Impact of Road and Traffic Characteristics on Driver Behaviour and Safety Using Data from Smartphones

Eleni-Konstantina Frantzola
Civil & Transportation Engineer, National Technical University of Athens,
5 Iroon Polytechniou Str, Zografou Campus, GR-15773, Athens, Greece Tel: +30 210 772 1575;
Fax: +30 210 772 1454
Email: elinafra@gmail.com

Armira Kontaxi*
Ph.D. Candidate
School of Civil Engineering, National Technical University of Athens,
5 Iroon Polytechniou Str, Zografou Campus, GR-15773, Athens, Greece Tel: +30 210 772 1575;
Fax: +30 210 772 1454
Email: akontaxi@mail.ntua.gr

George Yannis, Ph.D.
Professor,
School of Civil Engineering, National Technical University of Athens,
5 Iroon Polytechniou Str, Zografou Campus, GR-15773, Athens, Greece Tel: +30 210 772 1326;
Fax: +30 210 772 1454
Email: geyannis@central.ntua.gr

Word Count: 5489 words + 3 tables (250 words per table) = 6239 words Submitted [February 28, 2020]

*Corresponding Author

ABSTRACT

The Internet of Things (IoT) constantly offers new opportunities and features to monitor and analyse driver behaviour in direct assessment and improvement of driver behaviour and safety.
Aim: The objective of the current research is to exploit large-scale spatio-temporal data from smartphone sensors to investigate harsh events throughout the selected road axes, specifically at the road segment and junction levels.
Methods: Drivers were recorded for one year using an innovative smartphone application and data were imprinted spatially using Geographical Information Systems in order to indicate patterns of the accumulation and ranking of the harsh events regarding time variation. Their measurements are complemented by additional traffic parameter measurements from the Traffic Management Centre of Athens in order to acquire a measure of the surrounding traffic compared to each driver. Subsequently, log-linear regression models were developed in order to associate harsh events numbers with road geometry parameters.
Results and conclusions: Results reveal an increase in harsh event counts if average surrounding traffic speed decreases in the respective areas. Furthermore, in segments as the average occupancy increases, there is an increase in harsh accelerations, and as the average speed increases, fewer harsh breakings occur. It appears that traffic characteristics have the most statistically significant impact on the frequency of harsh events. Regarding driver behaviour collected data, the increase of maximum of event speed causes an increase in harsh events. Finally, a strong correlation between harsh events and time variation was found, indicating an overall increase of harsh events during nighttime.

Keywords: road safety; driver monitoring; driver behaviour; spatio-temporal analysis; smartphone application

INTRODUCTION

Every year, thousands of people lose their life or are seriously injured due to road crashes. The investigation of potential influencing factors that affect the likelihood of road crash occurrence has been of major interest in the scientific field of transport research. Some of the most important risk factors recognized in the road safety literature are human factors (speeding, distracted driving, driving under the influence of alcohol etc.), unsafe road infrastructure, unsafe vehicles and inadequate law enforcement of traffic laws (1, 2). Among them, human factors are likely to be the most crucial cause of road traffic fatalities and injuries every year and therefore the importance of studying how these factors can affect traffic risk is high.

Technological progress, especially in telematics, provides new potential for driver monitoring through smartphone applications and the respective data collection and processing (3, 4). Smartphones have the advantage of being programmable and feature a wide array of sensors, which have now become standard equipment that can be utilized for transport studies (such as accelerometer, digital compass, gyroscope, GPS, microphone and camera) and enable sensing applications, even without user engagement.

Naturalistic Driving experiments consist a thorough research method for the observation of everyday driving behaviour of road users (5). In this framework, a recent survey (6) aimed to identify aggressive and dangerous driving profiles using data collected from smartphone sensors. The dataset included more than 10,000 trips made in 129 urban and suburban environments. Trips were categorized into six separate groups (safe behavior, aggressive behavior, risky behavior, distracted behavior, aggressive/risky behavior, aggressive/distracted behavior) with increasing importance in terms of safety. The results showed that about 50% of the trips were classified as "safe trips", while 23.5% of the trips were realized by drivers driving above the speed limit and only 7.5% of trips were characterized as distracted.

The literature indicates that driving behaviour tracking systems appear to improve road safety level (7, 8). A relevant study (8) investigated driver behaviour using a data recording system adapted inside the car called Drive Diagnostics. The purpose was to assess the behaviour of drivers by measuring how they drive by examining specific parameters: the system collected data such as the acceleration of the vehicle, speed, position, fuel consumption, travel time, etc., in order to measure specific parameters such as frequency of change lanes, brakes and accelerations. Drivers were then classified into three categories, depending on the profile of the data. The results showed that the participants' initial exposure to the experiment had a significant effect on improving their driving behaviour and road safety and their access to system data helped them understand the errors they made on the road. However, the effect of Drive Diagnostics had disappeared by the end of five months.

The aim of another similar research (9) was to investigate the correlation between IVDR (In-vehicle data recorders), road network characteristics and accidents, and to examine the ability to predict events and identify high - risk locations on the road network. The study database included 3,500 sections of highway in Israel for which IVDR automated events were combined with road infrastructure and accidents. Negative binomial regression models were used that were adjusted for the relationships between road characteristics and driving events. Different types of events showed different relationships with the characteristics of the infrastructure and the different impacts on accidents on the different types of roads. Better road conditions were associated with a decrease in "braking" events and an increase in "speed alert" events, while road constraints and node closeness were associated with an adverse effect on events.

Another related study (10) highlighted the potential for monitoring and evaluating driver behaviour through continuous on-board data collection (OBD - On Board Diagnostics) and smartphones. The results were transferred to a user-friendly smartphone application. The study concluded that a scoring point system that monitors and provides feedback to the driver could be a very good incentive for most drivers to improve their driving behaviour.

One of the most important purposes of accident analysis is to identify and produce information to help policy makers to respond appropriately by selecting informed measures to prevent and reduce accidents. Spatial statistical mapping is the key to understanding the spatial and temporal occurrence of accidents, and spatial statistics include a set of techniques for describing and modeling spatial data (11). Spatial statistical analysis related to traffic accidents can be performed in a spatial database containing all the desired information and generating data levels from the available sources obtained from the field. Many of the factors contributing to accidents operate on a spatial scale (e.g. land use policy, demographics, infrastructure functioning).

Two research studies (12, 13) analyzed spatial patterns of road crash frequency for the state of Pennsylvania in the U.S. and developed the IRAS (Intelligent Road Accident System) system through which Malaysian police can monitor real-time road accidents, respectively. With respect to the IRAS system, it is based on GIS (Geographical Information System) and telecommunications technologies as it uses GPRS to transfer data to the Control Center. The purpose of a similar research (14), was to evaluate and represent the accident hotspots in a city in South India by modeling real accident location information in conjunction with various spatial attributes using spatial statistics in geographical information technology.

Furthermore, in a naturalistic driving experiment (15), consisted of 90 drivers, authors attempted to explore the relationships between driver characteristics and traffic safety-related events, and between traffic safety-related events and crash involvement while mitigating some of those limitations. Although authors concluded that there is still much to learn about the factors affecting the positive correlation between safety-related events and crashes, they found that a Bayesian multivariate Poisson log-normal model is shown to be useful to quantify the associations between safety-related events and crash risk while controlling for driver characteristics.

Regarding time variation, a set of studies have been investigating the differences in driving behaviour and safety through a variety of methods, namely driving simulators studies, in-depth analyses and Naturalistic Driving experiments. More specifically, researchers have found that driving during the night brings noticeable change in driver behaviour. Relative research was conducted to compare the behaviour of the driver's speed during the day and night, identifying important factors that affect the behaviourspeed under different lighting conditions (16). Speed profiles were recorded for 40 drivers under simulated daytime and nighttime driving conditions. The results proposed new velocity forecasting models, differentiated for daytime and nighttime driving, highlighting the effects of different geometric forecasts under different visibility conditions.

Speed distributions were examined (17); they were developed using environmental parameters related to both road geometry and lighting conditions. The authors compiled a comprehensive database in which geometric features, speeds, brightness and brightness were available for both night and day conditions. The results showed that the average velocities and deviations from the mean are significantly influenced by changes in the lighting parameters for the different conditions (sunny, cloudy and dark).

Drawing conclusions from past research and seeking to further build on them, the objective of this paper is to investigate driver behaviour and examine possible differences between daytime and nighttime by using data from smartphone sensors. A secondary goal is to examine the extent to which driving characteristics recorded by smartphone sensors, traffic metrics (velocity, traffic, occupancy) and geometrical road characteristics determine the driving behaviour of each driver in terms of harsh events (acceleration and braking), depending on its position on the road and more specifically on whether it is moving at a junction or on a road segment. In order to achieve the above objective, appropriate data analysis methods will be implemented.

METHODS

Data Collection and Processing

For the purpose of this research, data from three separate sources were analyzed. The study area consisted of two urban expressways that were examined, Mesogeion Avenue and Vouliagmenis Avenue, in Athens mainly due to the comparable number of traffic lanes and the separation of the two directions that they feature. Specifically, driving behaviour data, traffic characteristics of the two urban expressways under study as well as road geometry characteristics were collected. The first dataset concerned the driving behaviour of approximately three hundred (300) drivers in Athens and was collected using smartphone device sensors. The second one consisted of traffic characteristics which were collected through twenty-six (26) inductive loops, installed by the Traffic Management Centre of Attica Region in specific measuring positions on the two urban expressways under study. Finally, the third dataset was formed by geometry characteristics of the two urban expressways which were collected using the online mapping service provided by Google Maps. The data collection process is described in more detail in the following sections.

More specifically, an innovative data collection system using a Smartphone Application that has been developed by OSeven Telematics was exploited (www.oseven.io). OSeven implements ~~A~~advanced Machine Learning ~~techniques~~ algorithms ~~have been implemented~~ to detect harsh events and speeding, identify the trip transport mode, and recognize whether the user is a driver or a passenger. Consequently, the OSeven platformhelps drivers understand their weak points and motivates them to improve their driving behaviour.

The data recording is initiated automatically in the smartphone app when a driving status is recognized, and again it stops automatically when a non-driving status is recognized. For every trip a driver completes, a large amount of data is recorded, transmitted through Wi-Fi or cellular network and valuable critical information such as metrics, features, highlights and driving score is produced in order to evaluate driving profile and performance. The exposure indicators that are available include indicatively duration (seconds), total distance (mileage), trip duration, type(s) of the road network used, given by GPS position and integration with map providers e.g. Google, OSM, (highway, rural or urban environment) time of the day driving (rush hours, risky hours) and weather conditions. Moreover, the driving indicators which can reliably quantify the risk associated with a specific driving behaviour are the following: speeding (distance and time of driving over the speed limit and the exceedance of the speed limit), driver distraction (caused by smartphone use during driving), number and severity of harsh events number and severity of harsh events (braking and acceleration), harsh cornerings, driving aggressiveness (e.g. braking, acceleration). The basic operating frame of the data flow is shown in Figure 1.

Figure 1: The OSeven data flow system

It must be noted that since privacy and security consist two of the main principles, in the field of telematics, the OSeven platform has very clear privacy policy statements for the end users covering the type of data collected, the reason data is collected for, the time that data is stored and the procedures for data security based on encryption standards for data in transit and at rest. All this is done using state-of-the-art technologies and procedures in compliance with GDPR. In this framework all data has been provided by OSeven Telematics in an anonymized format.

Approximately three hundred (300) drivers participated in the smartphone naturalistic driving experiment in Athens and in time from August 2016 to November 2017 leading to the creation of two large databases of harsh accelerations and decelerations with thousands of events each. More specifically, during this period 219,757 harsh accelerations and 102,918 harsh decelerations were recorded. For each harsh event that was recorded, respective descriptive variables (speed of the event, maximum speed difference in two seconds during the event (a form of range), distance of the event) were recorded.

The second part of the data collection phase was implemented at the Traffic Management Center (TMC) of Attica Region. Specifically, 26 available measuring positions of traffic volume and occupancy were found in Vouliagmenis Ave. and Mesogeion Ave., installed by the TMC. It was decided that the utilized traffic parameters would be the average hourly traffic flow (normalized as average hourly traffic flow/lane, measured in vehicles/hour), average occupancy (measured as absolute percentage) and average aggregated speed (measured in kilometers/hour). It was assumed that the above parameters on the road sections which did not have a measuring position were equal to the average of the corresponding traffic sizes of the previous and subsequent road sections which had a precise measurement position. The junction traffic parameters were equal to the average of the traffic parameters of the ascending and the descending road section which their flows connected to. Finally, additional geometric characteristics of the two road axes were investigated. The respective data were collected with the assistance of the online mapping service Google Maps. Variables such as the number of entrances and exits of each junction under study, the number of outgoing and ingoing traffic lanes to and from the junction, the presence or absence of access roads, the number of right exits and entrances of the road segment, the presence or absence of bus lanes etc. were collected via visual inspection.

Spatial visualization of the collected data was then carried out through a Geographic Information System (GIS) application (ArcMap 10.3) which resulted in the visualization on the road network. In order to achieve the categorization of the harsh events in each numbered spatial unit that was defined, a geoprocessing model was used. In this way, each spatial unit included the frequency of harsh acceleration or harsh deceleration and the variables that characterized any harsh event. Since harsh events were not analyzed in isolation but as total frequencies depending on where they occurred, the values of the descriptive variables (minimum number, maximum number, standard deviation, range, mean) of the harsh events that occurred in the same spatial unit were calculated through the geoprocessing model.

Finally, in order to combine the driving behaviour high resolution data from the intelligent mobile phone sensors and the traffic sizes measured by the inductive loops placed by the TMC, the traffic measurement locations were added in the map under processing, based on their Cartesian coordinates. Consequently, a dataset was produced that imprinted the events of harsh accelerations and decelerations in Athens, the geometric characteristics of the road network and the measurement positions of traffic characteristics. Regarding the investigation of the time variation safety behaviour of the driver, the data were separated to daytime (07:00-22:00) and nighttime (22:00-07:00). A part of the junctions, segments and harsh events as drawn using the GIS program is shown in Figure 2. The purpose of this procedure was to investigate the driving behaviour with the main focus on the driver's position on the road.

Figure 2: Map of harsh events during daytime (left) and nighttime (right) on Mesogeion Avenue

Analysis Method

The data obtained from the processes described above, were analysed using multiple linear and lognormal regression models. Specifically, regression models were developed to model how parameters of driving behaviour, traffic characteristics and road geometry characteristics influence the frequency of harsh events in a defined spatial unit during daytime and nighttime, respectively. The basic equation (**Equation 1)** of the multiple normal regression models is the following:

$$y_i = b_0 + b_1 * x_{1i} + b_2 * x_{2i} + \dots + b_v * x_{ni} + e_i \quad (1)$$

Linear regression is a well-known simple technique used to model a linear relationship between a continuous dependent variable and one or more independent variables (18). In the analysis under consideration, the dependent variables were considered to be continuous due to the large number of events that were observed in the defined spatial units. The log-linear (log-normal) regression was applied as it described better the aggressive behaviour of drivers in both road segments and junctions. Both approaches were calibrated using the Ordinary Least Squares method.

In this case, eight statistical regression models forecasting harsh events were developed, as shown below:

- Model 1: frequency of harsh accelerations at junctions during the daytime
- Model 2: frequency of harsh accelerations at junctions during the nighttime
- Model 3: frequency of harsh brakings at junctions during the daytime
- Model 4: frequency of harsh brakings at junctions during the nighttime
- Model 5: frequency of harsh accelerations at road segments during the daytime
- Model 6: frequency of harsh accelerations at road segments during the nighttime
- Model 7: frequency of harsh brakings at road segments during the daytime
- Model 8: frequency of harsh brakings at road segments during the nighttime

Specifically, models for daytime and nighttime were designed to contain the same variables in order to enable model comparison. Then, all models were separately examined on whether the numerical results quantifying relationships between variables were of reasonable quality. For the correlation coefficient R^2, which indicates the percentage of the dependent variable that is explained by the independent variables, the highest possible values were sought.

It is important to mention that the values and signs of linear and lognormal regression coefficients (bi) must be reasonably explainable. Also, the value of the t-statistic control and the statistical significance level should be acceptable (> 1.7). Lastly, the elasticity (ei) that shows how responsive one variable is to a change in another but also the relevant influence elasticity (ei*) used for quantifying the influence of each individual variable is calculated and examined, in order to allow for the comparison between the influence of different variables within a single model. In particular, point estimates of elasticities (ei) are provided by the following **Equation 2**, for each value (i) in the sample:

$$e_i = (\Delta y_i / \Delta x_i) * (x_i / y_i) = \beta_i * (x_i / y_i) \qquad (2)$$

RESULTS AND DISCUSSION

The final dataset consisted of several types of variables, including parameters extracted from the naturalistic driving experiment (OSeven variables) as well as parameters extracted from the TMC and geometry characteristics of roads extracted from Google Maps. The maximum (MAX), minimum (MIN), the range (RANGE), the standard deviation (STD) and the mean (MEAN) value of the variables was calculated for all OSeven parameters. The models were the result of a series of tests in which many mathematical models were developed that included combinations of all the variables recorded.

Log-linear regression was preferred after several trials, in order to avoid negative projections from the models. Therefore, for all models, the dependent variable was the logarithm of the frequency of harsh events to traffic and the independent variables were the ones described below. Table 1 provides a description of the independent variables that were found to be statistically significant in the multiple lognormal regression models

TABLE 1 Description of the parameters used in the models

Independent Variables	**Description**
V	Average Traffic Velocity (km/h)
O	Average Traffic Occupancy (%)
Speed_Diff	Speed difference that caused the event (km/h)
Event_Speed	Speed of vehicle at the start of the event (km/h)
Distance	Distance of accelerometer
No.Left Entrances	Number of left entrances in junction
No.Right Exits/Entrances	Number of right exits from the segments and entrances to the road segments
Length	Length of road section (m)
Sideway	Presence (1) or absence (0) of sideway

The analysis was conducted using SPSS Statistics (19). In Tables 2 and 3, the parameters of the driving experiment and the multiple lognormal regression results are presented. The final models presented below are the best of multiple tests performed and therefore predict the dependent variable as best as possible. Correlation tests were performed, to avoid entering in the model any strongly correlated pair of independent variables. The coefficients of each model are described in the tables.

Results appear on Table 2 (for junctions) and Table 3 (for road segments) that follow:

TABLE 2 Summary table of the models developed for junctions

	Harsh accelerations								Harsh brakings							
	day				night				day				night			
	β	t	ei	ei*	β	t	ei	ei*	β	t	ei	ei*	β	t	ei	ei*
Constant	-1,234	-2,195			-0,113	-0,116			-0,347	-0,782			1,570	1,799		
V	-0,034	-4,064	1,063	-4,070	-0,056	-3,497	2,431	-5,081	-0,051	-6,263	1,149	-4,672	-0,069	-4,670	2,000	-11,690
MAX_Speed_Diff	0,037	2,411	-0,487	1,867	0,064	4,053	-0,897	1,875								
MAX_Event_Speed	0,007	2,379	-0,261	1,000	0,011	3,349	-0,478	1,000	0,008	2,999	-0,246	1,000	0,005	2,352	-0,171	1,000
No Left_Entrances	0,137	2,209	-0,080		0,173	2,981	-0,118									
Sideway									0,173	2,024	-0,045		0,170	2,036	-0,049	
R^2	**0,661**				**0,67**				**0,621**				**0,486**			

TABLE 3 Summary table of the models developed for road segments

	Harsh accelerations								Harsh brakings							
	day				night				day				night			
	β	t	ei	ei*	β	t	ei	ei*	β	t	ei	ei*	β	t	ei	ei*
Constant	-2,320	-7,810			-1,772	-4,887			-1,530	-2,555			1,074	1,317		
O	0,068	3,605	-0,427	2,202	0,377	4,876	-1,517	3,748								
V									-0,013	-1,733	0,390	-3,164	-0,040	-2,833	3,648	-16,480
MAX_Event_Speed	0,006	2,575	-0,194	1,000	0,004	1,876	-0,405	1,000								
MIN_Speed_Diff									-0,038	-1,998	-0,450	3,610	-0,030	-2,272	-1,150	5,175
MIN_distance	-0,302	-4,212	0,357	-1,839	-0,276	-4,447	0,773	-1,910	-0,239	-3,217	0,260	-2,099	-0,210	-3,873	0,510	-2,254
No Right Exit/Entrance	0,075	2,487	-0,073		0,103	3,691	-0,307									
Length									0,001	2,385	-0,455	1,000	0,001	1,815	-0,222	1,000
R^2	**0,565**				**0,676**				**0,571**				**0,502**			

From the results reported on Tables 2 and 3, several conclusions can be drawn:

- The determining factor affecting the driver behaviour when moving at a junction is average traffic velocity, due to its highest elasticity. More specifically, as the average speed of traffic on a junction increases, there is a decrease of harsh event frequencies.
- The average occupancy and average traffic speed also have a major impact on driver behaviour on a road segment, according to the calculated elasticities. As the mean occupancy at segments increases, so do the harsh accelerations in that area. However, as the average traffic velocity increases the harsh brakings decrease.
- An increase in the maximum speed of events causes an increase in harsh events occurring on a junction and a road segment. The maximum event speed was not found to be statistically significant for the prediction of frequency in harsh braking in road segments.
- The increase of the maximum of the speed difference that caused the event indicated an increase in harsh accelerations in junctions, whereas the increase of the minimum of the speed difference

decreases the frequency of harsh brakings in a road segment.

- The last variable that was collected from smartphone sensors and has an impact on driver behaviour is the distance of the occurring event. An increase in the minimum value of distance causes decrease of harsh events in segments.
- Regarding the geometrical characteristics of road, the results indicated that: as the length of the road segment increases, more harsh brakings occur. Moreover, as the number of left entrances in a junction and the number of right exits and entrances in a segment increase, more harsh accelerations occur. Finally, the existence of a sideway showed a positive effect on the occurrence of harsh brakings in a junction.
- The existence of a constant with a negative coefficient indicates the presence of additional unrecognized factors that cause a decrease in the frequency of harsh events.
- In all models, the maximum possible correlation between the dependent and the independent variables was achieved, with the R^2 coefficient ranging from 0.486 to 0.676. This value is satisfactory.

The sensitivity analysis that was conducted confirms the above results. As shown in figure 3 as the average traffic speed increases the frequency of harsh decelerations in a road segment decreases.

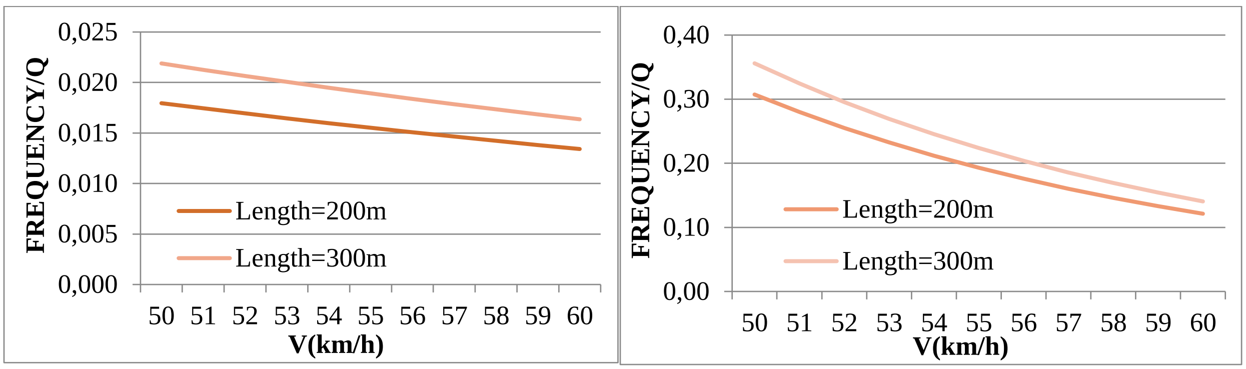

Figure 3: Sensitivity analysis for harsh brakings in segments during daytime (left) and nighttime (right)

Additionally, the sensitivity analysis displayed in Figure 4 shows to what extent the increase of the maximum event speed, influences the frequency of harsh accelerations in a junction. It is found that the higher the increase in max event speed the more harsh accelerations occur.

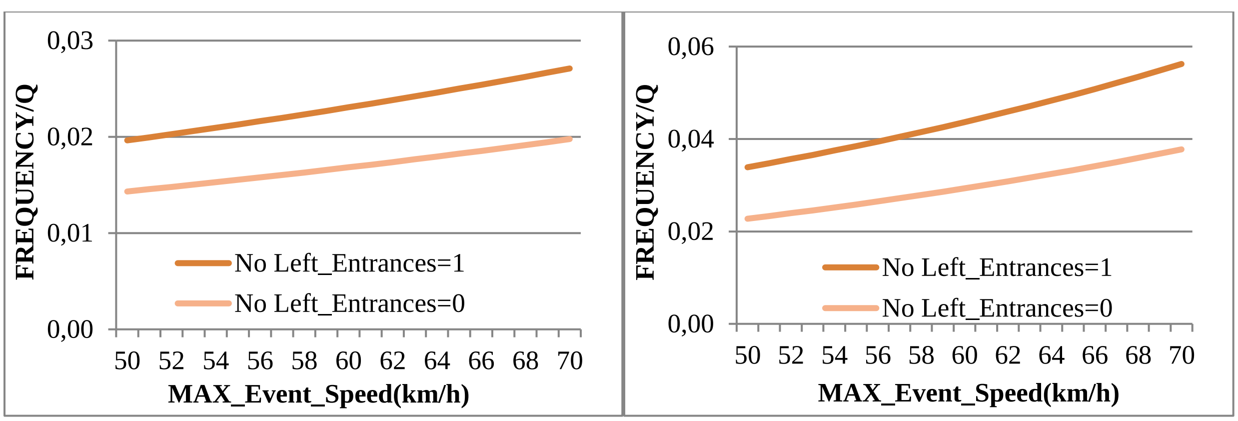

Figure 4: Sensitivity analysis for harsh accelerations in junctions during daytime (left) and nighttime (right)

After analyzing the models and their results, a comprehensive view of all models emerges. All models include: an independent variable that relates to a geometrical characteristic of the region, a traffic variable and finally, one or two parameters describing the behaviour of the driver, as recorded from the smartphone sensors. Overall, it appears that the traffic variables have the highest impact on the occurrence of harsh events. For most of the variables, coefficient values seem to change between daytime and nighttime in a similar direction for both acceleration and braking events.

Regarding the R^2 of all eight regression models, it is high enough to indicate that the examined independent variables can predict the dependent one. Moreover, the residuals in all models were found to be normally distributed, which means that the assumption is valid. As an example, the normal probability plot of the residuals for harsh accelerations and brakings in road segments during daytime is presented in Figure 5.

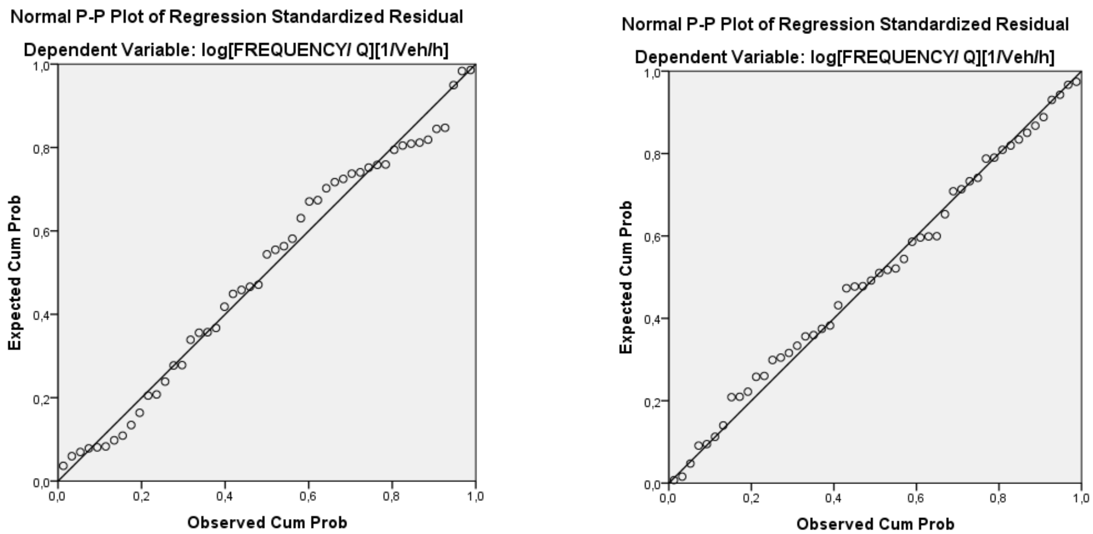

Figure 5: Normal plot of regression residuals of harsh accelerations (left) and harsh brakings (right) in road segments during daytime

It is believed that the current research has captured the majority of the parameters that can meaningfully and informatively describe and predict harsh event frequency. This was primarily achieved by taking into consideration three main pillars of road safety: (i) driver behaviour, (ii) traffic parameters and (iii) road network/geometric characteristics. However, the examination of additional independent variables, either not recordable or partially recorded and thus, excluded from this analysis, may provide more insight on the examined dependent variables.

CONCLUSIONS

The aim of this paper was to identify the safety behaviour of the driver using data collected from smartphone sensors, traffic and road geometry characteristics. The data were depicted spatially using GIS. Events of harsh driver behaviour (acceleration and deceleration) were mapped to delimited segments and junctions of two avenues in Athens, Greece, during daytime and nighttime, by categorizing the data from smartphones based on the time of the event. For the analysis, eight log-linear regression models were developed. The results of this study indicate that traffic parameters (speed and occupancy) have the most statistically significant impact on the frequency of harsh events compared to road geometry characteristics and driver behaviour data.

However, it should be mentioned that the present analysis has some limitations. The lack of data on driver characteristics (e.g. age, gender, etc.) in this study limits the potential of additional explanation and more correct detection of the factors that influence the behaviour of the driver. It is important to highlight that there is little or no previous experience on analyzing and predicting harsh events in road segments and junctions through microscopic driving behaviour metrics collected from smartphone devices, and therefore the results of the present research cannot be directly compared to those of existing literature.

A number of observations can be made also for the data collection system; the use of smartphone sensors alone has some limitations compared to full vehicle instrumentation in naturalistic driving experiments (e.g. driver reaction time, brake response, video data). The main purpose of the models developed is to identify parameters that affect the driver behaviour, regarding also the time variation. It is possible that rapid developments in smartphone technologies may boost the possibilities for additional driving behaviour metrics and additional analyses. The expected increase in the use of modern smartphones may be seen as an opportunity to exploit the wealth of data that can be made available by smartphone sensors, in order to find new ways to identify the key risk factors while driving and improve road safety.

The use of driver behaviour feedback services via smartphones, could potentially allow more drivers to be informed and improve their behaviour by reporting their driving performance. The results of this survey may be generalized to apply to areas other than the spesific research area. Before any generalization, however, adjustments should be made for possible variations in the road environment and traffic. For future research, it would be useful to have an analysis based on even more data such as driver characteristics (e.g. gender, driving years, age, etc.), vehicle characteristics, the condition of the road or the weather conditions by determining the month by where the harsh event took place.

ACKNOWLEDGMENTS

The authors would like to thank OSeven Telematics, London, UK for providing all necessary driver behavior data exploited to accomplish this study. The authors also like to thank the Traffic Management Centre of Attica for providing all traffic parameter data exploited to accomplish this study.

This research is co-financed by the European Union - European Regional Development Fund (ERDF) and Greek national funds through the Operational Program "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) of the National Strategic Reference Framework (NSRF) - Research Funding Program: BeSmart - Multi-modal driver behaviour and safety support system on the basis of smartphone applications.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Eleni-Konstantina Frantzola Author, Armira Kontaxi Author, George Yannis Author; data collection: Eleni-Konstantina Frantzola Author; analysis and interpretation of results: Eleni-Konstantina Frantzola Author, Armira Kontaxi Author, George Yannis Author; draft manuscript preparation: Elina Frantzola Author Armira Kontaxi Author, George Yannis Author. All authors reviewed the results and approved the final version of the manuscript

REFERENCES

1. WHO: Global status report on road safety by World Health Organization, 2018

2. Peden, M. World report on road traffic injury prevention, 2004

3. Vlahogianni, E.I., and E.N. Barmpounakis. Driving analytics using smartphones: algorithms, comparisons and challenges. *Transport. Res. C: Emerg. Technol*. 2017 Jun 1;79:196-206.

4. Ziakopoulos, A., Kontaxi, A., Yannis, G., Fortsakis, P., Kontonasios, K.-N. & Kostoulas, G., (2020b). Advanced driver monitoring using smartphone applications: The BeSmart project. Proceedings of 8th Transport Research Arena TRA 2020

5. Ziakopoulos, A., Tselentis, D., Kontaxi, A., & Yannis, G. (2020a). A critical overview of driver recording tools. Journal of Safety Research.

6. Mantouka EG, Barmpounakis EN, Vlahogianni EI. Identification of driving safety profiles from smartphone data using machine learning techniques. *Safety Science*. 2019 Feb 11.

7. Roetting M, Y.H. Huang, J.R. McDevitt, D. Melton. When technology tells you how you drive—truck drivers' attitudes towards feedback by technology. Transportation Research Part F: Traffic Psychology and Behaviour. 2003 Dec 1;6(4):275-87.

8. Toledo, T., Musicant, O., & Lotan, T. In-vehicle data recorders for monitoring and feedback on drivers' behaviour. *Transportation Research Part C: Emerging Technologies*, 2011. 16(3), 320-331.

9. Gitelman, V., Bekhor, S., Doveh, E., Pesahov, F., Carmel, R., & Morik, S. Exploring relationships between driving events identified by in-vehicle data recorders, infrastructure characteristics and road crashes. *Transportation research part C: emerging technologies*,2018. 91, 156-175.

10. Yannis, G., Tselentis, D., Papadimitriou, E., & Mavromatis, S. Star rating driver traffic and safety behaviour through OBD and smartphone data collection. In Proceedings of the International Conference on Artificial Intelligence and Intelligent Transport Systems, Technical University of Brest, Brest, Belarus., 2016

11. Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. Accident Analysis & Prevention, 135, 105323.

12. Aguero-Valverde, J., & Jovanis, P. P. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, 2006. 38(3), 618-625.

13. Mahmud, A. and Zarrinbashar, E. Intelligent GIS-Based Road Accident Analysis and Real-Time Monitoring Automated System using WiMAX/GPRS. *International Journal of Engineering*, 2008 2(1), pp.1-7.

14. Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia-Social and Behavioural Sciences*, 2011. 21, 317-325.

15. Wu, K. F., Aguero-Valverde, J., & Jovanis, P. P. (2014). Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. Accident Analysis & Prevention, 72, 210-218.

16. Bella F, Calvi A, D'Amico F. Analysis of driver speeds under night driving conditions using a driving simulator. *Journal of safety research*. 2014 Jun 1;49:45-e1.

17. Bassani, M., Catani, L., Cirillo, C., & Mutani, G. Night-time and daytime operating speed distribution in urban arterials. *Transportation research part F: traffic psychology and behaviour*, 2016. 42, 56-69.

18. Washington, S. P., M. G. Karlaftis, F. L. Mannering, 2010. Statistical and econometrics methods for transportation data analysis. Chapman and Hall/CRC.

19. IBM Corp. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp., 2015