Accident Analysis and Prevention 118 (2018) 221–235

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions

Loukas Dimitriou[a,*], Katerina Stylianou[a], Mohamed A. Abdel-Aty[b]

[a] Lab. for Transport Engineering, Dept. of Civil and Environmental Engineering, University of Cyprus, 91, Aglanzias Av., 2111 Nicosia, Cyprus
[b] Dept. of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, 32816-2450, United States

A R T I C L E I N F O

Keywords:
Rear-end crashes
Crash potential
Locational analysis
Near-crash behavior
Multinomial Logit model

A B S T R A C T

Rear-end crashes are one of the most frequently occurring crash types, especially in urban networks. An understanding of the contributing factors and their significant association with rear-end crashes is of practical importance and will help in the development of effective countermeasures. The objective of this study is to assess rear-end crash potential at a microscopic level in an urban environment, by investigating vehicle-by-vehicle interactions. To do so, several traffic parameters at the individual vehicle level have been taken into consideration, for capturing car-following characteristics and vehicle interactions, and to investigate their effect on potential rear-end crashes. In this study rear-end crash potential was estimated based on stopping distance between two consecutive vehicles, and four rear-end crash potential cases were developed. The results indicated that 66.4% of the observations were estimated as rear-end crash potentials. It was also shown that rear-end crash potential was presented when traffic flow and speed standard deviation were higher. Also, locational characteristics such as lane of travel and location in the network were found to affect drivers' car following decisions and additionally, it was shown that speeds were lower and headways higher when Heavy Goods Vehicles lead. Finally, a model-based behavioral analysis based on Multinomial Logit regression was conducted to systematically identify the statistically significant variables in explaining rear-end risk potential. The modeling results highlighted the significance of the explanatory variables associated with rear-end crash potential, however it was shown that their effect varied among different model configurations. The outcome of the results can be of significant value for several purposes, such as real-time monitoring of risk potential, allocating enforcement units in urban networks and designing targeted proactive safety policies.

1. Introduction

Traffic safety performance is relied on measuring crash frequency and thus road safety analysis has traditionally been undertaken using crash data. It has been shown that rear-end crashes are one of the most frequently occurring types; 3.98 million rear-end crashes were reported in the U.S. roadways in 2013, which accounted for the 34% of the total reported crashes, and resulted in 2200 fatalities (National Safety Council, 2015). From this perspective, the identification of contributing factors to rear-end crashes, such as human behavior, roadway characteristics or environmental conditions, and the investigation of their impact on these type of crashes is of practical importance in order to develop effective countermeasures for improving network safety levels. Rear-end crashes are defined as a type of crash where the rear side of a vehicle is hit by the front side of the following vehicle (Singh 2003).In car-following situations vehicles interact with adjacent vehicles and it has been identified that crashes are potentially raised when interactions between vehicles become unstable (Oh and Kim, 2010).Rear-end crashes are usually attributed to insufficient spacing between consecutive vehicles; the proper distance kept between consecutive vehicles should provide the driver of the following vehicle time to recognize a potential hazard ahead and engage in an appropriate action (such as braking, or lane changing) if needed. Generally, drivers can minimize the risk of being involved in a rear-end crash by maintaining an adequate space cushion that is appropriate for the driving conditions (Abdel-Aty and Abdelwahab, 2004).

Traditional rear-end crash analysis has focused on identifying factors that contribute to rear-end crash counts by developing models using aggregated traffic data, geometric characteristics and crash data collected over a prolonged time period. However, in addition to long periods of time required for collection, crash data are often faced with reliability, quality and quantity problems. Therefore, the use of alternative indicators to crashes, known as surrogate measures, have been widely used in road safety analysis to reflect road safety. In this point of

* Corresponding author.
E-mail addresses: lucdimit@ucy.ac.cy (L. Dimitriou), stylianou.p.katerina@ucy.ac.cy (K. Stylianou), M.Aty@ucf.edu (M.A. Abdel-Aty).

https://doi.org/10.1016/j.aap.2018.02.024
Received 27 April 2017; Received in revised form 20 December 2017; Accepted 23 February 2018
Available online 02 March 2018
0001-4575/

view, identifying rear-end potentials with the use of surrogate measures in the network which could potentially lead to rear-end crashes, would be a proactive measure toward mitigating transportation problems in the network.

The objective of this study is to estimate real-time rear-end crash potential in urban networks, and to identify the effect of various traffic parameters on this potential. This paper introduces a thorough investigation of car-following decisions (speed and time headway) and vehicle-by-vehicle interactions through disaggregated loop detector data, aiming to systematically reveal the underlying patterns and the contributing factors of possible rear-end crashes. The scope of this paper stands for the extrapolation and adaptation of the extensive and thorough research related to crash risk potential that has been mainly focusing on highways, to the more complex case of signalized and congested urban networks by using real data typically collected in Traffic Control Centers (here, from the urban network of Nicosia, Cyprus). Several traffic parameters at the individual vehicle level, locational characteristics and operational conditions are introduced, and assessed. Traffic volume and vehicle speed are collected from inductive loop detectors, along with individual vehicle information such as time of arrival, vehicle type and lane of travel. Rear-end crash potential is estimated based on stopping distance calculation and subsequently four distinctive cases of rear-end crash potentials are developed (a) no rear-end crash potential, (b) rear-end crash potential when the leading vehicle is stationary; (c) rear-end crash potential when the leading vehicle is braking; (d) rear-end crash potential where the distance kept is not sufficient to avoid a collision with the leading vehicle in the next 10 s if no action is taken. An exploratory analysis is employed to observe the effect of location (lane width, lane of travel, location in the network), traffic parameters (traffic flow, average speed, moments of speeds) and vehicle configuration (vehicle type, leading/following type) on rear-end crash potential. A preliminary statistical investigation of the above variables with respect to driving behavior and crash potential is followed by a model-based analysis. In particular, alternative Multinomial Logit (MNL) regression models are developed to identify the significance of the aforementioned parameters on the rear-end crash potential, proving detailed and robust functional form explaining risk driving behavior. The extensive datasets used in this study are based on more than 400,000 vehicles and the experimental setting in terms of the model structure, selection and analysis, offers valuable results for monitoring and policy-making purposes in realistic urban traffic systems. The results of the above described analytical approach contribute important exploratory and explanatory insights on both the design as well as the operational traffic characteristics that effect crash risk potential in congested signalized urban networks, possibly valuable for controlling/management purposes, an objective that exceed the scope of the present study.

The structure of the paper starts with a presentation of the relevant research background on rear-end crash investigation. Then the methodological approach is presented, offering the analytical framework that the following experimental results are based on. In the final section, the results are discussed in detail while several directions for future research are given.

2. Background

Significant research effort has been undertaken to analyze the characteristics and contributing factors of rear-end crashes. A number of studies investigating rear-end crashes, with different perspectives (e.g. crash propensity, injury severity, car-following decisions etc.), and modelling techniques has been presented in recent years. Kim et al. (2007), developed a modified negative binomial regression model to estimate rear-end risk using Washington State data, reporting among others that urban areas have an increased rear-end crash probability. Yan et al., (2005) used binary logistic models to investigate rear-end crashes at signalized intersections utilizing Florida traffic accident data, in order to identify the risk factors related to traffic environment, driver characteristics and vehicle type. Also, the generalized estimating equations were used by Wang and Abdel-Aty (2006) to model rear-end frequencies at signalized intersections, showing high correlations between the longitudinal and spatially correlated rear-end crashes. Ahmed and Abdel-Aty (2013), used a machine learning technique, namely, Stochastic Gradient Boosting (SGB), in order to develop a framework for real-time risk assessment on a freeway in Colorado. The general estimates system (GES) databases were used by Abdel-Aty and Abdelwahab (2004) to study the effect of the lead vehicle's size on the rear-end crash configuration, by developing nested logit models in order to estimate the probabilities of four types of rear-end collisions. Shi and Abdel-Aty (2015), utilized a Microwave Vehicle Detection System (MVDS) consisting of 275 detectors and developed a methodology of data mining and Bayesian inference techniques to implement real-time crash prediction models.

Furthermore, several studies have investigated the connection between crash risk and operational characteristics (mainly traffic flow). A study conducted by Golob and Recker (2004), examined how the types of urban freeway accidents are related to traffic flow using accident data and corresponding loop detector data. This study found that median traffic speed and temporal variation in speed are strongly related to the type of collision. Christoforou et al. (2011) proposed a framework of multivariate probit models to assess the effects of traffic variables on crash type, and findings showed that crash type can almost exclusively be defined by the prevailing traffic conditions shortly before the crash occurrence. Xu et al. (2013), utilized the sequential logit model to link crash likelihood at different severity levels, and showed that traffic flow characteristics contributing to crash occurrence vary across different crash severity levels. It is evident that traditionally research has focused on the macroscopic measurements of total counts of crashes and total flows to develop safety functions. However, an increased interest over time has been on individual vehicle interactions and behavioral analysis relationship with crashes. Hu et al. (2004), proposed a probabilistic model to predict traffic accidents based on 3-D model-based vehicle tracking. Hourdos et al. (2006), established a relationship between quickly evolving real-time traffic conditions and crash likelihood, by utilizing a state-of-the-art infrastructure which allowed the video capture of 110 live crashes. Oh and Kim (2010) proposed a methodology to estimate rear-end crash probabilities by using vehicle trajectories obtained from video surveillance mechanisms. In this study a statistical methodology was developed to estimate rear-end crash potential using time-to-collision (TTC) measures. Kim et al. (2016) explored the association of crash propensity and micro scale driving behavior by integrating rear-end crash data experienced in a freeway with micro-scale driving behavioral data gathered by an in-vehicle sensing system. The findings of this study showed that 85% of all rear-end crashes occurred within 2000 feet of the on-ramp gore and a strong association between rear-end crash rates and the propensity of hard deceleration exists. These studies concretely offer evidence on the relationship among traffic conditions and crash risk, valuable both for strategic planning as well as for real-time applications. In the last decade real-time crash prediction has also been an active field in safety research. In real-time safety analysis, the use of dynamic traffic data enables researchers to investigate crash likelihood over a short period of time. Real-time analysis assumes that the occurrence of a crash is because of a short-term turbulence right before the crash, and this turbulence can be identified by real-time data (Wang et al., 2017a,b). For example Abdel-Aty et al. (2012) examined the relationship between visibility-related crashes and real-time traffic data using two data sources, namely loop detectors and Automatic Vehicle Identification (AVI) sensors. The study showed that the average speed along with the coefficient of variation in speed at 5–10 min prior to a crash significantly affects the likelihood of visibility related crashes. Similarly, Ahmed et al. (2014) used real-time weather data collected from airport stations to investigate the viability of this type of data in real-time crash risk assessment in locations with recurrent fog problems, and showed that these are good predictors of increased risk on highways. In a similar study Wu et al. (2017) developed a Crash Risk Increase Indicator (CRII) to explore the differences of crash risk between fog and clear weather conditions and showed that the proposed indicator performed well in evaluating the increase of crash risk under fog conditions, as well as that 5-minute volume during fog conditions and the lane position are also important factors of crash risk increase. Another study, by Theofilatos (2017), investigated the accident likelihood and severity utilizing real-time traffic and weather data collected from urban arterials in Athens. Contrary to the previously mentioned studies, here weather parameters were not found to have a direct influence on accident likelihood or severity. Other results of the study showed that variations in traffic significantly influence accident occurrence.

Wang et al. (2015) presented a multilevel Bayesian logistic regression model for crashes at expressway weaving segments using crash, geometric, Microwave Vehicle Detection System (MVDS) and weather data, and showed that the speed at the beginning of the weaving segment, the speed difference between the beginning and the end of the segments and the logarithm of volume present a significant impact on the crash risk of the following 5–10 min for weaving segment. This study also showed that wet pavement condition also increases crash ratio. Similarly, Kwak and Kho (2016) developed real-time crash risk perception models for different segment types and traffic flow states in Korean expressways. The results of the study showed that traffic flow characteristics contributing to crash risk differ by segment type and traffic flow state. More specifically, during uncongested traffic, high density and disruptions were shown to be the main contributing factors to crash risk, whereas during congested traffic speed variation and queue formations were shown to increase crash risk. A recent study analyzed real-time risk for expressway ramps using not only traffic and geometric characteristics but also socio-demographic and trip generation predictors (Wang et al., 2017a,b), and showed that socio-demographic and trip generation parameters have a significant impact on the real-time crash risk.

Another approach was by Xu et al. (2015) who developed a real-time crash risk model with limited data in China by using Bayesian meta-analysis and Bayesian inference approach. The meta-analysis combined the results of previous studies regarding real-time traffic variables which were found to increase crash risk, which were later used as informative priors in the Bayesian inference process. A different approach was by Fang et al. (2016) who proposed a new method to model the real time crash likelihood based on loop data through schematic eigenvectors. The results showed that both the eigenvectors and eigenvalues significantly impact accident likelihood and the proposed logistic model has the advantage of avoiding multicollinearity. Generally the findings of the aforementioned studies show that the use of dynamic, real-time data can provide adequate information regarding crash precursors. Despite this large body of safety modeling research, a common issue is that crash data quantity and quality are often unsatisfactory and not suitable for analysis, therefore, several research attempts have been undertaken to overcome the shortcomings associated with crash data. Multiple authors have advocated the use of alternative indicators, also known as surrogate measures, which reflect road safety. For example, Oh et al. (2006) proposed a methodology of incorporating a traffic conflict technique and an advanced inductive loop detector surveillance system, where a fuzzy-clustering algorithm was employed to estimate rear-end crash risk based on the loop detector data. The proposed framework was implemented on a freeway section and it was found that 35% of the car-following events passing the loop detector station, based on safe stopping distance, would be identified as unsafe. Duan et al. (2013) investigated safe distance headways between consecutive vehicles using a simulator experiment and proposed a conceptual model to predict how drivers maintain their distance headway. Wang et al. (2016) investigated drivers' collision avoidance behaviors when exposed to rear-end collision risks. In this study a driving simulator was used to examine the effects of differing levels of lead vehicle decelerations and headways on the following drivers' avoidance behavior and the results showed that as situational urgency increased, drivers released the accelerator and braked to maximum quicker. Another approach was by Weng and Meng (2012) who used time-to-collision (TTC) as a surrogate safety measure to estimate rear-end crash potential in the work zone merging areas. Additionally, Meng and Qu (2012) applied an inverse Gaussian regression model to TTC data, in order to estimate rear-end vehicle crash frequencies in road tunnels. Li et al. (2014) developed a rear-end collision risk index for evaluating the risk related to kinematic waves near freeway recurrent bottlenecks, utilizing aggregated traffic data from loop detectors. This study reported that the likelihood of a rear-end collision is highest when traffic approaching from upstream is at capacity state while traffic downstream is highly congested. Ghanipoor Machiani and Abbas (2016) combined real-time safety analysis techniques and surrogate safety measures to develop a novel real-time safety measure for signalized intersections, named safety surrogate histograms (SSH), and showed that this new measure can be used as complementary to existing surrogate measures.

The aforementioned studies provide an insightful understanding of rear-end crash occurrence and their contributing factors mainly on highways. However, it can be identified that the research efforts so far are limited in investigating this phenomenon in urban environment. Hence, the present study aims at the investigation of rear-end crash potential with the use of surrogate safety measures in dynamic urban traffic networks. To our knowledge the present study is among the first to suitably extend the investigation of near-crash phenomena in safety analysis within the urban environment.

3. Methodological approach

Vehicles travelling in the network interact on the same and adjacent lanes, generating car-following and lane changing events respectively. In order to develop a methodology for evaluating rear-end crash potential in urban networks, this study utilizes disaggregate vehicle information (vehicle-by-vehicle data) and car following concepts. In this study, rear-end crash potential is estimated based on stopping distance utilizing car-following decisions (speed and time headway). Based on this estimation, four cases of rear-end crash potentials are developed, providing an understanding of how car-following decisions could potentially lead to a rear-end crash.

Given the above estimation, the effect of traffic measures, locational conditions and leading vehicle size is then investigated. In the concept of the urban environment, it was assumed that traffic measures, such as flow, average speed and speed moment (i.e. standard deviation), would influence car-following decisions and therefore could result in a rear end crash. It was also assumed that location (intersections, links, etc.), as well as lane of travel (based on carriageway design) would influence rear-end crash potential. Lastly, traffic composition and more specifically leading vehicle size is assumed as a contributing factor to rear-end crashes, since drivers' attention is mainly devoted to the leading vehicle when driving.

3.1. Rear-end crash potential estimation technique

In order for a collision to be avoided, the stopping distance of the leading vehicle should be greater than the stopping distance of the following vehicle to prevent a collision, if the vehicles' course and speeds remain unchanged. Research focused on driving behavior in different states of traffic, has shown that a major contributor to rear-end crashes is a short headway, which does not allow for the following driver to react in an appropriate manner in the case of a sudden break of the leading vehicle (Taieb-Maimon and Shinar, 2001).The proposed rear-end crash estimation here is based on car-following decisions (speed and time headway) and is restricted to two-vehicle crashes

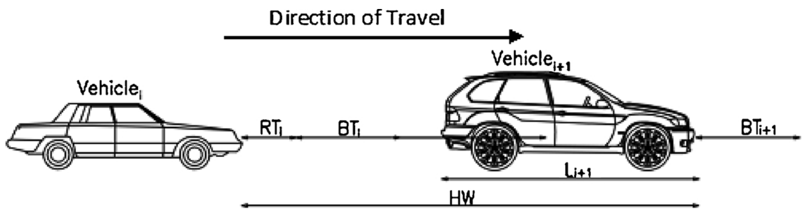

Fig. 1. Notations Used in the Car Following Setup.

proceeding straight at intersections, extending the preliminary results of Stylianou et al. (2017).

In an example situation, assume that two consecutive vehicles are travelling in the same lane. Let Vehicle $i+1$be the leading vehicle with a speed$V_{i+1}$and vehicle i be the following vehicle with a speed V_i, and between the two vehicles exists a time headway, HW. By using this microscopic analytical setting and following the procedures described by American Association of State Highway and Transportation Officials-AASHTO (2011) guidelines for estimating safe stopping distances, rear-end crash potential can be estimated. The explanatory variables include speed, time headway, reaction time, vehicle deceleration rate, brake time and vehicle length. As so, the rear-end crash potentials are derived as shown in Eqs. (1a)–(1d) and in Fig. 1:

$$\text{Case\_0: no rear-end crash potential} \tag{1a}$$

$$\text{Case\_1: } HW - L_{i+1} - RT - BT_i < 0 \tag{1b}$$

$$\text{Case2: } HW - L_{i+1} + BT_{i+1} - RT - BT_i < 0 \tag{1c}$$

$$\text{Case3: } S_i - S_{i+1} - S_{HW} < 0 \tag{1d}$$

where:

HW: Temporal Headway (s)
L: Vehicle length (4.5 m for private vehicles and 8.0 m for HGV)
RT: Driver's Perception and Reaction Time (= 3 s)
BT: Brake Time (s) calculated as $\frac{V}{dr}$
V: Speed (km/hr)
dr: Deceleration Rate of Vehicle (= 0.5g m/s^2)
S: Distance Travelled in time t (km) calculated as Vt, t = 10s
S_{HW}: Distance Headway, calculated as $V_i \times HW$

The above rear-end crash potential estimation represents four cases derived by whether drivers would engage in a potential rear-end crash based on stopping distance, if their course and speed remain unchanged. As such, Case_0 indicates the case where a safe distance is kept from the leading car and therefore there is no rear-end crash potential; Case_1 indicates the case where the distance kept from the leading vehicle is not sufficient to stop up to the leading vehicle's location (e.g. something is dropped from it, or suddenly avoids an obstacle); Case_2 indicates the case where there is not sufficient distance for the following vehicle to stop when the leading vehicle is breaking; and Case_3 indicates the case where the distance kept is not sufficient to avoid collision with the leading vehicle in the next 10 s if no action is taken by both vehicles. Based on the above classification of driving style based on leading-following records, it can be assumed that cases 1 to 3 reflect different rear-end crash potential states.

Rear-end crashes are more common at signalized intersections as the signal change increases the diversity of actions taken by the driver (Yan et al., 2005). The proposed rear-end crash potential estimation here, uses instantaneous individual speeds (and time headways) and as such crash potential remains invariant in regard to signal effect, suggesting that even at the yellow or red phase of the traffic lights, individual vehicle speeds (and time headways) are considered to be adequate indicators in order to capture an instantaneous rear-end crash potential. The investigation based on the above four driving cases, could be valuable for identifying the mix of driving styles in terms of rear-end crash potential. Though, in order to better understand the interrelations among the parameters that influence driving style, a model-based analysis is also conducted. The details of this type of analysis are presented in the next section.

3.2. Model based analysis – significant variable identification

In order to better understand the multiple interrelations that affect the driving styles that are captured by the disaggregate information on leading-following vehicle records, a suitable model could be used. In brief, here the Multinomial Logit model (MNL) a commonly used unordered discrete modeling approach, is adopted to identify the statistically significant variables in rear-end crash potential. The probabilities of the four rear-end crash potentials based on drivers' car following decisions are modelled. By employing a MNL it is able to examine the possible correlation among the unobserved effects of the four rear-end crash potential cases. Given that each individual has a feasible choice set (C_n) then there is$J_n \le J$ number of feasible choices for each individual. Hence, the probability of individual n choosing choice i in the choice set C_n is given by (Ben-Akiva and Lerman (1985)):

$$P_n(i) = P(X_n\beta_i + e_{in} \ge X_{nj} + e_{jn}), \quad \forall j \in C_n \tag{1}$$

where$i$ denotes the choice, n denotes the individual, $\boldsymbol{X}$ is a vector of individual specific characteristics, $\boldsymbol{\beta}$ is a vector of estimable coefficients and e is the error term. Assuming that e_n is logistically distributed, the Multinomial Logit Model (MNL) is derived as follows:

$$P_n(i) = \frac{e^{\beta_i X_{in}}}{\sum_{j=1}^{J} e^{\beta_j X_{jn}}} \tag{2}$$

One of the most common concerns for the MNL is the Independence of Irrelevant Alternatives property (IIA), which states that for a specific individual the ratio of the choice probabilities of any two alternatives is entirely unaffected by any other alternative (Ben-Akiva and Lerman, 1985). The MNL specifications regarding the IIA have been well documented in other studies (see Abdel-Aty and Abdelwahab, 2004 for example) and are therefore not repeated. Here, the four rear end crash potential cases are considered to be independent since the data corresponds to instantaneous observations of different pairs of vehicles. Though, alternative model settings could be employed (e.g. incorporating information of more than two vehicles into the model), but ultimately the decisions of two (leading-following) drivers are those that may lead to a rear-end crash. In the next section, the results from a realistic large-scale experiment on the estimation of crash potentials based on the earlier described methodological approach are presented.

4. Application and results

4.1. Data collection and preparation

The experimental setup for testing the earlier proposed methodological approaches is based on a realistic large-scale urban area, namely, a data collection from the urban network of Nicosia, Cyprus. Detailed disaggregate traffic data from 15 Inductive Loop Detectors (ILD) is used (Fig. 2), where individual vehicle records are collected. Each loop detector covers either one or two lanes in each direction (depending on the carriageway design), totaling 46 lanes. The total volume collected for one typical day was 409,920 observations. Loop

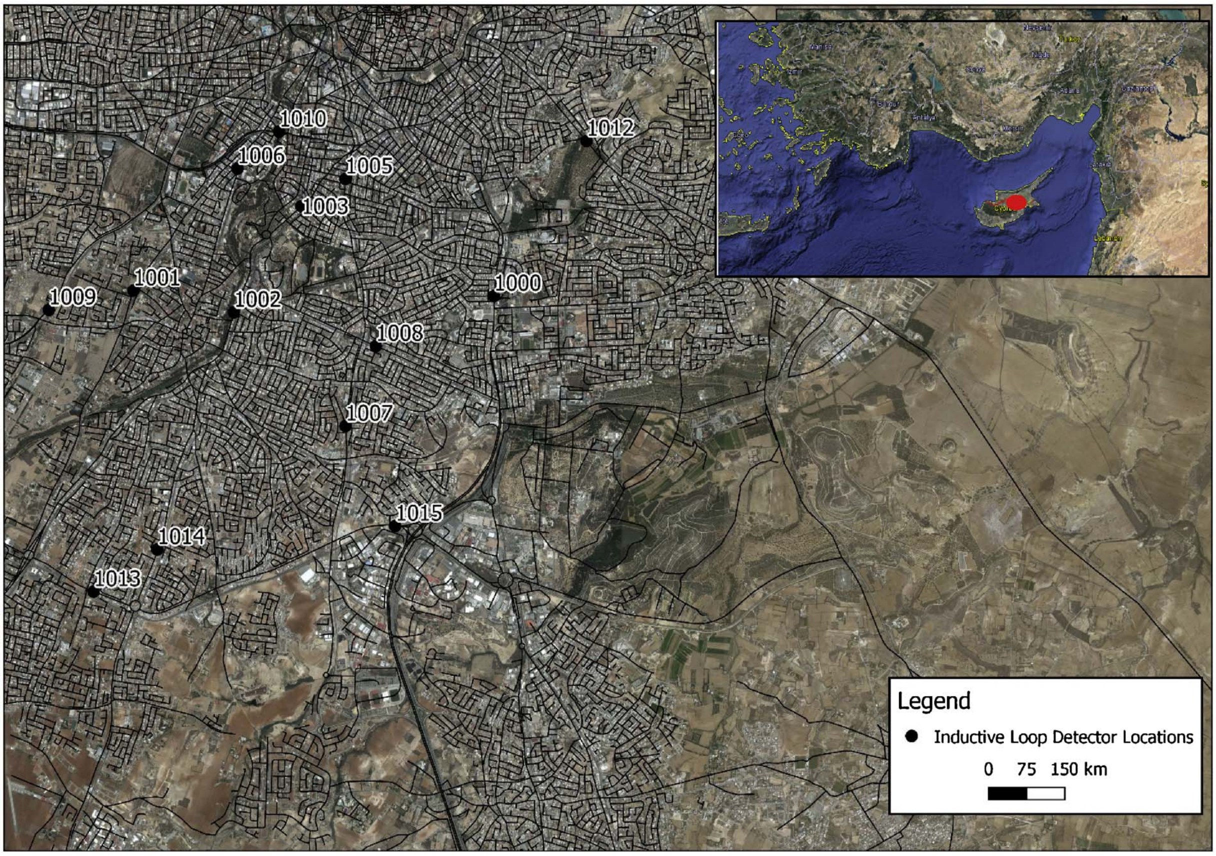

Fig. 2. Inductive Loop Detector Locations.

Table 1
Inductive Loop Detector Location Characteristics.

ILD Number	Carriageway Design	Number of Lanes covered by ILD	Lane Width (m)	Distance from upstream major Intersection (m)		Distance from downstream major Intersection (m)		Distance from nearest upstream minor Intersection (m)		Distance from nearest downstream minor Intersection (m)	
1000	Dual (a)	4	3.50	S	255	S	255	P	155	P	115
1001	Single	2	3.15	S	350	S	635	P	63	P	32
1002	Dual (b)	4	3.30	S	400	S	280	P	86	P	20
1003	Single	2	3.15	R	410	S	130	P	200	P	10
1005	2 + 1 (a)	3	3.30	S	520	S	340	P	35	P	8
1006	Single	2	3.50	S	85	S	150	P	50	P	80
1007	Single	2	3.30	S	315	S	430	–	–	–	–
1008	2 + 1 (b)	3	3.30	S	160	S	140	–	–	–	–
1009	Dual (a)	4	3.30	S	115	S	550	–	–	–	–
1010	Dual (a)	4	3.30	S	140	S	66	–	–	–	—
1012	Single	2	3.15	S	247						
1013	Dual (a)	4	3.30	R	670	R	380	P	65	–	–
1014	Single	2	3.15	R	550	S	1200	P	100	P	50
1015	Dual (a)	4	3.30	S	650	R	145	P	40	–	–

Dual (a): Dual Carriageway with built median island.
Dual (b): Dual Carriageway with painted median.
2 + 1 (a): One lane per direction with an additional dedicated right turning lane.
2 + 1 (b): One lane per direction with an additional dedicated right turning lane and built median island.
S: signalized intersection.
R: Roundabout.
P: Priority Intersection.

detectors cover locations near signalized intersections, near roundabouts or midway in links. The location and the geometric characteristics of the location for each ILD are presented in Table 1.

The disaggregate, individual vehicle by vehicle (VBV) data, obtained from the ILDs, contain valuable information about several parameters. Spot speed and time headways were the primary variables used to explore vehicle interactions, as the combination of these two parameters reflect driving style.

Other traffic parameters were also used in order to capture other traffic characteristics effects, such as traffic flow, average speed and speed standard deviation. Following previous research findings on crash prediction, it has been shown that 5-minute aggregation of the

Table 2
Explanatory Variables and summary statistics.

Variable	Description	Statistics		
		Total	Mean	Std. Dev
Volume	Total number of vehicles in 24hours	409,920		
Flow	Number of vehicles aggregated in 5 minute intervals (veh/5min)		51.10	16.08
Spot Speed (Speed)	Individual vehicle speed (km/h)		48.67	16.10
Average Speed (Av_Speed)	Average vehicle speed aggregated in 5 minute interval (km/h)		50.00	15.90
Speed standard deviation (STD_Speed)	Vehicle speed standard deviation aggregated in 5 minute interval		9.14	4.69
Headway	Temporal Headway between two consecutive vehicles (s)		8.87	27.57
Location	Location of Loop Detector in the network	15 ILDs		
Lane	Lane of travel for each individual vehicle (Lane_1, Lane_2, Lane_3, Lane_4)	46 Lanes		
Vehicle Type	Type of vehicle (1: Passenger Car, 2: HGV)	Type1:399,929 Type2: 9,991		

data provides an acceptable accuracy and omits information noise (Abdel-Aty et al., 2012, Xu et al., 2013, Oh et al., 2001).As so, additionally to the individual information (speed and headway), the average speed, traffic flow and speed standard deviation in 5-minute aggregation were developed. Other measures used in the analysis, such as lane of travel and vehicle type were directly obtained from the ILDs. Table 2 summarizes the variables used in this study and their basic statistics.

4.2. Modelling results and discussion

4.2.1. Rear-end crash potential estimation

According to Eq. 1(*a–d*) the classification of traffic into the four leading-following cases can be estimated, which in turn reflect the rear-end crash potential for each location. It should be mentioned that the rear-end crash potential cases are mutually exclusive. In Fig. 3, the results of the classification of all vehicles' records for one day from all count locations in the urban network are presented. In total, it is shown that 33.6% of the observed car-following events were classified as Case_0, 19.4% as Case_1, 46.3% as Case_2 and the remaining 0.7% were classified as Case_3. These values represent that 66.4% of total observed car following events passing the loop detectors would be identified as potentially unsafe driving.

Other than the statistics of the rear-end crash potentials for a typical day, the daily pattern of the rear-end crash potential throughout the day was also investigated. A more detailed investigation of traffic classification per traffic lane is presented in Fig. 4, in which a representative selection of three distinctive locations in the examined network is offered. Loop detector 1000 is located at a dual-carriageway, in the middle of a link, 1008 is located upstream of a major intersection and covers three lanes and 1012 is located in a single carriageway downstream of a significant intersection. It can be observed that from midnight until around 6 a.m., the dominant rear-end crash potential case was Case_0. This is mainly due to the fact that at this time period the flow is low and as such limited car-following settings occur. Between morning peak hours (07:00) and afternoon peak hours (18:00) the dominant case was Case_2. A similar pattern is presented with Cases 1 and 3, where occurrences of these cases are low during nighttime but increase during working hours. The results presented, provide evidence that rear-end crash risk exposure is related to flow states, and more specifically higher flows could potentially present higher instances of rear-crash potential. Overall, it is interesting to note the fact that the percentage of risky driving style is significant during rush hours (for example reaching levels around 30% for Case_2), which could be a surrogate measurement for possible rear-end crashes that suggests in-depth investigation (which follows) for better understanding the factors of such risk-prone driving behavior.

Since the above classification is based on assumptions regarding the drivers' perception and response times as well as vehicles deceleration rates, a sensitivity analysis (SA) with respect to the values of these two variables should be performed. The rear-end crash potential analysis was performed again, for three values of Perception Reaction Time (RT:

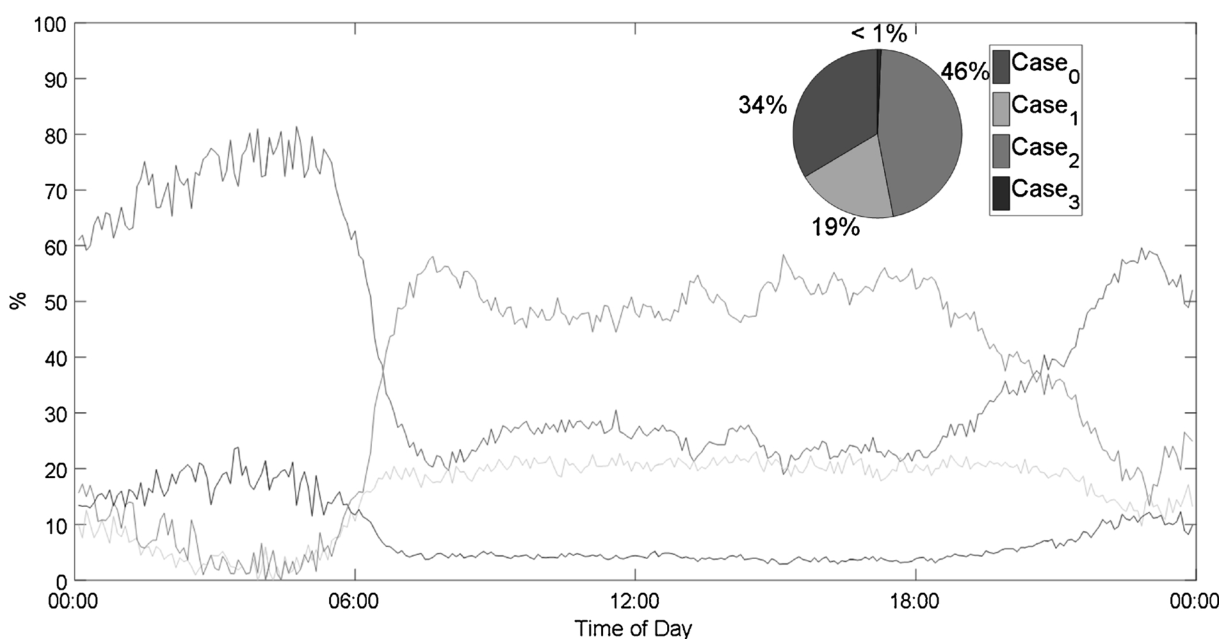

Fig. 3. Rear-end crash potential analysisfor all vehicle records throughout a typical day.

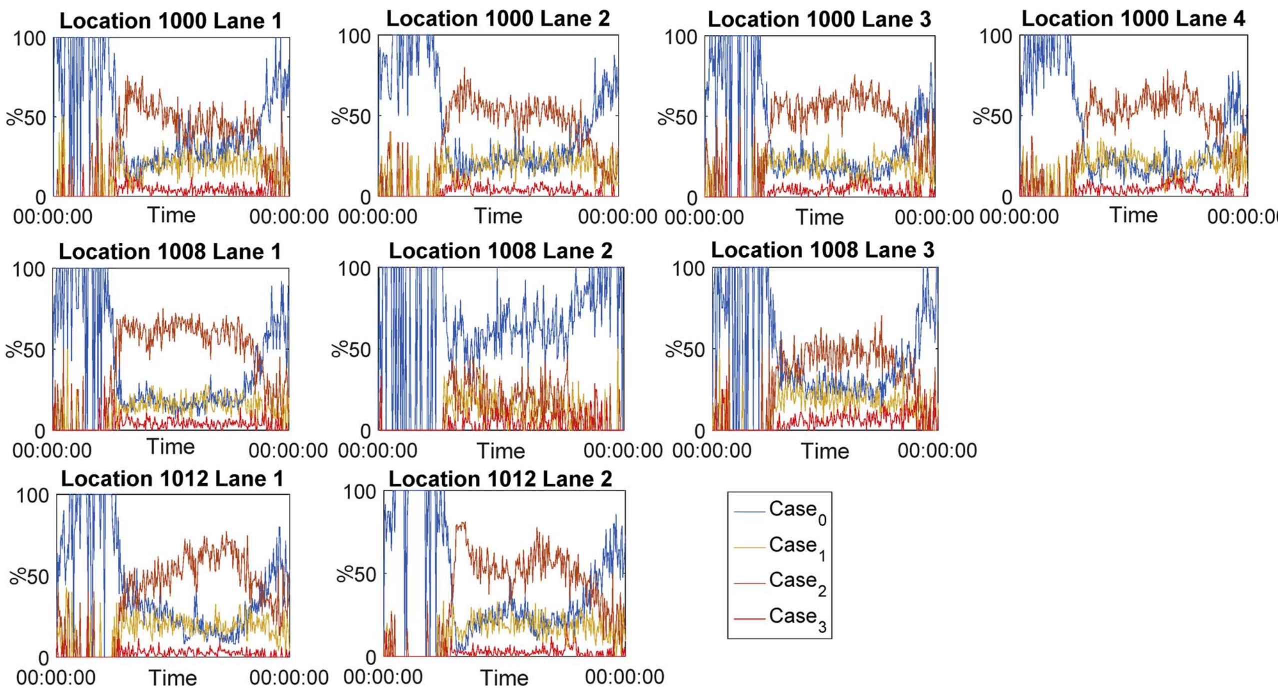

Fig. 4. Rear-end crash potential analysis per lane at three distinctive locations.

1 s, 2 s, 3 s) and three values for Vehicle Deceleration Rate (DR: 0.1 g, 0.3g, 0.5 g), as presented in Fig. 5.

As it can be observed, the higher the deceleration of the vehicle the lower were the potential of engaging in a rear-end crash, and especially in Case_1 here, where the leading vehicle is considered to be stationary and as such the crash probability is solely based on the following vehicle. Similarly, the SA showed that the lower the reaction time the higher is the potential of avoiding a rear-end crash. In the present study, typical reaction times of 3 s and deceleration rates of 0.5 g are assumed.

4.2.2. Traffic parameter effect investigation

Based on the previous analysis rear-end crash potential was exposing dynamic and fluctuating characteristics. Therefore, in the following sections the results of a correlation investigation on the effect of various traffic parameters on car-following decisions and the rear-end potential are presented.

4.2.2.1. Traffic characteristics. The traffic characteristics assessed in this study correspond to traffic flow, average vehicle speed and speed standard deviation. Fig. 6 presents the daily profile of flow, mean speed and speed standard deviation, computed over the 5-min aggregation interval. The statistical results showed that the mean speed of the total sample was 49 km/h with a standard deviation equal to 16.0 km/h. Fig. 6 indicates the typical flow pattern for urban networks where flows are lower during nighttime and higher during day time. Also, mean speeds are higher and speed values present a higher variation during nighttime in comparison to daytime.

In accordance to the rear-end crash potential analysis presented above, it is clear that rear-end crash potential cases 1–3 (i.e. cases where rear-end crash has increased probability) are presented when traffic flow is higher, as seen when comparing Figs. 4 and 6. This is an expected result as in higher flows (and subsequently increased density values), vehicles have more interactions and therefore rear-end crash propensity increases. In previous real-time safety studies, vehicle count was found to be the most common exposure for crashes (e.g. Yu and Abdel-Aty, 2013).

Based on traffic flow fundamentals, speed is (linearly) correlated with density with higher density values resulting in reduced speed values. In general, traffic volume affects traffic speeds, the variation of speed of the travelling vehicles and the car following decisions of the

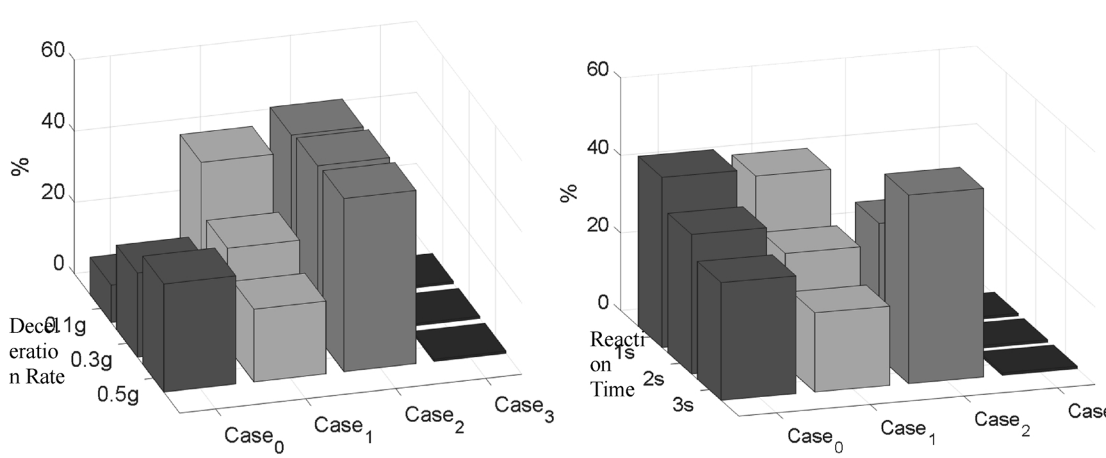

Fig. 5. Sensitivity Analysis Results (a) Deceleration Rate; (b) Reaction time.

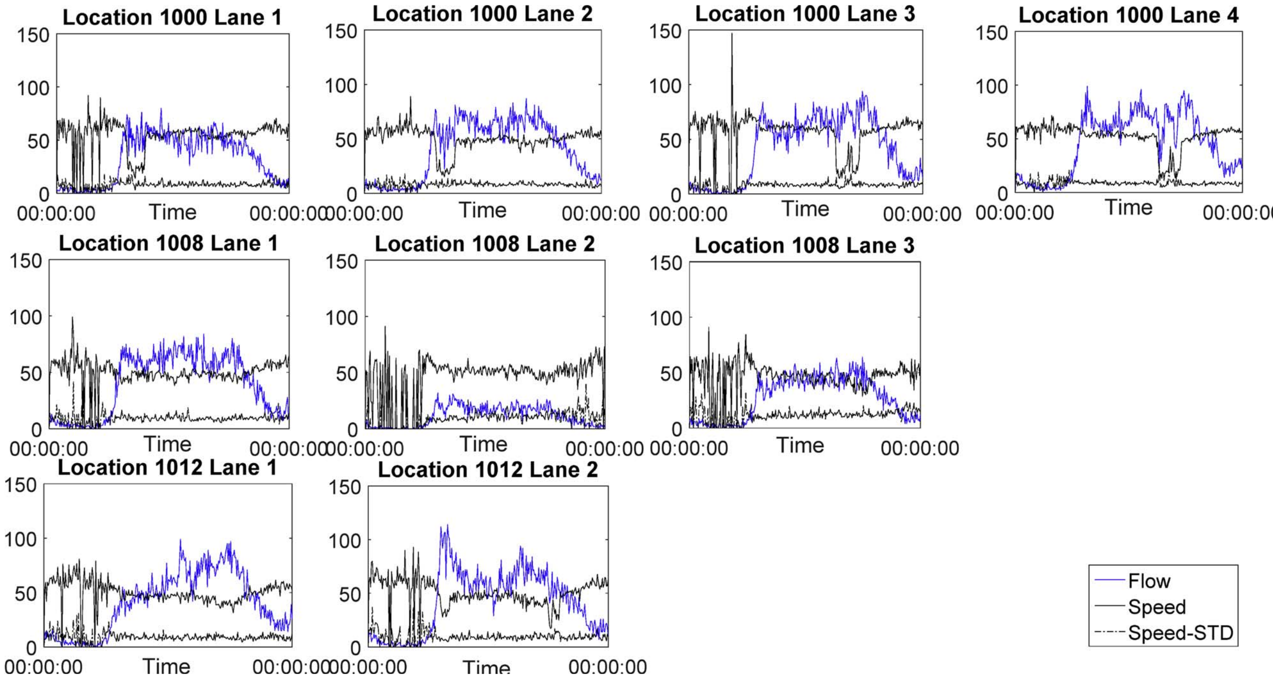

Fig. 6. Speed, Flow and Speed Standard Deviation Diurnal Profiles per lane for three distinctive locations.

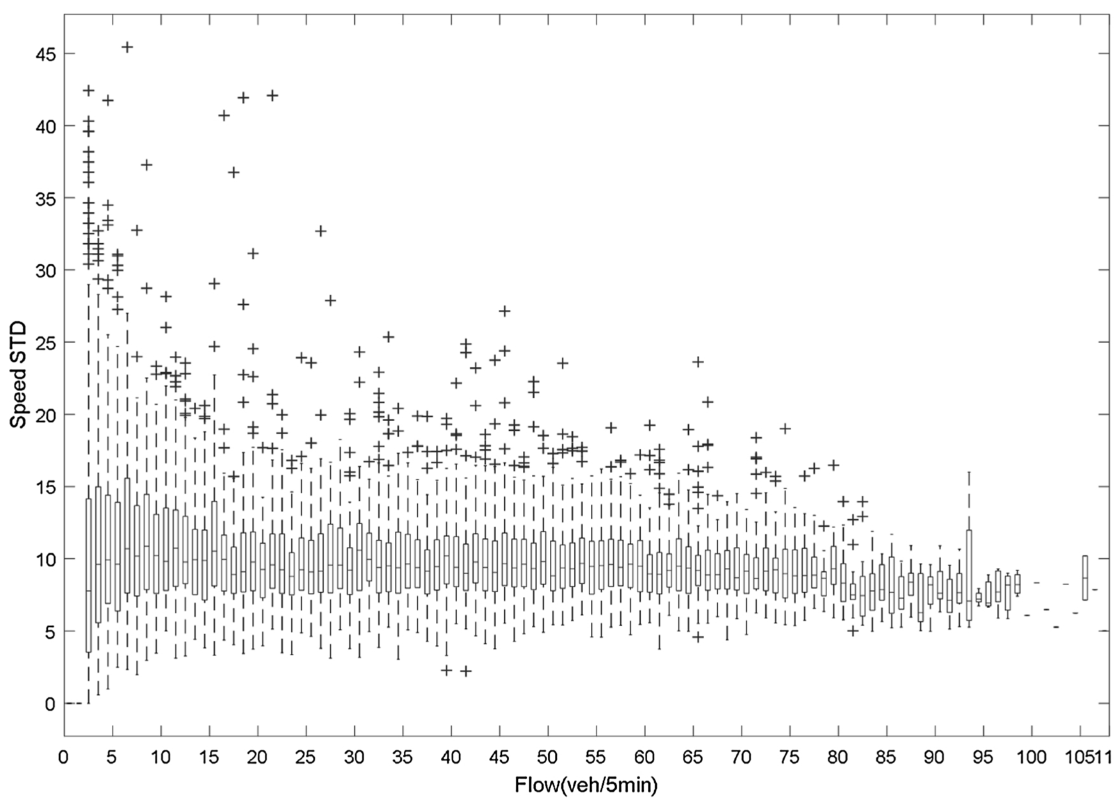

Fig. 7. Flow-Speed STD relationship for the complete dataset.

drivers. In addition to flow, speed and speed standard deviation daily profiles (presented above), the relationship between flow and speed standard deviation is investigated as it was considered as a good indicator of vehicle interaction and turbulence in regard to the traffic state. Fig. 7 presents the relationship between flow and speed standard deviation aggregated in the 5-min interval for the complete sample. Additionally, in Fig. 8 the relationship between flow and speed standard deviation per lane of three distinctive locations is presented, exposing the same evidence as of the complete sample. The flow-speed standard deviation diagram indicates that speed variation is higher at lower values of traffic flow, but reduces as traffic flow increases.

Moreover, Fig. 9 indicates that the rear-end crash potentials are distributed in relation to the variation of speed. It is shown that rear-end crash potentials form a major cluster at higher speed standard deviation values, and during the hours where traffic flow is increased. This result provides evidence that in the urban network, rear-end crash

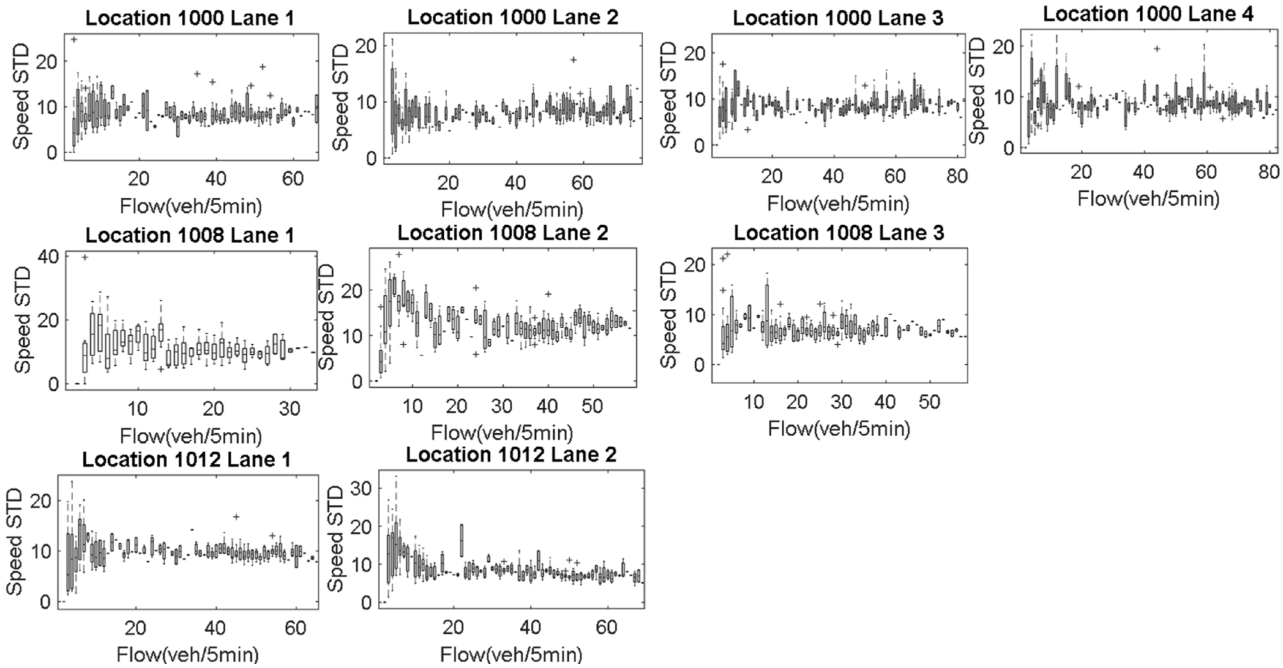

Fig. 8. Flow-Speed STD relationship per lane for three distinctive locations.

potential tends to be evident primarily under conditions of varying speed. Similar results have been reported by Wang et al. (2017), where the average daily standard deviation of speed (aggregated in 1-minute intervals) had a positive effect on crash frequency.

4.2.2.2. Locational conditions. A reasonable assumption that should be investigated with respect to driving style corresponds to the particularities of each location (e.g. geometric characteristics, number of lanes etc.). In order to observe any possible behavioral changes, the nonparametric test of equality, Kolmogorov-Smirnov two sample test (K–S test) was employed between the two lanes of each dual carriageway in the examined network. The K-S test was performed for both the car-following decisions considered in this study, namely speed and temporal headway. In all the tests performed the null hypothesis that the two samples come from the same continuous distribution was rejected. Although this does not translate to a directly observable behavioral effect, it does provide evidence that driving differs between lanes of travel. The effect of lane of travel is better observed in the following modeling analysis.

The effect of location in the urban network was also studied. Fig. 10 presents the occurrence of the rear-end crash potential for the different locations studied here. Each subplot corresponds to one of the rear-end crash potential cases as described in Section 3.1, and it is a representation of the occurrence of each case through time and in the different locations. The resulting image corresponds in a 288-by-15 grid, where 288 is the number of rows (288 five-minute intervals in a day) and 15 is the number of columns (15 examined locations). Therefore, in order to draw conclusions regarding the locational effect on rear-end crash potential from this analysis the columns in the same image should be compared. The percentage of each case is presented in different colors, with the blue color indicating 0% and the red color indicating 100%.

It is shown that the occurrence of each of the rear-end crash potential cases are different through the alternative locations. For relating specific locational characteristics with rear-end crash classification, Fig. 10 should be checked along with Table 1. This spatial analysis shows that rear-end crash potentials are expected increased near signalized intersections.

4.2.2.3. Leading vehicle effect. Lastly, the effect of leading vehicle size was investigated. Loop detector technology allowed for the identification of vehicle type in real time which provided valuable information for identifying possible differentiation in car-following decisions under mixed traffic conditions. The effects of leading vehicle type on speed and headway of the following vehicle in this study were found to differ according to the vehicle's size. It was shown

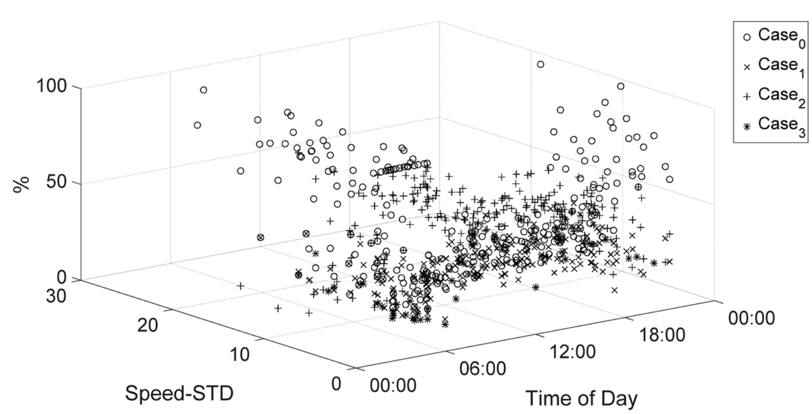

Fig. 9. Rear-end crash potential relationship with Speed Standard Deviation – Location 1000 Lane 1.

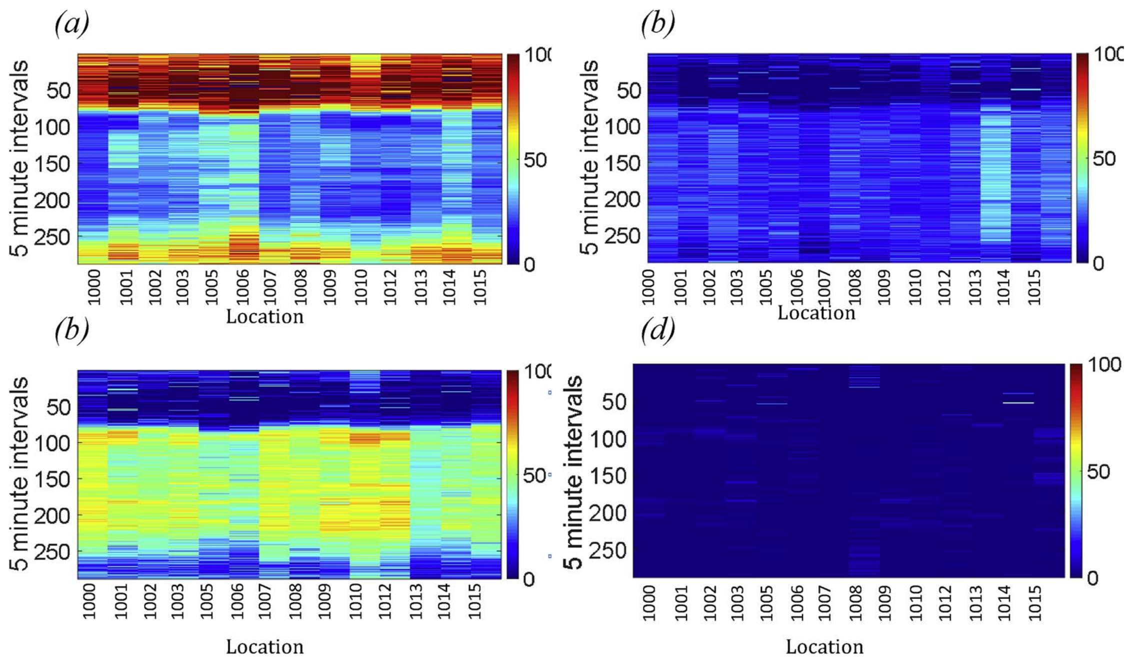

Fig. 10. Rear-end crash potential for different locations (a) Case_0; (b) Case_1; (c) Case_2; (d) Case_3.

that speeds are lower and headways higher when Heavy Goods Vehicles (HGV) lead. Four configurations of car-following settings were examined based on following-leading vehicle setup:

(i) Private Vehicle – Private Vehicle (PV);
(ii) PV– HGV;
(iii) HGV – PV;
(iv) HGV-HGV.

Fig. 11 presents mean speeds and headways for the aforementioned configurations. Fig. 11*a* shows that mean speed was lower when the leading vehicle was a HGV compared to when the leading vehicle was a PV. Mean headway was found to be higher in the configuration of PV-HGV, compared to PV-PV. However, in the cases where HGVs were following, the mean time headway was found lower when a HGV was leading compared to when a PV was leading. Two distinctive peaks are presented in Fig. 11*b* in the configuration of HGV-PV, and one at HGV-HGV. These represent situations where the aforementioned car-following configurations were present within a large time gap. As per the results of this analysis, the effect of vehicles configuration affects their speed and time headways, which in turn are important in crash risk classification.

4.2.3. Multinomial Logit model

The results of the exploratory correlation analysis provided earlier regarding rear-end crash potential, showed that car-following decisions are affected by traffic measures and states, locational characteristics, and vehicle size/type. Moreover, it has been pointed out that there are multiple stochastic and dynamic inter-dependencies among the variables reflecting those characteristics, possibly reflecting the decision-making process of the multiple drivers-classes in realistic circumstances. In order to further analyze datasets with the above characteristics, a model-based investigation has been undertaken. In particular, a Multinomial Logit Model (MNL) is used, which can be seen as a suitable model, able to handle stochastic multivariate datasets and identify statistically significant variables, treating characteristics of heterogeneity, variability and randomness. Moreover, in order to capture the dynamics of the rear-end crash potential, two model settings have been investigated: an overall diurnal static setting, followed by an intra-day dynamic extension.

In particular, in the model development phase, the constant specific variables were set as the rear-end crash potential cases, the alternative specific variables were the car following decisions (speed and temporal headway), and finally the individual specific variables were set as the lane of travel, loop detector location, leading vehicle type, flow (aggregated in a 5-minute interval), mean speed (aggregated in a 5-minute interval), and speed standard deviation (aggregated in a 5-minute interval). Several models (thirteen in total) were evaluated in order to investigate the significance of the variables on rear-end crash potentials. In the procedure of developing the alternative multinomial logit

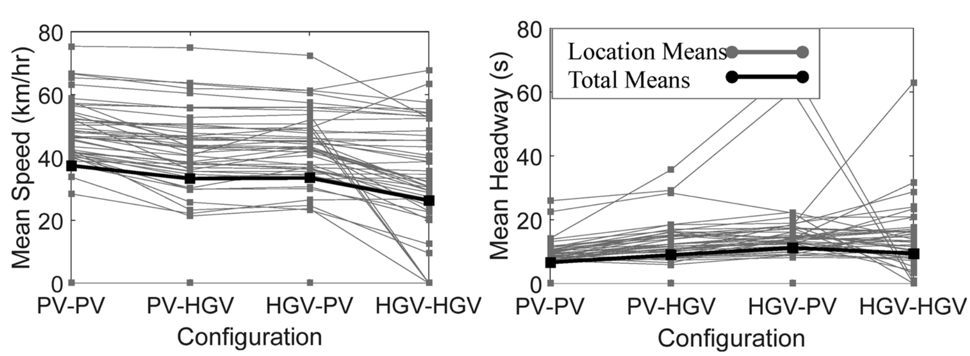

Fig. 11. Leading Vehicle Effect (a) Mean Speed (b) Mean Headway. (For interpretation of the references to colour in text, the reader is referred to the web version of this article).

Table 3
Multinomial Logit Model Results (t-values in parenthesis).

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9	Model10	Model11	Model12	Model13
C1:Intercept	−0.5631 (-43.7842) ***	−15.579 (-14.9388) ***	−15.998 (-15.0657) ***	−0.6247 (-23.1151) ***	−0.3804 (-16.1857) ***	−6.1112 (-5.5071) ***	−1.7445 (-55.5648) ***	−0.7871 (-26.6196) ***	−1.7031 (-42.6385) ***	−1.6803 (-42.3003) ***	−0.7223 (-19.8249) ***	−17.060 (-15.8192) ***	−6.6846 (-5.9838) ***
C2:Intercept	1.4082 (129.1338) ***	7.8327 (8.7760) ***	7.8311 (8.7542) ***	2.2257 (84.7342) ***	1.8505 (93.8497) ***	6.0045 (6.4158) ***	−4.101 (15.7622) ***	0.7288 (28.5860) ***	−0.7222 (21.7709) ***	0.6857 (20.8204) ***	1.5000 (44.1073) ***	4.1090 (4.4878) ***	4.4227 (4.6809) ***
C3:Intercept	−1.3716 (-73.238) ***	−3.1002 (-1.9405)	−3.3784 (-2.1132) *	−1.0567 (-25.6308) ***	−2.2194 (-70.4202) ***	−12.2170 (-7.3284) ***	−1.4776 (-33.3503) ***	−3.5968 (-82.0430) ***	−2.8688 (-48.9158) ***	−2.8469 (-49.1272) ***	−3.2120 (-58.7778) ***	0.1955 (0.1208)	−5.8414 (-3.4738) ***
Speed	−0.0155 (-251.1762) ***	−0.0154 (-250.8632) ***	−0.0155 (-250.5381) ***	−0.0155 (-250.8414) ***	−0.0155 (-250.9261) ***	−0.0146 (-235.0377) ***	−0.0146 (-235.0070) ***	−0.0149 (-239.4775) ***	−0.0146 (-235.1978) ***	−0.0146 (-235.1033) ***	−0.0149 (-239.2033) ***	−0.0149 (-238.8472) ***	−0.0146 (-234.8135) ***
Headway	−0.000027 (-3.0485) **	−0.000027 (-3.0489) **	−0.000027 (-3.0404) **	−0.000027 (-3.0401) **	−0.000026 (-2.9887) **	−0.000029 (-2.9628) **	−0.000029 (-2.9682) **	−0.000031 (-3.1740) **	−0.000029 (-2.9744) **	−0.000029 (-2.9754) **	−0.000031 (-3.1713) **	−0.000031 (-3.1650) **	−0.000029 (-2.9727) **
C1:Lane	0.0749 (14.6275) ***	0.0733 (14.4528) ***	0.0736 (14.5183) ***	0.07457 (14.6970) ***	0.00749 (14.7524) ***	0.0360 (6.9444) ***	0.0352 (6.7765) ***	0.0653 (12.7510) ***	0.0353 (6.7927) ***				0.0372 (7.1559) ***
C2:Lane	−0.0134 (-3.1307) **	−0.0133 (-3.1108) **	−0.0089 (-2.0870) *	−0.0089 (-2.6814) *	−0.0145 (-2.4428) *	−0.0421 (-9.5946) ***	−0.0415 (-9.4600) ***	−0.0377 (-8.6562) ***	−0.0382 (-8.6713) ***				−0.0340 (-7.7029) ***
C3:Lane	0.03179 (4.2232) ***	0.0315 (4.1866) ***	0.0338 (4.4902) ***	0.03419 (4.5378) ***	0.0283 (3.7025) ***	−0.0280 (3.6628) ***	0.02737 (3.5757) ***	−0.0040 (-0.5307)	0.0138 (1.7877) .				0.0174 (2.2621) *
C1:Location		0.0152 (14.4089) ***	0.0153 (14.4809) ***			0.0044 (3.9429) ***						0.0162 (15.1572) ***	0.0049 (4.4590) ***
C2:Location		−0.0064 (-7.1990) ***	−0.0056 (-6.2696) ***			−0.0055 (-5.9753) ***						−0.0026 (-2.8528) ***	−0.0029 (-3.0405) ***
C3:Location		0.00172 (1.0817)	0.0023 (1.4522)			0.0106 (6.4367) ***						−0.0033 (-2.1065)	0.0034 (2.0255) *
C1: Vehicle			0.0261 (2.2348) *	0.0313 (2.6814) **							0.0222 (1.9083))	0.0171 (1.4682)	0.0111 (0.9436)
C2:Vehicle			−0.4024 (-34.0238) ***	−0.4046 (-34.1864) ***							−0.4191 (-35.1017) ***	−0.4183 (-35.0516) ***	−0.4146 (-34.8335) ***
C3:Vehicle			−0.1555 (-8.5713) ***	−0.1546 (-8.5286) ***							−0.2011 (-10.9607) ***	−0.1996 (-10.8798) ***	−0.2122 (-11.6080) ***
C1:Flow						−0.0099 (30.4894) ***		0.0076 (23.8087) ***	0.0099 (29.8682) ***	0.0103 (31.1466) ***	0.0083 (25.5441) ***	0.0083 (25.7581) ***	0.0098 (29.6482) ***

(continued on next page)

Table 3 (*continued*)

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9	Model10	Model11	Model12	Model13
C2:Flow						0.0199 (71.1994) ***		0.0191 (68.2873) ***	0.0192 (6,707,552) ***	0.0189 (67.2475) ***	0.0191 (68.1865) ***	0.01901 (68.0782) ***	0.0193 (67.9360) ***
C3:Flow						−0.0181 (38.1354) ***		0.0239 (49.6741) ***	0.0221 (44.8363) ***	0.0222 (45.2111) ***	0.0241 (50.1040) ***	0.02411 (50.1073) ***	0.0222 (44.8961) ***
C1:Av_Speed						0.0152 (31.4151) ***	0.0159 (34.3079) ***		0.0158 (34.0259) ***	0.0165 (35.7965) ***			0.0147 (30.2594) ***
C2:Av_Speed						0.00082 (2.0595) *	0.0002 (0.6113)		−0.0002 (-0.7077)	−0.00079 (-2.0752) *			−0.00009 (-2.2064) *
C3:Av_Speed						−0.0169 (-25.1829) ***	−0.0162 (-24.3659) ***		−0.0123 (-18.1720) ***	−0.012 (-18.1423) ***			−0.0131 (-18.7819) ***
C1:STD_Speed					−0.0189 (-9.1930) ***			−0.01091 (-5.2342) ***	−0.0033 (-1.5738)	−0.0023 (-1.1232)	−0.0093 (-4.4940) ***	−0.0127 (-6.0723) ***	−0.0047 (-2.2064) *
C2:STD_Speed					−0.0467 (-27.1850) ***			−0.0269 (-15.2964) ***	−0.0266 (-15.1205) ***	−0.0275 (-15.6923) ***	−0.0263 (-14.9607) ***	−0.0255 (-14.4428) ***	−0.0245 (-13.7985) ***
C3:STD_Speed					0.0845 (34.4738) ***			0.1111 (43.0696) ***	0.1016 (39.0500) ***	0.1018 (39.1344) ***	0.1126 (43.6038) ***	0.1131 (43.6298) ***	0.1023 (38.9062) ***
Log-Likelihood	−362220	−361910	−361020	−361300	−360930	−357470	−357570	−357630	−356430	−356590	−356980	−356740	−355470
McFadden R^2	0.173	0.174	0.176	0.175	0.176	0.184	0.184	0.184	0.186	0.186	0.185	0.186	0.189
X^2	151780	152390	154070	153500	154350	161260	161070	160940	163340	136020	162140	162630	165160

- Statistically Significant Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1.

models all the variables were considered for inclusion; the development of the models followed a sequential and additive rationale where firstly the effect of locational characteristics was investigated (lane and location), then, vehicle characteristics were added (leading vehicle type), and lastly the traffic measures were added (flow, speed, average speed, speed standard deviation). The last model consisted of all the independent variables for comparative purposes.

The base case was selected as Case_0 therefore its constant specific can be regarded as 0 in the models presented. In these MNLs the coefficients of the variables signify the difference in the utility of the corresponding rear-end crash potential case, by altering its attribute while holding the others constant, thus indicating the specific variable's influence on the rear-end crash potential case. The estimation results for the alternative MNL models tested here are summarized in Table 3.

Analysis of the results indicates that in the alternative models, some factors were not found to be statistically significant determinants of rear-end crash potential. The coefficient estimates for the variables of Models 1,4,5,6 and 7 were all found to be statistically significant at the level of 5%. Also, McFaddens R^2 for the aforementioned MNLs ranges form 0.17–0.19 indicating an acceptable goodness-of-fit of the proposed models to the data. Alternative selection of variables sets has been tested in models 8-12, exposing the contribution of each variable in explaining the risk perception as this is captured in driving classification. It is noted that Model 13 was estimated including all the explanatory variables. From this complete model, the only variable that was not found to be significant was vehicle type when specific to rear-end crash potential Case_1.

The MNL results presented in Table 3 consist of the complete dataset covering a period of 24 h. In addition to these modelling approaches and since the preliminary analysis indicated that rear-end crash potentials vary throughout the day (Fig. 3), Model5, was chosen to be recursively (hourly) estimated over the day, providing supplementary findings, particularly on the model stability. The chosen model consists of lane of travel and speed standard deviation as individual specific variables. The selection of Model5 was based on its straightforward structure, and further on the previous analysis which indicated that lane and speed standard deviation have an effect on the rear-end crash potential. The results from the KS-Test presented in Section 4.2.2 indicated that drivers' car following decisions differ between the different lanes of travel in the network; the modeling results presented in Table 3corroborate the fact that lane of travel is a significant variable in rear-end crash potential. Additionally, the negative coefficients of Speed_STD specific to Cases 1 and 2 reflect the fact that the higher the variation of speed the lower the probabilities of these cases to be present, whereas the positive sign for Speed_STD when specific to Case 3 shows that the presence of this rear end crash potential is increased when speed variation increases. The same results are evident when the model is estimated for each hour of the day, as presented in Fig. 12.

Fig. 12 presents the hourly coefficient values and their variation through the day, for the selected Model 5. A backward elimination process was followed in order to identify significant variables for the hourly model, and as such, only the statistically significant, at the level of 0.1, coefficients are presented here. The variable coefficients were scaled to the $(-1, 1)$ range and conversion values are presented in the legend's parentheses. The coefficients for the variables as calculated for the static diurnal model are also presented in Fig. 12, with dashed lines. As can be seen from Fig. 12, for the hourly model, the coefficients of the intercepts and speed were found to be statistically significant for each hour of the day. On the contrary the coefficients of Headway, Lane and Speed-STD were not found to be significant for every hour. Interestingly enough, the coefficients of Speed-STD were systematically found to be significant between 10:00 and 14:00, a period that can be characterized as intermediate in terms of traffic state conditions.

The coefficient estimate signs were found to be the same across all the variables for each hour implying a systematic relationship between the examined variables and rear-end crash potential. The coefficient of lane was only found statistically significant for one hour of the day (17:00) and presented a positive sign when specific for all rear-end crash potentials. The coefficient of Speed-STD presented a negative sign when specific to Cases 1 and 2 and a positive sign for Case 3 in the hourly analysis. The coefficient of Speed-STD in the utility was systematically found higher for the Case_3 alternative. This result shows that when the three rear-end crash potential cases are compared, increasing Speed-STD is associated with an increased presence of Case_3.

5. Conclusions and outlook

Rear-end crashes have been identified as a frequent type of traffic crash and have long been studied in the road safety science, thus the investigation and understanding of the significant contributing factors to this type of crashes is of practical importance. It is very important to understand the effects of driving behavior, vehicle type interactions as well as road characteristics, on rear-end crashes. Using an extensive database of disaggregate vehicle-by-vehicle information, obtained from inductive loop detectors, this study investigated the rear-end crash potential in an urban network. In this study, rear-end crash potential

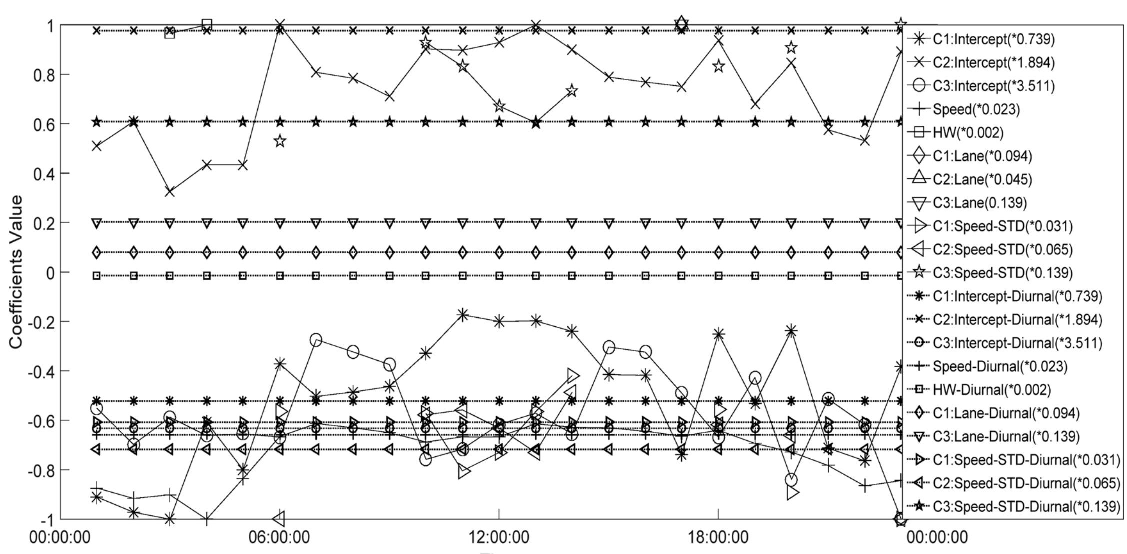

Fig. 12. Model 5 coefficient values time series.

was estimated by the development of a technique based on stopping distance. Second, an exploratory analysis was conducted to investigate the effect of traffic measures, locational conditions, and vehicle types on the estimated rear-end crash potential and driver car-following decisions and lastly a Multinomial Logit Model analysis was developed to identify the significant variables. The emphasis of this study is the investigation of the rear-end crash potential in the more complex and dynamic urban environment, rather than the freeways.

Rear-end crash potentials here were estimated based on stopping distance by using individual car following decisions, namely speed and temporal headway, obtained from inductive loop detectors. Rear-end crash potentials were allocated into four cases based on the type of rear-end crash. Here, Case_0 indicated the case where there was no rear-end crash potential; Case_1 indicated the case where the distance kept from the leading vehicle is not sufficient when the leading vehicle is stationary; Case_2 indicated the case where there is not sufficient distance for the following vehicle to stop when the leading vehicle performs an emergency brake; and finally Case_3 indicated the case where the distance kept is not sufficient to avoid collision with the leading vehicle in the next 10 s if no action is taken. The analysis showed that overall 66.4% of the observations were estimated as rear-end crash potentials.

In the current study, by combining car-following concepts with extensive realistic datasets from multiple location within the urban environment, it was able to quantitatively provide evidence on the rear-end crash potential and in overall rear-end crashes risk assessment. In particular, it was found that rear-end crash potentials were more evident during daytime-working hours, where flows (and vehicle interactions) are higher, rather than nighttime (although the speeds are higher). It was also shown that other than flow, the standard deviation of speed was found to affect rear-end crash potential. The analysis also showed that rear-end crash potential differs in relation to the location in the network and specifically, it was reported that more rear-end potentials were expected closest to signalized intersections. The results also highlighted the difference of drivers' behavioral characteristics in relation to the traffic configuration as it was shown that speeds were lower and headways higher when Heavy Goods Vehicles (HGV) led.

The modeling results highlighted that the proposed explanatory variables were significantly associated with rear-end crash potential, however their effect varied through the different models that were estimated. In total, thirteen Multinomial Logit Models (MNL) were estimated, with different combinations of the explanatory variables. In the procedure of estimating the MNLs all the explanatory variables were tested for inclusion, but not all were found to be significant. Model 13, which was the structure utilizing all the explanatory variables, provided the highest McFadden R^2 (0.189) but not all the variables were found to be significant; vehicle type when specific to rear-end crash potential Case_1 was not found to be significant, which is expected as Case_1 rear-end potential is invariant of the leading vehicle. However, all the other variables were found to be significant at the 0.05 level. The modeling results provided evidence that more simplified models can still provide a good performance. One model structure (Model 5) was separately estimated for each hour of the day. The results of this analysis showed that the variables' effect was consistent throughout the day (coefficient sign remained the same), however their significance was not. This finding shows that rear-end potential is sensitive to the real-time traffic characteristics and as such this fact should be taken account of when estimating hazardous conditions.

This study has some limitations which should be mentioned as to avoid misinterpretation of the presented results. The research is based on disaggregate vehicle-by-vehicle information obtained from loop detectors, and therefore, instantaneous observations of speed and temporal headway. Such data, lacking continuity through time, could therefore result in the over-reporting of rear-end crash potential, as traffic conditions vary especially in the urban network. Rear-end crash potential here is estimated based on the data collected, but not validated through an actual crash database, therefore more complete datasets are desirable to support the rear-end crash potential estimation proposed here.

The findings of this study provide evidence that monitoring the distance between consecutive vehicles in real-time situations could be valuable in the identification of near-crash phenomena and crash prediction. The results depict that traffic management authorities can benefit from the wealth of collected data to develop tools aiming in the mitigation of crash risk and it is believed that using the evidence outlined in this study, researchers and policymakers can identify real-time 'hazardous' traffic conditions and mitigate crash potentials.

Further research is desirable to enhance the proposed rear-end crash potential estimation and better understand the association of the significant variables. Additionally, more modelling structures should be investigated and estimated to provide a better understanding of the rear-end crash potential contributing factors and more information in seeking risk mitigation countermeasures. It is explicitly noted as a concluding remark that the proposed analytical procedure (and as so the estimated model too) focus on the explanatory investigation of the factors that effect on crash risk potential, not on predicting crash risk. As so, in this paper the extensive database of almost 400,000 vehicle observations from various locations in the urban network under study was utilized in order to capture the significant factors that effect on potential rear-end collisions and to model the 'way' that these factors correlate with crash potential. A reasonable extension of this study could be the prediction of crash potential in similar circumstances. The ultimate task of this analysis could be the development and application of traffic control strategies in order to minimize crash potential both for rear-end collisions, as well as for other types of crashes.

References

American Association of State Highway and Transportation Officials-AASHTO, 2011. A Policy 29 on Geometric Design of Highways and Streets, fifth edition. Washington, DC.

Abdel-Aty, M., Abdelwahab, H., 2004. Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accid. Anal. Prev. 36 (3), 447–456.

Abdel-Aty, M.A., Hassan, H.M., Ahmed, M., Al-Ghamdi, A.S., 2012. Real-time prediction of visibility related crashes. Transp. Res. Part C Emerg. Technol. 24, 288–298. http://dx.doi.org/10.1016/j.trc.2012.04.001.

Ahmed, M.M., Abdel-Aty, M., Lee, J., Yu, R., 2014. Real-time assessment of fog-related crashes using airport weather data: A feasibility analysis. Accid. Anal. Prev. 72, 309–317. http://dx.doi.org/10.1016/j.aap.2014.07.004.

Ahmed, M., Abdel-Aty, M., 2013. A data fusion framework for real-time risk assessment on freeways. Transp. Res. Part C: Emerg. Technol. 26, 203–213.

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis. MIT Press, Cambridge, Mass.

Christoforou, Z., Cohen, S., Karlaftis, M., 2011. Identifying crash type propensity using real-time traffic data on freeways. J. Saf. Res. 42 (1), 43–50.

Duan, J., Li, Z., Salvendy, G., 2013. Risk illusions in car following: is a smaller headway always perceived as more dangerous? Saf. Sci. 53, 25–33.

Fang, S., Xie, W., Wang, J., Ragland, D.R., 2016. Utilizing the eigenvectors of freeway loop data spatiotemporal schematic for real time crash prediction. Accid. Anal. Prev. 94, 59–64. http://dx.doi.org/10.1016/j.aap.2016.05.013.

Ghanipoor Machiani, S., Abbas, M., 2016. Safety surrogate histograms (SSH): A novel real-time safety assessment of dilemma zone related conflicts at signalized intersections. Accid. Anal. Prev. 96, 361–370. http://dx.doi.org/10.1016/j.aap.2015.04.024.

Golob, T., Recker, W., 2004. A method for relating type of crash to traffic flow characteristics on urban freeways. Transp. Res. Part A: Policy Pract. 38 (1), 53–80.

Hourdos, J., Garg, V., Michalopoulos, P., Davis, G., 2006. Real-time detection of crash-prone conditions at freeway high-crash locations. Transp. Res. Rec.J. Transp. Res. Board 1968, 83–91.

Hu, W., Xiao, X., Xie, D., Tan, T., Maybank, S., 2004. Traffic accident prediction using 3-D model-based vehicle tracking. IEEE Trans. Veh. Technol. 53 (3), 677–694.

Kim, J., Wang, Y., Ulfarsson, G., 2007. Modeling the probability of freeway rear-end crash occurrence. J. Transp. Eng. 133 (1), 11–19.

Kim, S., Song, T., Rouphail, N., Aghdashi, S., Amaro, A., Goncalves, G., 2016. Exploring the association of rear-end crash propensity and micro-scale driver behavior. Saf. Sci. 89, 45–54.

Kwak, H.C., Kho, S., 2016. Predicting crash risk and identifying crash precursors on korean expressways using loop detector data. Accid. Anal. Prev. 88, 9–19. http://dx.doi.org/10.1016/j.aap.2015.12.004.

Li, Z., Ahn, S., Chung, K., Ragland, D., Wang, W., Yu, J., 2014. Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks. Accid. Anal. Prev. 64, 52–61.

Meng, Q., Qu, X., 2012. Estimation of rear-end vehicle crash frequencies in urban road tunnels. Accid. Anal. Prev. 48, 254–263.

National Safety Council, 2015. Injury Facts®, 2015 Edition. Itasca, IL: Author. Library of Congress Catalog Card Number: 99-74142.
Oh, C., Oh, J., Ritchie, S., 2001. Real-time estimation of freeway accident likelihood. In: Presented at the 80th Annual Meeting of the Transportation Research Board. Washington DC.
Oh, C., Park, S., Ritchie, S., 2006. A method for identifying rear-end collision risks using inductive loop detectors. Accid. Anal. Prev. 38 (2), 295–301.
Oh, C., Kim, T., 2010. Estimation of rear-end crash potential using vehicle trajectory data. Accid. Anal. Prev. 42 (6), 1888–1893.
Shi, Q., Abdel-Aty, M., 2015. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transp. Res. Part C: Emerg. Technol. 58, 380–394.
Singh, S., 2003. Driver attributes and rear-end crash involvement propensity. NHTSA Report DOT-HS-809-540. U.S. Department of Transportation, Washington, D.C.
Stylianou, K., Dimitriou, L., Abdel-Ady, M., 2017. Investigating rear-end collision potential at signalized urban networks based on disaggregated spatial sensor data. In: Presented at the 96th Annual Meeting of the Transportation Research Board. Washington DC.
Taieb-Maimon, M., Shinar, D., 2001. Minimum and comfortable driving headways: reality versus perception. Hum. Factors 43 (1), 159–172.
Theofilatos, A., 2017. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. J. Saf. Res. 61, 9–21. http://dx.doi.org/10.1016/j.jsr.2017.02.003.
Wang, L., Abdel-Aty, M., Lee, J., Shi, Q., 2017a. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. Accid. Anal. Prev. 1–7. http://dx.doi.org/10.1016/j.aap.2017.06.003.
Wang, L., Abdel-Aty, M., Shi, Q., Park, J., 2015. Real-time crash prediction for expressway weaving segments. Transp. Res. Part C Emerg. Technol. 61, 1–10. http://dx.doi.org/10.1016/j.trc.2015.10.008.
Wang, X., Zhu, M., Chen, M., Tremont, P., 2016. Drivers??? rear end collision avoidance behaviors under different levels of situational urgency. Transp. Res. Part C Emerg. Technol. 71, 419–433. http://dx.doi.org/10.1016/j.trc.2016.08.014.
Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accid. Anal. Prev. 38 (6), 1137–1150.
Wang, L., Abdel-Aty, M., Lee, J., 2017b. Safety analytics for integrating crash frequency and real-time risk modeling for expressways. Accid. Anal. Prev. 104, 58–64.
Weng, J., Meng, Q., 2012. Rear-end crash potential estimation in the work zone merging areas. J. Adv. Transp. 48 (3), 238–249.
Wu, Y., Abdel-Aty, M., Lee, J., 2017. Crash risk analysis during fog conditions using real-time traffic data. Accid. Anal. Prev. 1–8. http://dx.doi.org/10.1016/j.aap.2017.05.004.
Xu, C., Wang, W., Liu, P., Li, Z., 2015. Calibration of crash risk models on freeways with limited real-time traffic data using Bayesian meta-analysis and Bayesian inference approach. Accid. Anal. Prev. 85, 207–218. http://dx.doi.org/10.1016/j.aap.2015.09.016.
Xu, C., Tarko, A., Wang, W., Liu, P., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accid. Anal. Prev. 57, 30–39.
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 (6), 983–995.
Yu, R., Abdel-Aty, M., 2013. Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes. Accid. Anal. Prev. 58, 97–105.