Rethinking transport
27–30 April 2020

TRA 2020 HELSINKI

Proceedings of 8th Transport Research Arena TRA 2020, April 27-30, 2020, Helsinki, Finland

Spatial analysis of driver safety behaviour using data from smartphones

Ilias Parmaksizoglou[a], Dimitrios I. Tselentis[a], George Yannis[a]*

[a]National Technical University of Athens
Department of Transportation Planning and Engineering
5, Iroon Polytechniou str. GR-15773, Athens, Greece

Abstract

The aim of this research is to conduct a spatial analysis of driver safety behavior using data from smartphones. More specifically, it is investigated how the number of harsh accelerations and decelerations per day, which are key elements of everyday driving, is influenced with both the road environment and road users' behaviour. Data were processed in a GIS computer software, resulting to the creation of new tables describing the phenomena observed on the signalized arterial studied (Leoforos Mesogeion) in node and link areas. Additionally, analytic maps were developed that aimed to indicate patterns of the accumulation and ranking of the harsh events in the selected road segments. Finally, four linear regression models were developed, which demonstrated speed as the most statistically significant factor in predicting the number of harsh events per day on a region basis.

Keywords: driver behaviour; harsh events; spatial analysis; GIS; linear regression

* Corresponding author. Tel.: +30-210-772-1326;
E-mail address: geyannis@central.ntua.gr

1. Introduction

According to World Health Organization (WHO, 2018), the total number of road fatalities worldwide continues to climb, reaching a high of 1.35 million in 2016. In the European Union, road crashes were the fifth cause of death in 2016, with roughly six people out of every 100,000 dying on the roads of the European Union due to a road crash. Road safety is a very complicated scientific field of transport research due to the random nature of crash occurrence both spatially and temporally. This paper is coping with the problem of analyzing driver behaviour from a spatial point of view.

Harsh acceleration, harsh breaking and harsh cornering events are three significant indicators for driving risk assessment (Johnson et al., (2011), Bonsall et al., (2005)) especially for evaluating driving aggressiveness. This is because they are strongly correlated with unsafe distance from adjacent vehicles, possible near misses, lack of concentration, increased reaction time, poor driving judgement or low level of experience and involvement in situations of high risk (e.g. marginal takeovers). The correlation between HA and HB events with driving risk has been highlighted in the scientific papers published by (Tselentis et al., (2017), Bonsall et al., (2005)) and it has been widely recognized by the insurance and telematics industry.

Regarding the data collection approaches used so far, the most common methodologies applied include driving simulators (Desmond et al., (1998), Lenné et al., (1997)), questionnaires (Matthews, et al., (1998)) combined with simulators and naturalistic driving experiments (Toledo et al., (2008), Birrell et al., (2014)), while the most common method of monitoring driving measures included recorders that relate to the car engine (Zaldivar et al., (2011), Backer-Grøndahl and Fridulv (2011)) and smartphones (Vlahogianni and Barmpounakis (2017)). Naturalistic experiments provide a wide perspective of understanding typical microscopic travel and driving behaviour. A naturalistic study can help (Regan et al., (2012)): a) determine accident risk, b) study the interaction between road/ traffic conditions and driver's behaviour, c) understand the interaction between car drivers and vulnerable road users, d) specify the relationship between driving pattern and vehicle emissions and fuel consumption, and many other aspects of traffic participation. The most popular devices for monitoring driving measures are recorders that relate to the car engine (Zaldivar et al., (2011), Backer-Grøndahl and Fridulv (2011)) such as on-board-diagnostics (OBD) devices and smartphones (Vlahogianni and Barmpounakis (2017)).

One of the first researches examined was conducted by Anderson (2007), who attempted to identify the most frequent location where road accidents have occurred in northern London in the following three ways: at the level of the administrative area, with the sum of the incidents within it at road level and finally creating a road accident density area after they were used as points through a kernel density estimator. Each method used gave a different picture of the problem with both advantages and disadvantages. An effort to illustrate the road safety problem in the road environment was undertaken by Ivan and Haidu (2012), who used the spatial means of road accidents for the period 2010 to May 2012 for the city of Cluj-Napoca in north-west Romania in order to identify the point representing the distribution. They also calculated their standard distance as well as the lack of standard error, which showed that the incident distribution is along the main road network.

In addition, Aguero-Valverde and Jovanis (2006) in their investigation of fatal road accidents and injuries in the Pennsylvania region, found that coefficient values from the variables related to the transport system are significant and are significantly approaching the values of the Negative Binomial Distribution model that coincides with the clustered pattern. Also, in the same survey, the age group of residents aged less than 15 years has been associated with fatal road accidents, while some others not (aged 65+). Another important research for linking a number of road accident factors was conducted by Noland and Quddus (2004) across England per electoral compartment. Although very few accidents occur in England than others, efforts are being made to reduce them further. Some of the variables used concerned the characteristics of the road network where it was found that first and second category roads were related to road accidents. Greater correlations were for road accidents with mild but also serious injuries. This probably occurs due to the fact that speed limits are often exceeded in these roads but also that the interaction of the vehicles tends to be higher (Noland and Quddus, 2004). The motorways were relatively safer despite the fact that the speed limit was higher (Noland and Quddus, 2004). In the same survey, road accidents were associated with junctions and roundabouts and only correlated with minor injuries was found.

The aim of this paper is to conduct a spatial analysis of driver safety behavior using data recorded from smartphone devices. More specifically, this research investigates how harsh accelerations and decelerations per day, which are key elements of everyday driving, interact with both the road environment and road users' behavior. Data were processed in a GIS computer environment, resulting to the creation of new tables describing the phenomena observed on the road map of the present study (Leoforos Mesogeion) in both node and link areas. Additionally, analytic maps were developed that aimed to indicate patterns of the accumulation and ranking of the harsh events in the selected road segments. Finally, four linear regression models were developed in order to identify the most statistically significant factor in predicting harsh events per day on a region basis.

The spatial visualization of the harsh events on a map helped to observe where events of different intensity level are accumulated on each road segment and therefore, to acquire a better picture of the problem studied. Additionally, four new datasets were created for the examined observations per area of the signalized arterial through map processing. In addition, the geometric characteristics of the road segments analyzed, taken from the Google Maps platform and based on-site autopsy, were also used as contributing factors. Using these datasets, a statistical analysis was performed, and four linear regression models were developed to determine the most important factors that can influence the harsh events (decelerations, accelerations) in junction areas and link areas for the signalized arterial to be examined. Due to the large number of variables, several acceptable models have been developed that have adequately described the characteristics of the research areas. However, for the better interpretation of the phenomena, the aim was to develop only two different models depending on the type of the area. This means that the parameters that affect harsh events in different regions are sought regardless of the type of incident. The statistical checks required to accept or reject mathematical models were an integral part of the results.

2. Experimental data collection

To achieve its objectives, this research makes use of driving behaviour data recorded through a specific mobile application. OSeven uses an integrated system for recording, collecting, storing driving behaviour data and data processing using advanced machine learning algorithms. The recorded data originate from the various sensors of smartphones and data fusion algorithms. The mobile phone application records the user’s behavior using the device sensors and uses a variety of APIs to read recorded sensor data and to be temporarily stored in the smartphone database. This innovative application, applied to large-scale data collection and analysis, presents new challenges by gathering a large amount of data for analysis. The anonymized data received from OSeven Telematics in the form of databases that contained a total of 194,850 observations from 319 users. The datasets exploited included data about each harsh event observation such as the user ID, the trip ID, the time of the event, the latitude and longitude, the driving speed at the beginning of the event, the type of road network, the accelerometer value and the speed difference between the start and end of the event.

Data capture automatically starts in mobile phone applications when a driving status is recognized and automatically stops when a non-driving state is recognized. The recorded data originate from the various sensors of smartphones and data fusion algorithms. The mobile phone application records the user’s behavior using the device sensors and uses a variety of APIs to read recorded sensor data and to be temporarily stored in the smartphone database. Upon completion of the route, the application transmits all collected data to the OSeven backend office central database via a suitable communication channel, such as a Wi-Fi network or a cellular network such as 3G / 4G network based on user settings. Data is stored using advanced data encryption and data security techniques, in accordance with national laws and EU directives on personal data protection.

Once the data is stored on the server for centralized processing and data size reduction, the collected data is converted into significant behaviors and road safety-related parameters. This is accomplished using two large data processing methods including two families of techniques, large data mining techniques and Machine Learning algorithms. It is noted that all ML algorithms developed for raw driving data analysis e.g. harsh event detection is company-owned and therefore all relevant information is kept confidential.

Through the above-described process, two large datasets with observations from the region of Attica, for harsh accelerations and harsh decelerations respectively, were acquired for this research. The two datasets were projected on a map containing the main roads of the Greek road network. The examined road arterial (Leoforos Mesogeion), was divided in node areas and link areas through very specific attributes.

For the link areas, the following parameters were considered:

- Each area is defined from the end of the previous node to the beginning of the next node.
- The shape of the area is polygonal according to the geometric characteristics of the road.
- Each section between two nodes is divided into an uphill area and a corresponding downhill area.
- Section width was defined by the estimated road arterial and width 12 meters to the right for the uphill and left for the downhill

For the node areas, the following parameters were considered:

- The area affected by the node is a circle.
- The center of the node is at the connection of Leoforos Mesogeion to its vertical axes.
- The radius of influence of each node is fifty meters.
- In the case of intersection of the node regions, the intersectional surfaces intersect at the two nodes forming non-circular regions

To analyze at a region level, a way to match each observation to their corresponding area had to be found. Through ArcMap iterative processes the division of observations by position in the road arterial was achieved. In addition, the frequency of observations in each area was easily calculated, and statistics such as the mean, standard deviation, maximum, minimum and range of driving behaviour concerning each defined area were extracted. Driving behaviour was described by information stored in every harsh event, such as the Event Speed of the observation, the Distance occurred in the harsh event and the Difference in Speed of the event.

Through this methodology the final tables that were statistically examined were created and contained data that described driver behaviour and the road environment in order to correlate both of these characteristics with the frequency of the harsh events per day. The four tables created described harsh accelerations/decelerations in node/link areas.

3. Methodological approach

3.1. Statistical analysis

In order to achieve the goals of the study, regression analysis was conducted through linear regression as the nature of the depended value (continuous), deemed it appropriate. The criteria used to evaluate a model of linear regression are the signs and values of the coefficients β_i of the equation, the statistical significance, the quality of the model and the error of the equation. As far as the coefficients of the equation are concerned, there should be a reasonable interpretation of their meanings. The positive sign of the coefficient indicates an increase of the dependent variable by the increase of the independent one. Instead, a negative sign implies a reduction of the dependent variable by increasing the independent one.

The significance of a variable in the model is determined by the t-ratio value. This index refers to each of the variables separately It is essentially the result of the division of the estimated value coefficient by its standard deviation. The standard deviation is a magnitude that shows the consistency with which the value of that factor has been calculated. This means that the actual value of the 95% confidence interval (significance level) lies within the center of the calculated value of the coefficient and ends this +/- the standard deviation. The greater the absolute value, the greater the influence of that variable on the final result. Depending on the level of significance to which the results of the survey are concerned, there are tables giving the t-ratio value above which the particular variable should be included in the model. Therefore, for a 95% confidence interval, a variable may remain in the template if the absolute value of the t-ratio of its factor is greater than 1.96.

The quality of the model is examined with the R^2 adjustment factor. The R^2 factor is used as an indicator of a good fit for the linear model. This coefficient expresses the percentage of variability of the variable Y explained by the variable X. It takes values from 0 to 1. The closer the R^2 value to the unit is, the stronger the linear relation of dependence of the variables Y and X is made. The R^2 factor has a comparative value, which means that there is no specific R^2 value acceptable or discardable, but between two or more models is chosen as more appropriate than the higher value of R^2. The coefficient R^2 can be used as a measure of the strength of the linear relationship regardless of whether X gets set values or is randomly variable. For the purpose of this study an appropriate R^2 is considered between the values of 0.4 – 0.75.

It should be mentioned that as part of the spatial analysis, a GIS program (ArcMap) was used, as due to the large volume of data, spatial processing without Geoinformatics tools would be virtually impossible. Furthermore, at the level of statistical analysis, the study used basic knowledge of statistics and the modelling was done through linear regression, with a special statistical software (IBM SPSS 21.0).

3.2. Variable correlation

Due to the large number of variables, several acceptable models have been developed that have adequately described the characteristics of the research areas. However, for the better interpretation of the phenomena, the aim was to develop only two different models depending on the type of the area. This means that the parameters that affect harsh events in different regions are sought regardless of the type of incident. The statistical checks required to accept or reject mathematical models were an integral part of the results.

The correlation between the variables was first examined. At this point, the maximum possible correlation between dependent and independent variables and zero correlation between the independent variables was sought. Absolute values of the correlation coefficients near the unit show a strong correlation, while values close to zero reveal a non-existent correlation between the variables. In fact, a correlation between two variables is considered small when the absolute value of the Pearson r correlation index is less than or equal to 0.5 ~ 0.6 ($r \leqslant 0.5 \sim 0.6$).

For each model examined, the criteria that should be met are:

- The above models have both Adjusted R^2 between 0.4 - 0.75, therefore describe well the dependent variables.
- Absolute t – ratio is greater than 1.7 for each independent variable
- The significance level of the variables is less than 5%
- Constant is small
- The variables introduced into the model and their symbols are explained logically

After checking correlation between variables different models were examined. Their quality was measured by the index R^2 but also by the residuals plots. A residual value is a measure of how much a regression line vertically misses a data point. Regression lines are the best fit of a set of data. A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.

Finally, the degree of influence of the independent variables on the dependent variable contained in the mathematical relationship of the model is quantified through the magnitude of the relative influence. The calculation of this magnitude is based on the theory of elasticity and reflects the sensitivity of the dependent variable to the change of one or more independent variables. Elasticity is dimensionless and does not depend on the units of measurement of the variables. In conjunction with the sign of the variables, it is possible to determine whether the increase of an independent variable results in an increase or decrease in the dependent one.

4. Results

4.1. Color grading of the observations

In order to assess the quality of the examined road arterial, a color grading was performed, with the probability of at least one harsh event per day for every link or node of the road environment as its factor. Although a dataset with more observations for the examined road arterial would possibly demonstrate more noteworthy conclusions, the results from the color grading were deemed appropriate. The true randomness of the observations on the arterial were displayed, with the probability of at least one harsh event per day on links and nodes, varying in a non-accumulative way from area to area. Finally, the areas that presented a higher probability for a harsh event were displayed in visual way. In the following characteristic map of a part of the examined road arterial, such areas were displayed with a red color, areas that were prone to harsh events, but with less probability for harsh events to occur were displayed with yellow color and areas with relatively low probability for harsh events were displayed with green color.

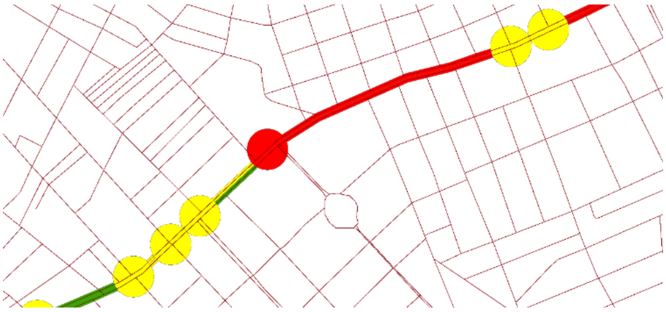

Fig. 1 Color grading of a selected area of the road arterial

4.2. Harsh events on nodes

After numerous tests performed, the best model to describe the factors causing harsh deceleration and acceleration events (Mod_Freq_Acc & Mod_Freq_Brk) in a node area features the independent variables of the typical deviation of event speed (STD_Event_Speed), the maximum speed difference (MAX_Speed_Diff) and the number of right exits from the node (Right_Exits) for both models, as shown in Table 1. The best models developed have an R^2 of 0.755 and 0.688, and an adjusted R^2 of 0.719 and 0.641 for the number of harsh decelerations and accelerations respectively.

Table 1. Model results for harsh decelerations per day in node areas.

	Coefficients	Bi	t - ratio	Significance
Harsh decelerations model	Constant	-0.112	-2.532	0.020
	Right_Exits	0.029	2.432	0.025
	STD_Event_Speed	-0.019	-6.647	0.000
	MAX_Speed_Diff	0.005	7.496	0.000
Harsh accelerations model	Constant	0.275	2.229	0.037
	Right_Exits	0.100	3.348	0.003
	STD_Event_Speed	-0.058	-4.987	0.000
	MAX_Speed_Diff	0.008	3.632	0.002

It is noticed that the increase in the right exits of a junction on Leoforos Mesogeion augments by 2.9 % the harsh decelerations per day. This is obviously logical, as the complexity of the node increases. Moreover, when the speed difference of a harsh event is maximized at the examined area, it is obvious that the risk of high speeds results in an increase in the sudden deceleration by 0.6 %. Finally, it is observed that as the standard deviation of the average event speed in kilometers per hour per area examined increases, the harsh decelerations per day decreases by 1.9 %. It can be observed that as in the example of slowdowns, the increase in right-wing exits at a junction on Leoforos Mesogeion increases by 10 % the sudden deceleration per day. Correspondingly, for every one kilometer per hour increasing the maximum speed difference per area, the harsh accelerations are increased by 0.8 %. Finally, as the standard deviation of the event speed increases the harsh accelerations the decrease by 6.8 %.

4.3. Harsh events on links

Working similarly to the work done for the node areas, after many tests the best model that correlates best variables to describe the factors that cause harsh decelerations and accelerations per day (Mod_Freq_Acc & Mod_Freq_Brk has the independent variables of minimum distance of event (MIN_distance), the existence of bus lane (Bus_Lane) and the length of link area (Length). The best models developed have an R^2 of 0.493 and 0.587, and an adjusted R^2 of 0.443 and 0.545 for the number of harsh decelerations and accelerations respectively.

It is found that the existence of the bus-lane increases the complexity of the road section and causes an increase of 3.1% in the harsh deceleration in the link areas. Furthermore, the increase of the length of the link increases the ability to develop speed for the driver and therefore makes sense to contribute to the increase of harsh deceleration by 0.008% for each meter of the road section. On the contrary, the increase in the minimum observation distance value means less harsh events for the area and reasonably reduces the harsh decelerations per day by 2.4% for each meter that increases the minimum value in the area.

Table 2. Model results for harsh decelerations and accelerations per day in link areas.

	Coefficients	Bi	t - ratio	Significance
Harsh decelerations model	Constant	0.0810	4.535	0.000
	Bus_Lane	0.0310	2.794	0.009
	Length	0.0001	1.900	0.067
	MIN_Distance	-0.0240	-3.540	0.001
Harsh accelerations model	Constant	0.0710	2.243	0.037
	Bus_Lane	0.0440	2.940	0.006
	Length	0.0002	4.056	0.000
	MIN_Distance	-0.0400	-3.126	0.004

As in the example of the harsh decelerations, the 4.4% increase observed per day due to the bus-lane is logical. Similarly, the increase in length increases the harsh accelerations by 0.022% for each meter of the road section. Finally, in harsh accelerations, the increase in the minimum observation distance value causes a decrease of 4% for each meter that increases the minimum distance value in the area.

5. Conclusions

This research comprised of two parts, those of spatial and statistical analysis. Regarding spatial analysis, the goal of this research would not be achieved without the use of a GIS computational environment. The visualization of the events occurred gives a clear advantage in understanding the current state of the area under examination. On the other hand, the visualization of the events and the construction of analytic maps also helped in the observation of patterns of accumulation and gradation in the phenomena and in making right decisions before statistical analysis. As for the statistical analysis, the determination of the factors that significantly influence the number of harsh events per day was achieved. Those factors were originating both from road characteristics and driver behavior data captured by the platform used.

As for the road characteristics, the increased number of right exits from a node leads to an increase in the number of harsh events per day, possibly due to the complexity of the road system. Additionally, when the length is increased in areas between nodes, a higher frequency of harsh events per day was observed. The existence of a bus-lane seems also to affect incrementally the harsh events frequency in link areas. With regards to driver behavior factors, the maximization of the speed difference contributes to an increase in the number of harsh acceleration events, as it indicates the presence of more intense phenomena in the junction areas. Respectively, an increase in the driving speed of harsh events (in absolute values) also causes a risk increase in the event of harsh decelerations. Conversely, increasing the standard deviation from the average speed per area seems to reduce the occurrence of harsh events in junction areas. Finally, the event distance was also observed to be significant in the indication of the amount of intensity of the harsh event and its minimization, proved to cause a reduction in the frequency of the harsh events and respectively the risk of the link areas.

It should be mentioned that spatial analysis has a high value when applied comparatively to similar road networks and this is the reason why several segments and nodes of the same arterial are analyzed. The results of this study can potentially be exploited to identify where further policy measures would help improve road safety. It would also be very interesting to exploit a larger sample of drivers in the future as well as to combine traffic flow data for the road under examination so that the event frequency weights also represents the traffic of the road segment examined.

5.1. Future research

The use of GIS programs can help significantly in identifying hazardous areas of the road network and, as a result, it can help in stronger and more targeted to policing in order to create safer road conditions for both car users and other drivers (motorcycles, bicycles) and pedestrians. The results of this study can be exploited by policing agents in the research area to identify where further policing would help improve road safety. In addition, the combination of data from smart phones in conjunction with GIS programs can create real-time network monitoring capabilities, which can give a clear lead to improved road safety, but also for better tracking of the road network and early intervention in case of need.

It would also be advisable to check the research road under examination (Leoforos Mesogeion) at the points where there was an increased chance of a harsh events and taking into account the important factors indicated by the statistical analysis to investigate any interventions in the geometry and the environment of it. Finally, it would be particularly useful to have the areas of increased chance of a harsh event indicated in this paper, to be correlated with existing black spot maps, in order to reinforce the arguments for intervention in the road environment.

It would be very interesting to observe the same variables in a larger sample of drivers. Specifically, if more driver data were captured and stored to the original databases, the results that would arose, would be more reliable with possibly stronger correlation between the variables. Additionally, it would be quite interesting to combine traffic flow data for the arterial under examination so that the event frequency weighing, is more representative of the traffic that the road examined has.

It would also be interesting to present an analysis based on more data such as driver gender, age, psychological status and vehicle characteristics (horsepower, vehicle type, age, etc.) as it would certainly lead to better conclusions and to more reliable and objective models. Also it would be beneficial to research a road arterial and analyze it by its temporal characteristics as well. This could help identifying the periods in the day most prone to harsh events, as well as to examine whether some variables are more important in a part of it. Finally, it would be very interesting to examine similar factors on road environments with different geometric characteristics and use, such as motorways, interchanges and urban road networks and see how they influence the development of harsh events.

Acknowledgements

The authors would like to thank OSeven Telematics, London, UK for providing all necessary data exploited to accomplish this study.

References

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

Anderson, T. (2007). Comparison of spatial methods for measuring road accident 'hotspots': a case study of London. Journal of Maps, 3(1), pp.55-63.

Backer-Grøndahl, Agathe, and Fridulv Sagberg. Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Safety Science Vol. 49, No. 2, 2011, pp. 324-330. https://doi.org/10.1016/j.ssci.2010.09.009

Birrell, Stewart A., Mark Fowkes, and Paul A. Jennings. Effect of using an in-vehicle smart driving aid on real-world driver performance. IEEE Transactions on Intelligent Transportation Systems Vol. 15, No. 4, 2014, pp. 1801-1810. https://doi.org/10.1109/TITS.2014.2328357

Bonsall, Peter, Ronghui Liu, and William Young. Modelling safety-related driving behaviour—impact of parameter values. Transportation Research Part A: Policy and Practice Vol. 39, No. 5, 2005, pp. 425-444. https://doi.org/10.1016/j.tra.2005.02.002

Desmond, Paula, Peter Hancock, and Janelle Monette. Fatigue and automation-induced impairments in simulated driving performance. Transportation Research Record: Journal of the Transportation Research Board No. 1628, 1998, pp. 8-14. https://doi.org/10.3141/1628-02

Ivan, K. and Haidu, I. (2012). The Spatio-Temporal Distribution of road accidents in Cluj Napoca. Geographia Technica, 2, pp.32-38.

Johnson, Derick A., and Mohan M. Trivedi. Driving style recognition using a smartphone as a sensor platform. In Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on, 2011 pp. 1609-1615. https://doi.org/10.1109/ITSC.2011.6083078

Lenné, Michael G., Thomas J. Triggs, and Jennifer R. Redman. Time of day variations in driving performance. Accident Analysis & Prevention Vol. 29, No. 4, 1997, pp. 431-437. https://doi.org/10.1016/S0001-4575(97)00022-5

Matthews, Gerald, Lisa Dorn, Thomas W. Hoyes, D. Roy Davies, A. Ian Glendon, and Ray G. Taylor. Driver stress and performance on a driving simulator. Human Factors Vol. 40, No. 1, 1998, pp. 136-149. https://doi.org/10.1518/001872098779480569

Noland, R. and Quddus, M. (2004). A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention, 36(6), pp.973-984.

OSeven, 2019, (www.oseven.io).

Regan, M. A., Williamson, A., Grzebieta, R., & Tao, L. (2012, August). Naturalistic driving studies: literature review and planning for the Australian naturalistic driving study. In Australasian college of road safety conference 2012, Sydney, New South Wales, Australia.

Sabel, C., Kingham, S., Nicholson, A. and Bartie, P. (2005). Road Traffic Accident Simulation Modelling - A Kernel Estimation Approach

Toledo, Tomer, Oren Musicant, and Tsippy Lotan. In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transportation Research Part C: Emerging Technologies Vol. 16, No. 3, 2008, pp. 320-331. https://doi.org/10.1016/j.trc.2008.01.001

Tselentis, Dimitrios I., George Yannis, and Eleni I. Vlahogianni. Innovative motor insurance schemes: a review of current practices and emerging challenges. Accident Analysis & Prevention Vol. 98, 2017, pp. 139-148. https://doi.org/10.1016/j.aap.2016.10.006

Vlahogianni, Eleni I., and Emmanouil N. Barmpounakis. Driving analytics using smartphones: Algorithms, comparisons and challenges. Transportation Research Part C: Emerging Technologies Vol. 79, 2017, pp. 196-206. https://doi.org/10.1016/j.trc.2017.03.014

WHO: Global status report on road safety by World Health Organization, 2018 (accessed 19 December 2018) https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.

Zaldivar, Jorge, Carlos T. Calafate, Juan Carlos Cano, and Pietro Manzoni. Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. In Local Computer Networks (LCN), 2011 IEEE 36th Conference on, pp. 813-819. IEEE, 2011. https://doi.org/10.1109/LCN.2011.6115556