, H. N. (2014) 'Do cooperative systems make drivers? car-following behavior safer?', *Transp. Res. part C Emerg. Technol.*, 41, 61–72.

Garber, N. J. and Ehrhart, A. A. (2000) 'Effect of Speed , Flow , and Geometric Characteristics on Crash Frequency for Two-Lane Highways', *Transp. Res. Rec. J. Transp. Res. Board*, 1717, 76–83.

Garber, N. J. and Gadiraju, R. (1989) 'Factors Affecting Speed Variance and Its Influence on Accidents', *Transp. Res. Rec. J. Transp. Res. Board*, 1213, 64–71.

Golob, T. F., Recker, W. W. and Alvarez, V. M. (2004) 'Freeway safety as a function of traffic flow', *Accid. Anal. Prev.*, 36, 933–946.

Google maps (2017), https://www.google.co.in/maps Accessed Jul. 31, 2017

Highways England (2017) *MIDAS Data* https://www.midas-data.org.uk/ Accessed Jul. 3, 2017.

Highways England (2017) *MIDAS Data*. https://www.midas-data.org.uk/ Accessed Jul. 3, 2017.

Huang, H., Zhou, H., Wang, J., Chang, F. and Ma, M. (2017) ‘Analytic Methods in Accident Research A multivariate spatial model of crash frequency by transportation modes for urban intersections’, *Anal. Methods Accid. Res.*, 14, 10–21.

Imprialou, M.-I. (2015) *Developing Accident-Speed Relationships Using a New Modelling Approach*. PhD thesis*,* Loughborough University, United Kingdom.

Imprialou, M.-I. M., Quddus, M. and Pitfield, D. E. (2016a) ‘Predicting the safety impact of a speed limit increase using condition-based multivariate Poisson lognormal regression’, *Transp. Plan. Technol.*, 39(1), 3–23.

Imprialou, M.-I. M., Quddus, M., Pitfield, D. E. and Lord, D. (2016b) ‘Re-visiting crash-speed relationships: A new perspective in crash modelling’, *Accid. Anal. Prev.*, 86, 173–185.

Kloeden, C. N., McLean, J. and Glonek, G. F. V. (2002) *Reanalysis of travelling speed and the risk of crash involvement in Adelaide South Australia*. Road safety research report (Australian Transport Safety Bureau), CR207, Australian Transport Safety Bureau.

Kockelman, K. M. and Ma, J. (2007) ‘Freeway Speeds and Speed Variations Preceding Crashes , Within and Across Lanes’, *J. Transp. Res. Forum*, 46(1), 43–61.

Kweon, Y. and Kockelman, K. M. (1996) ‘Safety Effects of Speed Limit Changes Use of Panel Models , Including Speed , Use , and Design Variables’, *Transp. Res. Rec. J. Transp. Res. Board*.

Lee, C., Hellinga, B. and Saccomanno, F. (2003) ‘Real-time crash prediction model for application to crash prevention in freeway traffic’, *Transp. Res. Rec. J. Transp. Res. Board*, (1840), 67–77.

Lemp, J.D., Kockelman, K.M. and Unnikrishnan, A., 2011. ‘Analysis of large truck crash severity using heteroskedastic ordered probit models’ Accid. Anal. Prev. 43(1), 370-380.

Levy, D. T. and Asch, P. (1989) ‘Speeding, coordination, and the 55-mph limit: Comment’, *Am. Econ. Rev.*, 79(4), 913–915.

Li, Y., Zheng, Y., Wang, J., Kodaka, K. and Li, K. (2018) ‘Crash probability estimation via quantifying driver hazard perception’, *Accid. Anal. Prev.*, 116(May 2017), 116–125.

Lord, D. and Mannering, F. (2010) ‘The statistical analysis of crash-frequency data : A review and assessment of methodological alternatives’, *Transp. Res. Part A*, 44(5), 291–305.

Ma, J. (2006) ‘Bayesian multivariate Poisson-Lognormal regression for crash prediction on rural two-lane highways’. PhD thesis, University of Texas, Austin

Ma, Z., Zhang, H., Steven, I., Chien, J., Wang, J. and Dong, C. (2017) ‘Predicting expressway crash frequency using a random effect negative binomial model: A case study in China’, *Accid. Anal. Prev.*, 98, 214–222.

Martensen, H. and Dupont, E. (2013) ‘Comparing single vehicle and multivehicle fatal road crashes : A joint analysis of road conditions , time variables and driver characteristics’, *Accid. Anal. Prev.*, 60, 466–471.

MetOffice (2016) *UK climate - Historic station data*. Accessed Jul. 2, 2016

Naik, B., Tung, L., Zhao, S. and Khattak, J. (2016) ‘Weather impacts on single-vehicle truck crash injury severity’, *J. Safety Res.*, 58, 57–65.

Park, E. and Lord, D. (2007) ‘Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity’, *Transp. Res. Rec. J. Transp. Res. Board*, (2019), 1–6.

Potts, I., Harwood, D. and Richard, K. (2007) ‘Relationship of lane width to safety on urban and suburban arterials’, *Transp. Res. Rec. J. Transp. Res. Board*, (2023), 63–82.

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(1), 27–45.

Shi, Q., Abdel-Aty, M. and Yu, R. (2016) ‘Multi-level Bayesian safety analysis with unprocessed Automatic Vehicle Identification data for an urban expressway', *Accid. Anal. Prev.*, 88, 68–76.

Solomon, D. (1964) 'Accidents on main rural highways related to speed, driver, and vehicle'.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003) 'WinBUGS user manual'.

Strömgren, P. and Lind, G. (2016) 'Harmonization with variable speed limits on motorways', *Transp. Res. Procedia*, 15, 664–675.

Stuster, J. (2004) *Aggressive Driving Enforcement: Evaluations of Two Demonstration Programs*.

Tanishita, M. and Wee, B. Van (2016) 'Impact of vehicle speeds and changes in mean speeds on per vehicle-kilometer traf fi c accident rates in Japan', *IATSSR*, 0–5.

Taylor, M. C., Lynam, D. A. and Baruya, A. (2000) 'The effect of drivers' speed on the frequency of road accidents', *Transp. Res. Lab.*, 56.

Wang, L., Abdel-Aty, M. and Lee, J. (2017) 'Safety analytics for integrating crash frequency and real-time risk modeling for expressways', *Accid. Anal. Prev.*, 104, 58–64.

Wang, L., Abdel-aty, M., Shi, Q. and Park, J. (2015a) 'Real-time crash prediction for expressway weaving segments', *Transp. Res. Part C*, 61, 1–10.

Wang, L., Shi, Q. and Abdel-Aty, M. (2015b) 'Predicting crashes on expressway ramps with real-time traffic and weather data', *Transp. Res. Rec. J. Transp. Res. Board*, (2514), 32–38.

Wang, X., Yu, R. and Zhong, C. (2016) 'A field investigation of red-light-running in Shanghai, China', *Transp. Res. Part F Traffic Psychol. Behav.*, 37, 144–153.

Wang, X., Zhou, Q., Quddus, M., Fan, T. and Fang, S. (2018) 'Speed, speed variation and crash relationships for urban arterials', *Accid. Anal. Prev.*, 113(November 2016), 236–243.

Wei, Z., Xiaokun, W. and Dapeng, Z. (2017) 'Truck crash severity in New York city : An investigation of the spatial and the time of day effects', *Accid. Anal. Prev.*, 99, 249–261.

Xu, C., Liu, P. and Wang, W. (2016) 'Evaluation of the predictability of real-time crash risk models', *Accid. Anal. Prev.*, 94, 207–215.

Xu, P., Huang, H. and Dong, N. (2018) 'The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research', *J. Traffic Transp. Eng. (English Ed.*, 5(1), 73–82.

Yu, R. and Abdel-Aty, M. (2014) 'Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data', *Accid. Anal. Prev.*, 62, 161–167.

Yu, R., Quddus, M., Wang, X. and Yang, K. (2018) 'Impact of data aggregation approaches on the relationships between operating speed and traffic safety', *Accid. Anal. Prev.*, 120, 304–310.

Zou, W., Wang, X. and Zhang, D., 2017. 'Truck crash severity in New York city: an investigation of the spatial and the time of day effects' *Accid. Anal. Prev.* 99, 249-261.

Zhu, X. and Srinivasan, S., 2011. 'A comprehensive analysis of factors influencing the injury severity of large-truck crashes' *Accid. Anal. Prev*, 43(1), 49-57.