CAPTURING THE EFFECTS OF TEXTING ON YOUNG DRIVERS BEHAVIOUR BASED ON COPULA AND GAUSSSIAN MIXTURE MODELS

Loukas Dimitriou, Ph.D.*
Lab. Of Transportation Engineering
Department of Civil and Environmental Engineering,
University of Cyprus,
75 Kallipoleos Str., P.O. Box 20537, 1678 Nicosia, Cyprus
Tel: +35722892286
E-mail: lucdimit@ucy.ac.cy

Katerina Stylianou, MSc
Lab. Of Transportation Engineering
Department of Civil and Environmental Engineering,
University of Cyprus,
75 Kallipoleos Str., P.O. Box 20537, 1678 Nicosia, Cyprus
Tel: +35722892286
Email: stylianou.p.katerina@ucy.ac.cy

George Yannis, Ph.D.
Department of Transportation Planning and Engineering
School of Civil Engineering,
National Technical University of Athens,
5, Heroon Polytechniou Str., 15773, Zographou Campus, Greece
Tel: +302107721326
E-mail: geyannis@central.ntua.gr

Abstract

This research effort aims to investigate the impact of texting on young drivers' behavior and safety based on data from driving simulator experiments, for different driving contexts, like motorways, urban and rural roads, during daytime and night, and for alternative weather conditions (‘clear sky’ and rain), offering a complete and comprehensive investigation of the effects of texting on driving behavior, able to provide evidence on policy-making. On that purpose, a driving simulator experiment was carried out in which 34 young participants drove in the above-described different driving scenarios. Initially, multivariate copula analysis was used in order to explore statistical inferences among variables, especially since it retains a parametric specification for bivariate dependencies and allows testing of several parametric structures to characterize them. Alternative copulae configurations have been tested, which showed that texting and other road and environmental characteristics affect young drivers behavior and in particular more than one outcome can occur at the same time. As a second step, Gaussian Mixture Modeling (GMM) was employed, demonstrating that the variables' pairs that presented the strongest correlations were lane excursion and speed, as well as speed and reaction time. GMMs application showed that drivers using mobile phones who were involved in a collision presented a different driving behavior compared to the drivers who were occupied but were not involved in a collision.

Keywords: Young Drivers; Driving Behavior; Texting; Driving Simulator; Multivariate Copula Analysis; Gaussian Mixture Models.

1. Introduction

In today’s society where mobile phones are one of the dominant means of communication and information, the use of mobile phones and especially texting (a distracting activity) has been acknowledged as a major concern for traffic safety. A recent study has reported that 70% of young drivers surveyed, initiate texts while driving, while 81% reply to texts, and even higher numbers (92%) read messages while driving (Atchley et al.,2005). Driving and the use of mobile phones have long been an area of interest in the field of road safety, as an acknowledged hazardous activity. The distraction posed by phones is only becoming worse as smartphone technology evolves and the range of available features expands. Social media are now another popular means of communication with social networking and media platforms amplifying the distraction caused by phones. Furthermore, with the development of application features for smartphones, for various social media platforms (e.g. Facebook, Twitter Pinterest, Instagram etc.), users are sent push notifications at any time. One of the most popular distraction features of cell phones though, is the capability to text message. Even though that this behavior is generally perceived as hazardous, surveys always indicate high numbers of drivers admitting that they read or text message while driving.

According to Charlene et al., 2012 in a Nationwide online survey in New Zealand, it was reported that over half of the responders send or read between 1 and 5 text messages while driving in a typical week, even though an 89% agreed that texting and driving generally impairs driving performance. However, Metz et al., 2014 while investigating the frequency of secondary tasks in driving in a naturalistic study, observed that especially demanding visual manual secondary tasks, including the handling of a mobile phone, occurred more often in a standstill situation of the vehicle.

To identify the impact of texting on driver behavior is a difficult task as the use of mobile phones is rarely reported. However, texting while driving is considered as a distracting activity which increases crash and near crash risk. A research developed by the Texas Transportation Institute, using a naturalistic driving approach, indicated that when reading or writing texts, drivers exhibited reductions in reaction time almost double that what was previously thought and it was also shown that nearly identical impairment in the reading and writing conditions occurs, thus suggesting that both these actions may be equally dangerous (Cooper et al., 2011). In accordance to Drews et al., 2009, in a driving simulator experiment, also indicated that participants responded more slowly to the onset of braking lights when engaged in text messaging compared with a free driving condition. In the 100-car naturalistic driving study it was reported that hand held wireless devices were associated with the highest frequency of distraction-related events for both incidents and near-crashes (Dingus et al., 2006).

Research has also shown that drivers engaging in visual and or manually complex tasks present a three times higher near-crash, crash risk than drivers who are attentive, and also drivers who glance away for a period higher than 2 seconds, double their crash risk (Klauer et al., 2006). Additionally, lane excursions are identified to be a common outcome related with texting and driving through the literature (Alosco et al., 2012, Crandall and Chaparro, 2012, McKeever et al., 2013). An experiment using an advance driving simulator in Australia, revealed that drivers' ability to maintain lateral position on the road and also to detect and respond to traffic lights was significantly reduced while texting (Hosking et al., 2009). In He et al., 2015, using a lane change task and smartphone technology, mutual interference of texting and driving was studied. Similarly, another driver simulator experiment reported a 66% of subjects overall exhibiting lane excursions while texting (Rumschlag, 2015).

Increased accident risk is also a common outcome when studying the effects of texting. In a naturalistic driving setting which lasted over a period of 1 year Farmer et al., 2014, reported that the risk of a near crash/crash event was approximately 17% higher when drivers were interacting with a mobile phone, due to action of reaching for/answering/dialing, which increased risk by three times. In accordance to the results of this study, Olson et al., 2009 reported that drivers were 23.2 times more likely to be involved in a safety-critical event while text messaging. Similarly, Drews et al., 2009 reported that an 86% of the collisions presented in the simulated environment were caused while participants were text messaging while operating the vehicle.

The above described results indicate that texting while driving has an important impact on driving behavior and overall road safety. These safety related concerns indicate the need to fully comprehend how texting and the use of mobile phones in general impact driving through continuous research and to further seek measures to eliminate such behaviors.

The influence of texting on driving behavior has been very well studied and provided consistent results regarding behavior while engaged in the secondary task. However, literature has focused on the individual outcomes texting imposes on driving behavior i.e. reaction time is reduced or the vehicle is off tracked while texting. A rising issue is whether these effects of texting occur simultaneously, and if yes which effects present the highest correlation. Within this context, the current study aims to collect and analyze information on driver behavior when faced with the secondary task of texting on a representative sample of young drivers by means of a driving simulator experiment and particularly, to examine the simultaneous effects of texting and driving while examining different road and traffic conditions (urban-rural environment, normal-increased traffic conditions), and environmental conditions (good weather, rainy weather, night time).

2. Methodology

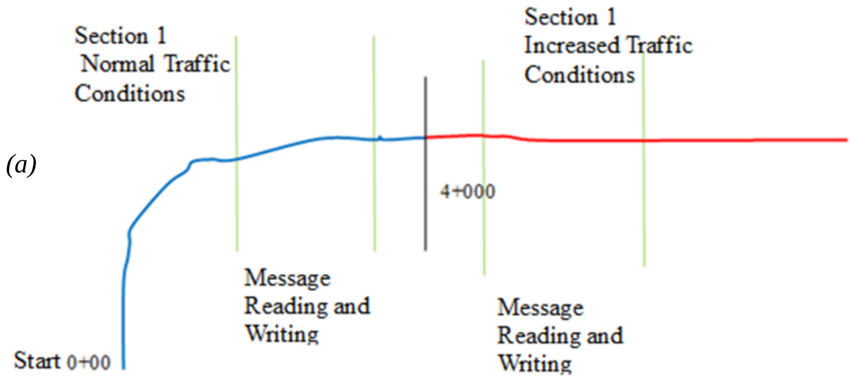
The impact of texting on driving was examined through a simulator experiment (Yannis et al., 2014, Christoforou, 2012, Gartzonikas, 2012). In this experiment 34 participants aged between 18 and 28 years of age, took part, with an average driving experience of 3.5 years. The sample of drivers consisted of 19 males and 15 females, and in order to be familiar with the device, each participant used their own mobile phone during the experiment. The discretization between touch screen and keyboard devices was 60% and 40% respectively. The participants of the experiment were first asked to complete a questionnaire about their personal driving characteristics with regards to texting. In general the data from the study comply with the literature, with 47% of the 34 participants stating that they often read or text message while driving.

The second part of the experiment constituted of the driver simulator experiment for the data collection. The participants first, had a practice drive in a random route for approximately 5 minutes, in order to get familiar with the simulator. Subsequently participants drove the same route for the same time period, for three times, where weather conditions or time of day were altered. Each drive, one in good weather, one in rainy weather and one at night (performed only in Scenario B) lasted approximately 5 minutes. Figure 1 presents the driving scenarios used in the current research. Scenario (A) included driving on a circular route comprising of two motorway sections where the first section comprised of normal traffic conditions whereas in the second section increased traffic conditions occurred. The speed limit for the motorway is 100 km/h. Scenario (B) included driving on a circular route as well, comprising of two rural sections separated by an urban section under moderate traffic. Speed limits for urban and rural section(s) are 50 and 70 km/h, respectively.

At predefined locations of the journey for both scenarios investigated, the surveyor sent and received text messages to and from the driver. In Scenario (A) while driving in normal traffic conditions the driver received a text message asking to provide directions to the University Campus. In increased traffic conditions the drivers were asked to reply to a text message asking for a cooking recipe. In Scenario (B) while driving in the rural sections, the drivers were asked to read a message thanking them for participating in an experiment and to write the first two lines of the national anthem in a text message. In the urban section the drivers were asked to reply to a message comprising of an approximately 30 character question. Furthermore, during the experiment incidents were scheduled to occur at various points of the selected route, for example the appearance of an animal or a pedestrian in the road. There was one incident scheduled for each of the drivers' actions in the different road sections, in each drive.

In particular, in Scenario B, for each drive i.e. good weather; rain; nighttime, one incident occurred in section 1 ("Rural") while the driver was reading a message, one incident occurred in the second section ("Urban") while reading the text message and one while composing it, and finally one incident occurred in the second Rural section while the participant was composing the text message. The incidents were not scheduled to occur at specific locations in order to avoid learning effects among the different drives. As presented in Figure 1 each Scenario is divided in distinct sections; free driving, message reading and writing and message reading or message writing. The database was composed by calculating average values of the selected variables in each of these sections, in the selected Scenario and environmental conditions, resulting in over 400 observations for each of the two Scenarios. The size of this database is adequate for examining the effect of texting on driving behavior.

A two point ordinal scale was used in the database to represent whether a collision occurred in each of the different drives: (0) no collision; (1) collision. Collisions would include crashes with other vehicles, pedestrians, animals or infrastructure. After a collision occurred the drive would stop for a few seconds and then the participant was able to continue from the same spot until the end of the drive.



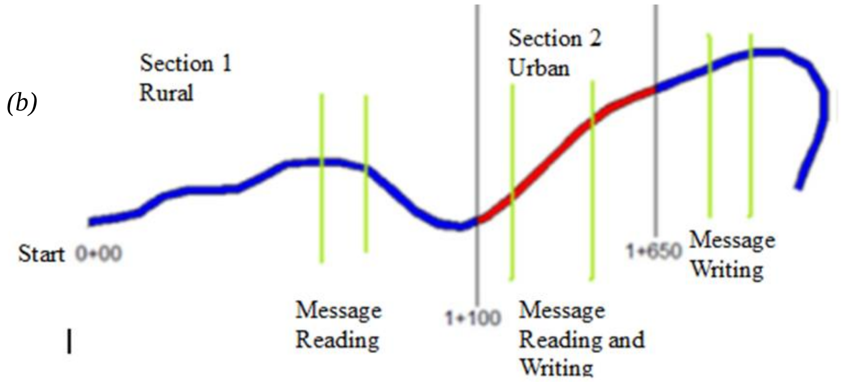


Figure 1. Driving Scenarios

The number of collisions in each Scenario and different conditions is presented in Table 1. In the full set of the data, the highest number of collisions was presented in the urban environment. Also, a higher number of collisions, was presented during the rainy weather conditions rather than the good weather. What is noteworthy of the results is the fact that 86% of the collisions occurred while the drivers were occupied by their phones. Table 1 also presents the use of a mobile phone with a touch screen. In both Scenarios it is clear that drivers using a mobile phone with a touch screen were involved in more collisions than those using a mobile phone with a keyboard.

Table 1. Number of no collisions and collisions in normal and increased traffic conditions and urban and rural environments under different environmental conditions

		Full Set	Good weather	Rainy weather	Nighttime
Scenario A					
Normal Traffic Conditions	No Collision	179	96	83	N/A
	Collision	21(19)(15)[1]	5(5)	16(14)	N/A
Increased Traffic Conditions	No Collision	168	93	78	N/A
	Collision	14(13)(8)	4(4)	10(9)	N/A
Scenario B					
Urban Environment	No Collision	181	61	56	64
	Collision	95(86)(55)	21(14)	40(16)	29(27)
Rural Environment	No Collision	206	82	59	65
	Collision	63(48)-30	10(10)	32(27)	21(14)

Variables for the study were extracted from the simulator's data recordings separately for each of the three circular routes (of each scenario) or from the survey questionnaire. This study sought to identify the variables, which are associated with driving behavior. Table 2 provides a description of the available parameters for modelling on the weather conditions (rainy; good), time of day (night time) traffic conditions (travel speed), geometry conditions (lateral position of the vehicle in the road; distance from left and right border and from the middle of the road), relationship with the preceding vehicle (time to headway), driver reaction to incidents (reaction time), type of mobile phone use (with touch screen or not) as well as driver actions (driving free; reading or composing a message).

Table 2: Variables' set for monitoring driver behavior

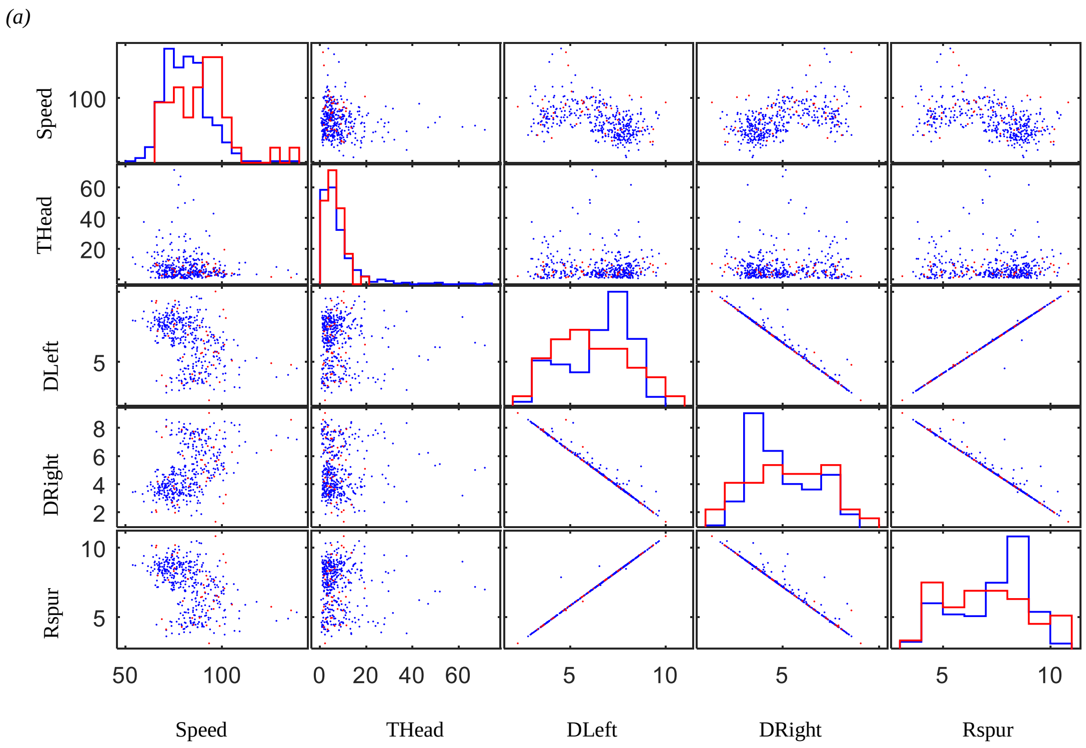
Rainy	Rainy Weather (1: yes, 0:no)
Good	Good weather conditions (1:yes,0:no)
Night	Driving During Night (1:yes, 0:no)
Speed	Vehicle Speed (km/h)
DLeft	Distance from the left road border (m)
DRight	Distance from the right road border (m)

---

[1] Total number of collisions (using a phone)(touchscreen)

Rspur	Track of the vehicle from the middle of the road (m)
THead	Time to Headway i.e. to collision with the ahead driving vehicle (s)
RT	Reaction Time (s)
Touch	Mobile phone with a touch screen (1:yes,0:no)
In_Free	Free driving in Urban environment (1:yes,0:no)
In_Read	Reading message in Urban environment (1:yes, 0:no)
In_Write	Composing message in Urban environment (1:yes, 0:no)
Out_Free	Free driving in Rural environment (1:yes,0:no)
Out_Read	Reading message in Rural environment (1:yes, 0:no)
Out_Write	Composing message in Rural environment (1:yes, 0:no)
Free_Q1	Free driving in Normal Traffic conditions (1:yes,0:no)
Read_Q1	Reading message in Normal Traffic conditions (1:yes, 0:no)
Write_Q1	Composing message in Normal Traffic conditions (1:yes, 0:no)
Free_Q2	Free driving in Increased Traffic Conditions (1:yes,0:no)
Read_Q2	Reading message in Increased Traffic Conditions (1:yes, 0:no)
Write_Q2	Composing message in Increased Traffic Conditions (1:yes, 0:no)

The scope of the current study focuses on monitoring and analyzing young drivers' behavior while texting when exposed in different driving conditions and while performing various tasks. Figure 2 presents correlation diagrams of 5 important variables, displaying all the bivariate scatterplots between the variables along with a univariate histogram for each variable. Depicted in red are the observations where a collision occurred throughout the experiment, whereas blue indicates the cases where no collision occurred. As can be seen from Figure 2 the actual data itself when plotted, does not give distinct information on driver behavior. In order to examine the driving patterns of the participants and to obtain insight on the correlations of different driving styles a copula analysis was employed. Then, data clustering using Gaussian Mixture Modeling (GMM) was utilized, in order to investigate stochastic nonlinear driving patterns. This two-step analysis presented in the current research, results in quantitative and qualitative information that can in later stage be used in simulation.



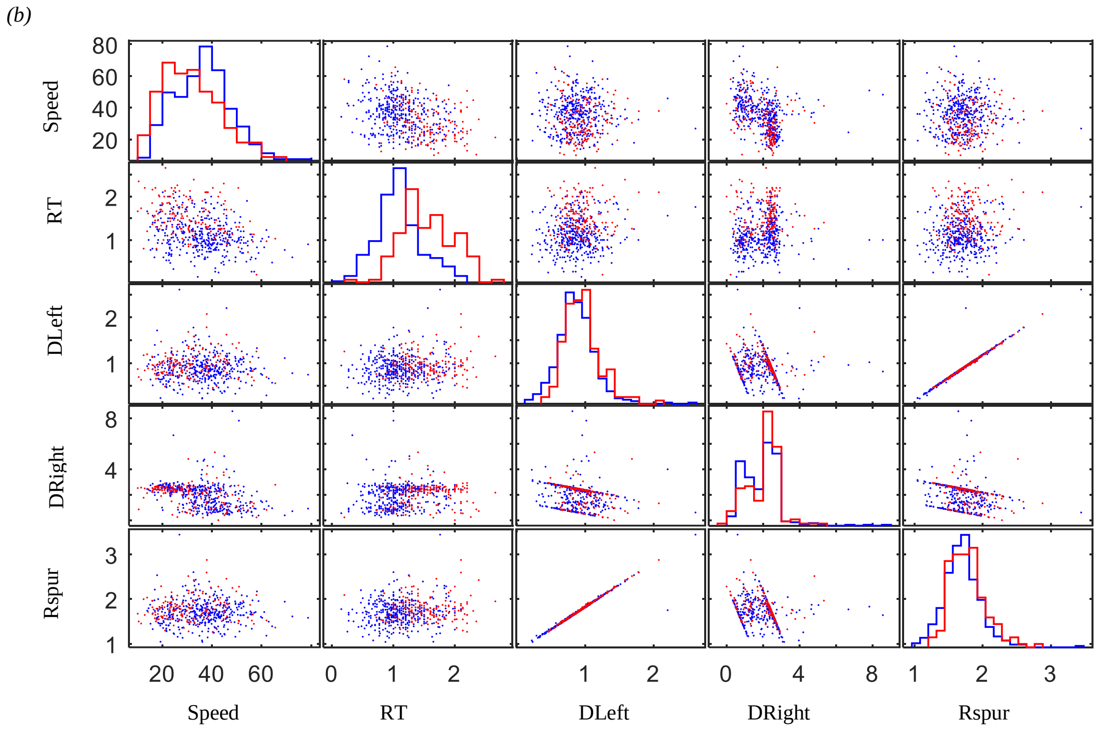


Figure 2: Correlation matrices for the two driving scenarios

2.1. *Copula Analysis*

The concept of the copula approach has been proven to be resourceful in economics and especially risk management. Copula function, as a solid tool to model dependence, provides understanding in the interaction of individual driver behavior components thus, providing inside to the nature of the entire traffic flow. The main purpose of copula is to describe the interrelation of several random variables (Schmidt, 2007).Copula theory and application is based on Sklar's theorem. From Schmidt, 2007 for a d-dimensional cumulative distribution function F with marginal F1,….,Fd, there exist a copula C, such that:

$$F(x_1, \ldots., x_d) = C(F_1(x_1), \ldots. F_d(x_d))$$

For all x_i *in* $[-\infty,\infty]$*,* $i = 1,\ldots.,d$*.*

If $F_1,\ldots.,F_d$ are continuous, then C is unique. That is, we can describe the joint distribution of $x_1,\ldots.,x_d$ by the marginal distributions $F_1,\ldots.,F_d$ and the copula C. Thus, from a modeling perspective, it appears that the information about the dependence is contained in the copula function, irrespectively of the variable's distributions, providing a natural framework for various investigations (Darsow et al., 1992). In essence, a copula is a device which reveals stochastic dependence relationships among variables with pre-specified marginal distributions (Bhat and Eluru, 2009), and as such provides an appropriate analytical tool. In this study a t copula was employed. From Demarta and McNeil, 2005, the d-dimensional random vector $X = (X_1,\ldots.,X_d)$ is said to have a multivariate t distribution with ν degrees of freedom, mean vector μ and positive-definite dispersion or scatter matrix Σ, denoted $X \sim t_d(\nu, \mu, \Sigma)$, under certain assumptions. The copula remains invariant under a standardization of the marginal distributions meaning that the copula of td (ν,μ,Σ) is identical to the copula of t_d (ν,d,P), with P denoting the correlation matrix implied by the dispersion matrix Σ (Demarta and McNeil, 2005).

In order to transform the data to the copula scale, the Kernel estimator of the cumulative distribution function was used. Figure 3 presents an example of the data as it is in its original form and how it transforms by using the Kernel smoothing factor estimate.

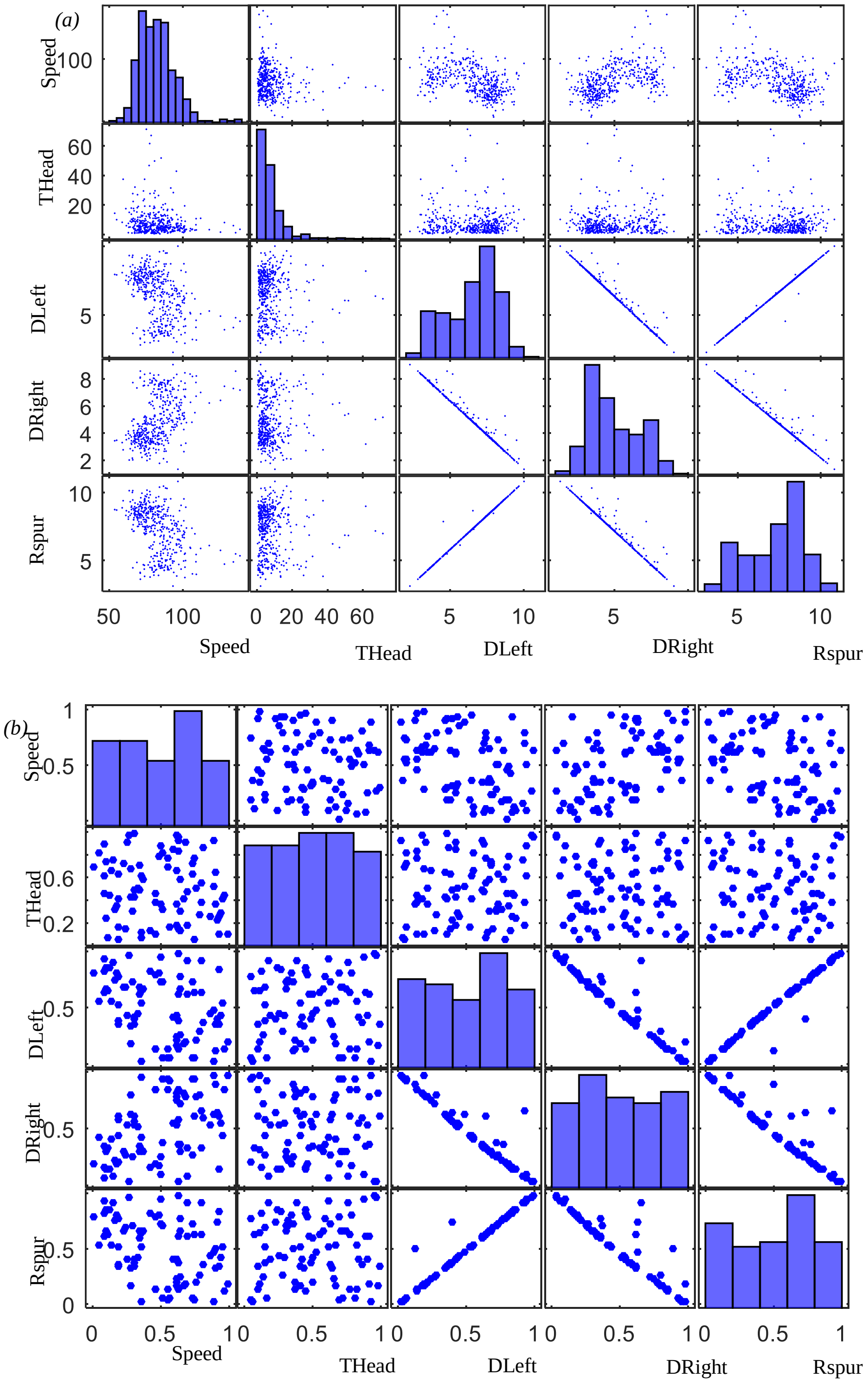


Figure 3: (a) Observations in normal traffic conditions without being involved in a collision; (b) Normalized observations in normal traffic conditions without being involved in a collision

2.2. Gaussian Mixture Models

Previous research in safety forecasting and crash frequency data analysis focused on describing the data using conventional probability density function (pdf), such as normal, exponential, gamma etc. (Lord and Mannering, 2010). However, if crash data follow a multinomial (e.g.

bimodal) distribution or mixed distributions there is no specific distribution available (Jin et al., 2013). In order to estimate pdfs of mixed distributions, mixture models are useful tools.

A Gaussian Mixture Model (GMM), which is the most well-defined type of the mixture models, is a parametric probability density function, represented by the weighted sum of M component Gaussian density, as given by Reynolds, 2009. In this study a clustering technique through GMMs was used to define relationships in the driving behavior of drivers when engaging in a secondary task. Cluster analysis is the autonomous identification of groups of cohesive data in a dataset (Fraley and Raftery 2002), following the assumption that d data generated by mixtures of multivariate normal densities are characterized by clusters centered at the means μi with increased density for points nearer the mean (Fraley and Raftery, 2002).The multivariate normal pdf of the GMM, can be presented using contour plots, displaying the isolines of constant probability density and the probability density, where for a bivariate normal distribution the contour Q=0 consist of one single point (μ1,μ2) where the density is maximum (Kalbfeisch, 1979).

3. Results

The impact of texting on young driver behavior in (Scenario A) Normal and Increased traffic conditions and (Scenario B) Urban and Rural environments, was analyzed first, through the employment of multivariate copula analysis and later through Gaussian Mixture Modelling. In particular, the multivariate copula analysis was performed for 22 categories of different traffic conditions and driver actions for the set of variables presented in the previous section. For each category, a t-copula was performed presenting the correlation matrix of the variables. The copula analysis examined the behavior of the drivers who were involved in a collision and subsequently the drivers who were not and identified the variables with the strongest correlations. Taking the results of the copula analysis, clustering through the use of Gaussian Mixture Models was then performed to present the relationships in the behavior of drivers while texting.

3.1. Copula Analysis

A separate t-copula was performed for each of the different categories of driving and weather conditions, which resulted in a correlation matrix of the examined variables. Through the correlation matrices it was easy to identify the significant variable pairs that would describe driver's behavior; the variable pairs with the strongest correlations for all of the categories were picked out and are presented below in Table 3.

In order to get an insight of the driving patterns of drivers when exposed in the same conditions, Table 3 presents the copula results which exhibited the strongest linear correlations in all the categories examined. As can be seen, the results present a variation of values of the linear coefficient ρ, ranging from $|0.75|$ indicating a strong correlation, to $|0.06|$ indicating no correlation at all. The wide range of ρ presented in the copula analysis presents the stochasticity and complexity of this type of data, which is associated with human behavior. Varying pairs of variables exhibited the strongest correlations for the multiple categories examined, implying that the operational patterns of drivers are not consistent and cannot be described in a simplistic manner.

It was shown that the variable pairs that presented the strongest correlations in the examined traffic conditions were the pairs of lane excursion and speed i.e. DLeft – Speed and DRight – Speed. This is not a surprising result, as literature has shown that drivers who text while driving present lower speeds and frequent lane excursions (Rumschlag et al, 2015). Further to this, driver's reaction time with speed (Speed – RT), also presented a high frequency in the results.

As presented in Table 1 the number of collisions in Scenario B was higher rather than Scenario A. In Scenario B it was shown that the drivers who were not involved in a collision while driving in a free state presented a high frequency of correlation between speed and reaction time. However, in the same scenario when examining the state where the driver was composing a text message, the strongest correlations were presented in speed and lane excursions.

Particularly, for the drivers who were not involved in a collision while composing a message, the results showed a higher frequency of correlation between speed and distance from the left border of the road (Speed – Dleft). The negative correlation between these two variables indicates the simultaneous increase (or decrease) between the driver's speed and decrease (or increase) of distance from the left border of the road. The highest correlation of the pair Speed – Dleft for this case (not involved in a collision while texting), was presented in the rural environment in rainy conditions (ρ = -0.57), indicating that in this category the drivers exhibited the most changes in their driving behavior regarding speed and lane tracking.

On the other hand, in the cases where drivers were involved in a collision while composing a text message, the most frequent correlation was found to be between speed and the distance from the right border of the road. Once again the negative correlation between speed and distance from the border implies the increase or decrease in speed in the same time of the decrease or increase of the distance respectively. The highest correlation in this case, again, was shown in the Rural environment while it was raining (ρ = -0.56). In the cases where the driver was reading a message different results were presented. In the case where the driver was not involved in a collision, the pairs speed-DLeft and speed-RT were equally presented.

Contrary to the case where the driver was composing a text message, in the case of reading a text message, speed and distance from the left border was presented in less of the categories examined, implying that composing a text message has a bigger effect on lane tracking and speed maintenance, than reading. For the case where drivers were involved in a collision, the variable pair that presented the highest correlations was speed and reaction time. The highest correlation of the two variables was presented in the rural environment while it was raining. In this case, ρ was equal to 0.64, whereas for the same conditions, while the driver was driving in a free state ρ was 0.35. This result shows that while reading a message the driver presents more variations in speed maintenance and reaction time rather when driving freely. When the drivers were driving during the night, it was shown that both for the set of drivers who were involved in a collision and the set that was not, the highest correlation was presented between speed and reaction time in both urban and rural environments. However, in the cases of collisions the linear correlation between the two variables was higher (ρ = -0.26 for both urban and rural environment).

When examining the impact of a touch screen, in the normal traffic conditions, the drivers who were occupied with their mobile phone (either composing a text or reading a text) presented the highest correlation in the variable pair of speed and distance from the left border (ρ=-0.48). Further to this, the negative correlation presented shows that when speed is increased, the distance from the left border is decreased and vice versa. In the urban environment the drivers who were occupied by a phone with a keyboard indicated a highest correlation in the variable pair of speed and distance from the left border, whereas the drivers using a phone with a touchscreen indicated a highest correlation in the pair of speed and distance from the right border. In the rural environment both the cases either with a phone with a keyboard or a touchscreen indicated a highest correlation between reaction time and distance from the right border, however in the case of the touchscreen phone the correlation was negative (ρ = -0.58).

Table 3: Results of the Copula analysis

			Driver Actions							
			Full Set		Free Driving Subset		Reading Text Subset		Composing Text Subset	
Scenario	Drive Description	Device	Selected Copula	ρ	Selected Copula	ρ	Selected Copula	ρ	Selected Copula	ρ
No Collision Subset										
Normal Conditions	Full Set		Speed/ DRight	0.50	Speed/ Dright	0.41	Speed/ DRight	0.39	Speed/ DLeft	-0.59
	Good Weather		Speed/ DRight	0.56	Speed/ DRight	0.53	Speed/ DLeft	-0.54	Speed/ DRight	0.63
	Rainy		Speed/	0.39	Speed/	0.30	Speed/	0.18	Speed/	-0.61

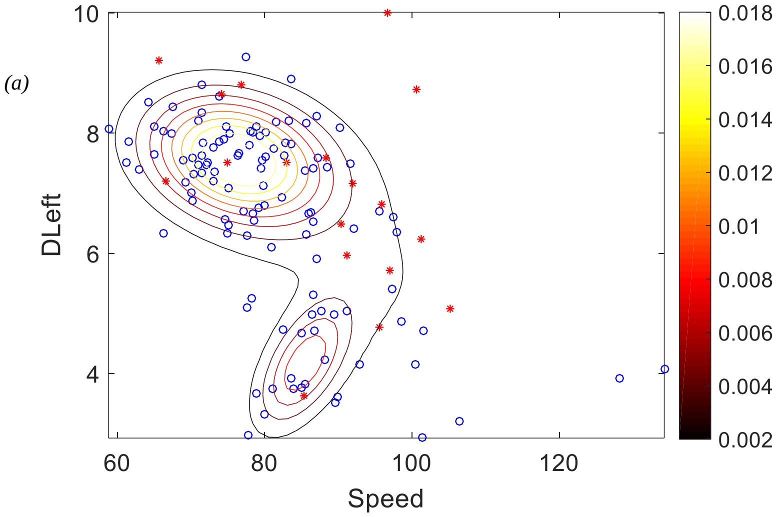
	Weather		DRight		DRight		DRight		DLeft	
Increased Conditions	Full Set		Speed/ DRight	0.47	Speed/ DRight	0.43	Speed/ DLeft	-0.35	Speed/ DRight	-0.45
	Good Weather		Speed/ DLeft	-0.53	Speed/ DRight	0.50	Speed/ DLeft	-0.43	Speed/ DRight	0.47
	Rainy Weather		Speed/ DRight	0.38	Speed/ DRight	0.36	Speed/ DRight	0.18	Speed/ DLeft	-0.40
Urban Env.	Full Set		Speed/ RT	-0.17	Speed/ RT	-0.16	Speed/ DLeft	-0.17	Speed/ DLeft	-0.14
	Good Weather		Speed/ RT	-0.21	RT/ DRight	-0.33	Speed/ DLeft	-0.26	RT/ DRight	-0.55
	Rainy Weather		RT/ rspur	0.27	RT/ DLeft	0.48	RT/ DLeft	0.35	Speed/ DRight	-0.60
	Nighttime		Speed/ RT	-0.23	RT/ rspur	-0.20	RT/ DRight	-0.10	Speed/ DLeft	-0.42
Rural Env.	Full Set		Speed/ DRight	-0.24	Speed/ RT	-0.16	Speed/ RT	0.30	Speed/ DLeft	-0.42
	Good Weather		Speed/ DRight	-0.36	Speed/ RT	-0.13	Speed/ RT	0.67	Speed/ DLeft	-0.34
	Rainy Weather		Speed/ DRight	-0.31	Speed/ RT	-0.35	Speed/ rspur	0.48	Speed/ DLeft	-0.57
	Nighttime		Speed/ RT	-0.26	Speed/ RT	-0.24	RT/ DRight	0.37	RT/ DRight	0.34
Collision Subset										
Normal Conditions	Full Set		Speed/ DLeft	-0.43	--	--	Speed/ DLeft	-0.60	Speed/ DLeft	-0.52
	Good Weather		--	--	--	--	--	--	--	--
	Rainy Weather		Speed/ DLeft	-0.43	--	--	Speed/ DLeft	-0.46	--	--
Increased Conditions	Full Set		Speed/ THead	-0.16	--	--	Speed/ DRight	-0.52	--	--
	Good Weather		--	--	--	--	--	--	--	--
	Rainy Weather		Speed/ THead	0.22	--	--	Speed/ DLeft	-0.73	--	--
Urban Env.	Full Set		Speed/ RT	-0.19	Speed/ DLeft	-0.40	Speed/ DLeft	-0.38	Speed/ DRight	-0.09
	Good Weather		Speed/ DLeft	-0.43	--	--	Speed/ DRight	0.58	Speed/ DRight	0.40
	Rainy Weather		RT/ DLeft	0.18	--	--	Speed/ rspur	-0.54	RT/ DLeft	0.19
	Nighttime		Speed/ RT	-0.26	--	--	RT/ DRight	0.65	Speed/ DRight	-0.42
		Touch	Speed/ DRight	0.28	N/A	N/A	RT/ DRight	-0.25	Speed/ RT	-0.20
		KB	Speed/ DLeft	-0.75	N/A	N/A	RT/ rspur	-0.19	RT/ DRight	0.30
Rural Env.	Full Set		Speed/ RT	-0.48	Speed/ RT	-0.65	Speed/ RT	-0.33	RT/ DLeft	0.32
	Good Weather		RT/ DRight	0.48	--	--	--	--	Speed/ DRight	-0.4
	Rainy Weather		Speed/ RT	-0.46	Speed/ RT	-0.74	Speed/ RT	-0.64	Speed/ DRight	-0.56
	Nighttime		Speed/ RT	-0.64	Speed/ RT	-0.71	RT/ DRight	0.50	RT/ DLeft	0.48
		Touch	RT/	-0.58	N/A	N/A	RT/	0.56	Speed/	-0.51

	KB	DRight RT/ DRight	0.58	N/A	N/A	DLeft RT/ DRight	-0.53	RT Speed/ RT	-0.45

3.2. Gaussian Mixture Models

The copula analysis results implied that the analysis of operational patterns of drivers when engaged in texting is a complex and unclear task. To have a better understanding of the data after the linear correlation analysis, a stochastic mixture of models is employed. The GMMs were performed for the same categories as described in the copula analysis. The modeling was based on the variable pairs that exhibited the highest correlation through the copula analysis, of the subset of drivers who were involved in a collision while using their mobile phone during the simulator experiment. GMMs were performed for $\rho \geq |0.30|$ for all the road conditions (normal-increased traffic conditions and urban-rural environment) and for the environmental conditions of rainy weather and nighttime. In the following figures, of the subset of drivers who were involved in collisions are depicted in red asterisks and the subset of drivers who were not involved in collisions with blue circles.

Figure 4 presents speed vs. distance from the left border for normal traffic conditions. When modeling all the observations together it is shown that observations of collisions overlap with the observations of no collisions. Two peaks are shown as the data seem to form two clusters, in which both observations of collisions and no collisions fall into. The main cluster exhibiting the highest probability density function (pdf) of around 0.018 mainly consists of observations of no collision while occupied by mobile phone use. When examining the two cases separately (b) when no collision occurred while texting and (c) a collision occurred while texting, different results are presented. In Fig. (4b) two main clusters are formed. The main cluster presented a distance from the left border between 6m and 8m while the speed was between 60km/h and 90km/h. The second cluster in Fig. (4b) presents a distance from the left border between 2m and 5m and a speed between 80km/h and 110km/h. These two clusters clearly represent the two lanes of the motorway. What the results show is that even though the drivers were using their mobile phone, they did not have great variations in speed and lane maintenance and as such avoided a collision. On the other hand, Fig. (4c) presents the cases where collisions occurred. The results indicate various lane excursions and especially in the cluster of higher speeds. The highest pdf in the case of a collision while using a phone in normal traffic conditions, is between 6m and 7m distance from the left border and a speed between 90km/h and 95km/h, or between 8m and 8.5m with a speed of 70km/h to 75km/h.



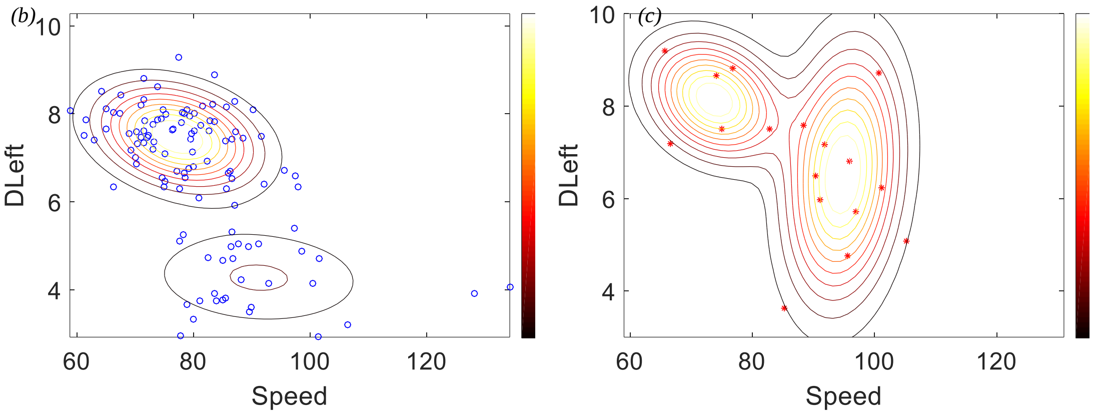


Figure 4: Normal Conditions (a) All observations; (b) No collisions; (c) Collisions

Figure 5 presents speed vs. DLeft for driving in normal traffic conditions while raining. When modeling all the observations three main clusters are formed in which both observations of collision and no collision fall into. In the majority of the cases were drivers were not involved in a collision Fig. (5b), they kept their distance from the left border of the road between 6m and 9m whereas the speed varied between 70km/h and 95km/h. The distance from the border and the respective speeds indicate that drivers who were using their phones but were not involved in collisions drove on both lanes of the motorway. However, in the cases where collisions occurred, the pdf was highest at 95km/h and 5m away from the left border of the road.

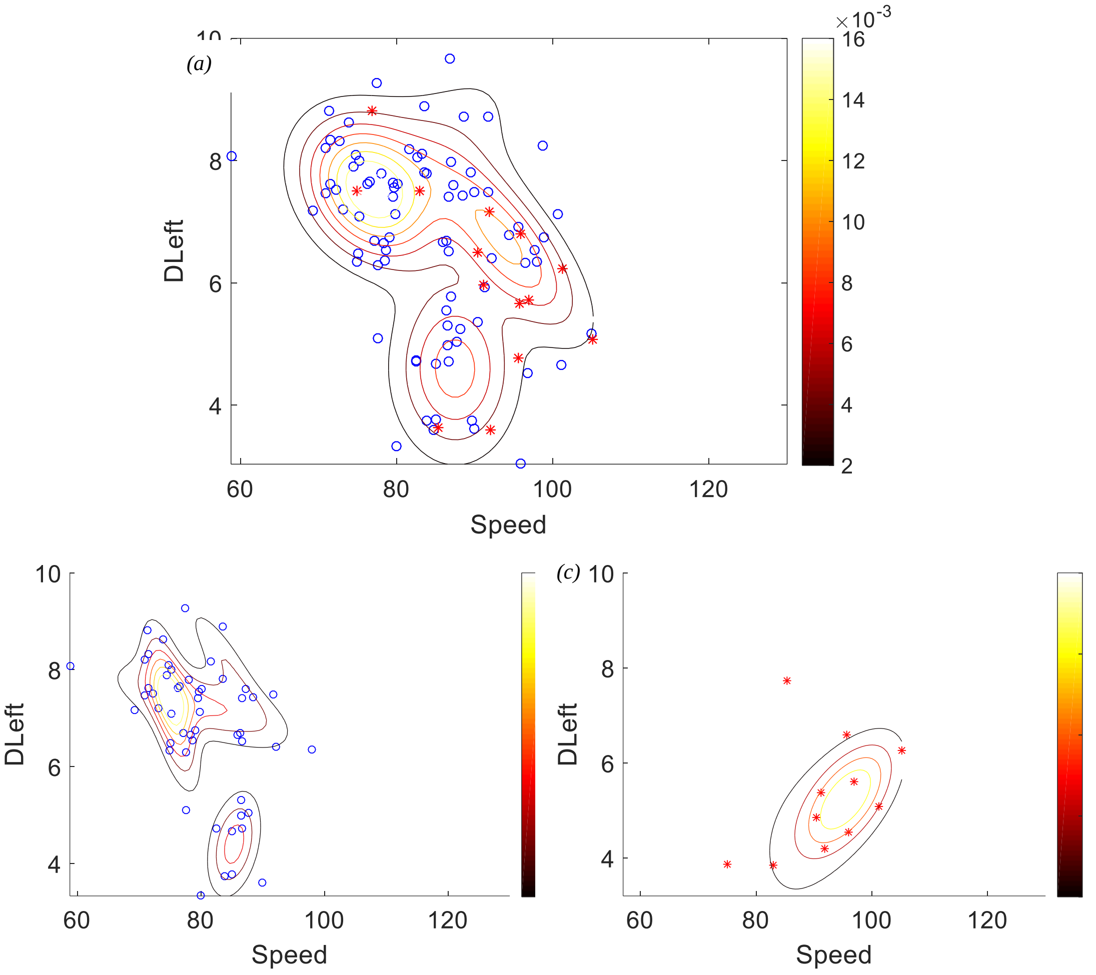


Figure 5: Normal Conditions and Rainy weather (a) All observations; (b) No collisions; (c) Collisions

Figure 6 presents the speed vs. reaction time in the rural environment. In this case Fig.(6a) it is clear that the observations of the collisions overlap with the observations of no collision. The main cluster presented in Fig(6b) shows that reaction time of the drivers who were using their phones but were not involved in a collision was around 1 second, while their speed varied between 30km/h and 40km/h, which is below the speed limit for the rural environment. On the other hand, in Fig. (6c) shows a higher variation in speed. The peak in the model shows that the reaction time of drivers who had a collision was around 1.2s to 1.5s but their speed varied between 25km/h and 50km/h.

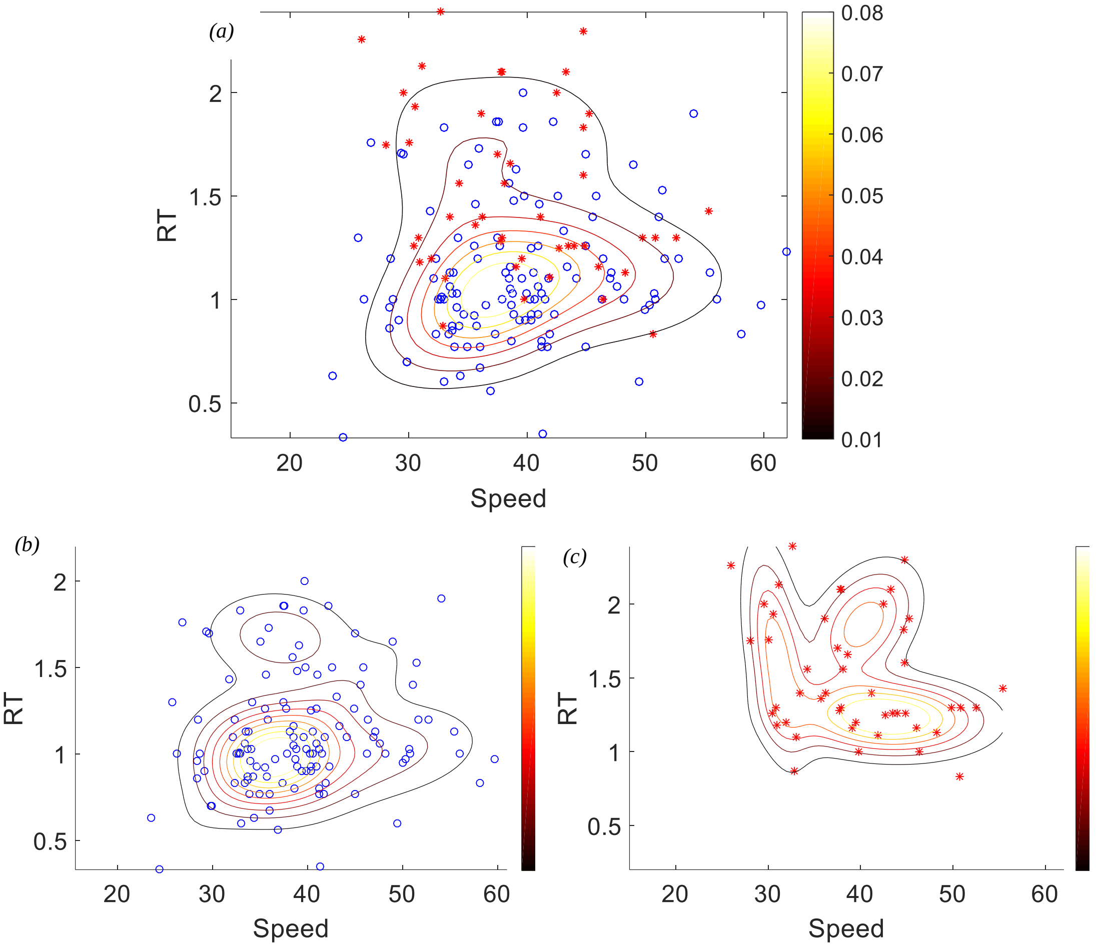


Figure 6: Rural Environment (a) All observations; (b) No collisions; (c) Collisions

In Figure 7 the speed vs. reaction time in a rural environment while raining is presented. When all the observations of the drivers using their phones are modeled, two peaks are presented. The two peaks present a pdf of 0.08 and 0.04 respectively. The cluster presenting the highest pdf presents a reaction time of around 1.2s with speed varying between 32km/h to 47km/h, which is below the speed limit for the rural area. In Fig. (7c) where the collisions are presented, two distinct peaks of equal pdf are shown. In the first cluster, the reaction time is between 1.5s and 2.2s with the respectful speed ranging between 15km/h to 50km/h, whereas in the second cluster the reaction time is between 1.2s and 1.5s and the speed between 20km/h to 50km/h. On the contrary, in Fig. (7b) one cluster presents an almost constant speed of 40km/h and a respectful reaction time of 1s.

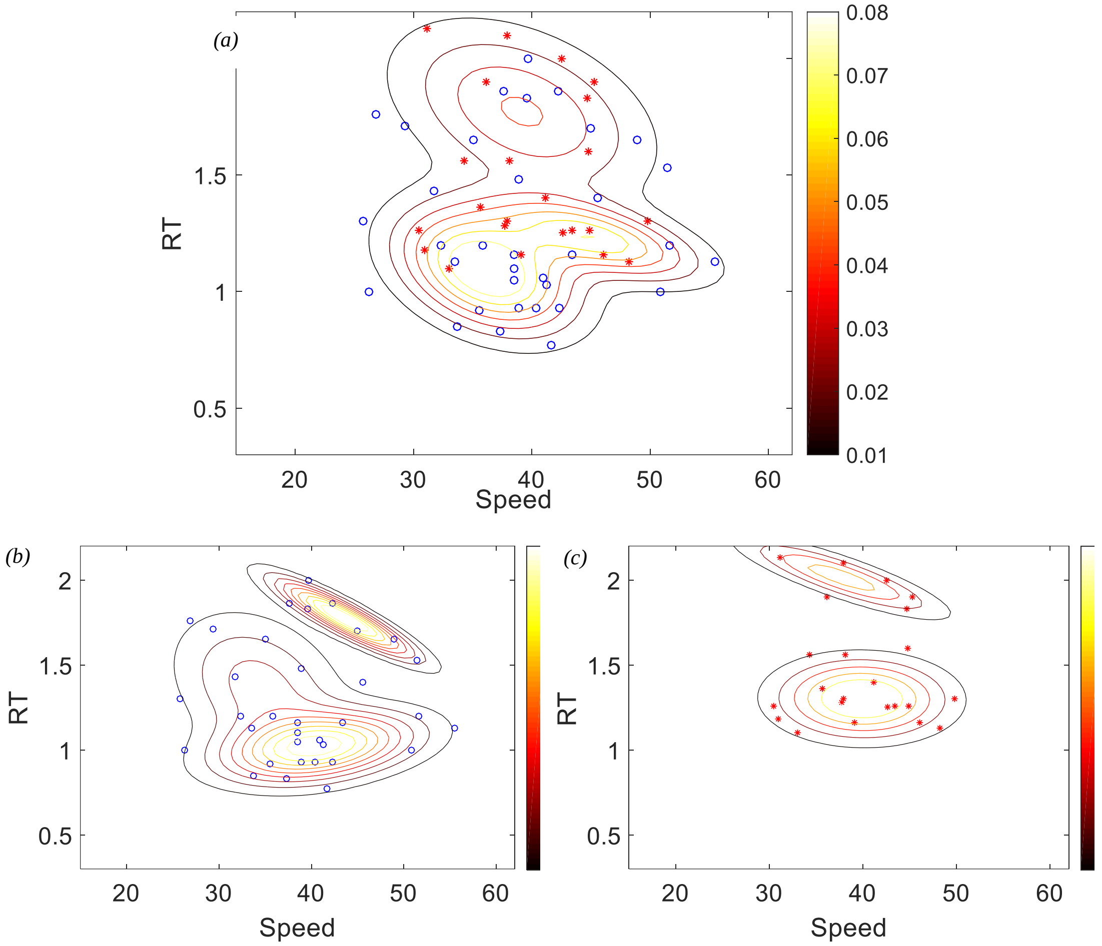


Fig. 7. Rural Environment and Rainy weather (a) All observations; (b) No collisions; (c) Collisions

Figure 8 presents the speed vs. RT in a rural environment in the night time. In Fig. (8b) one major cluster is presented, showing that within the range of 20km/h and 30km/h the reaction time was between 0.7s and 1.7s. However, as noted above this range of speed is quite low for a rural environment indicating the reduction of speed while the drivers were occupied. For the cases of collisions the reaction times of the drivers presented higher values.

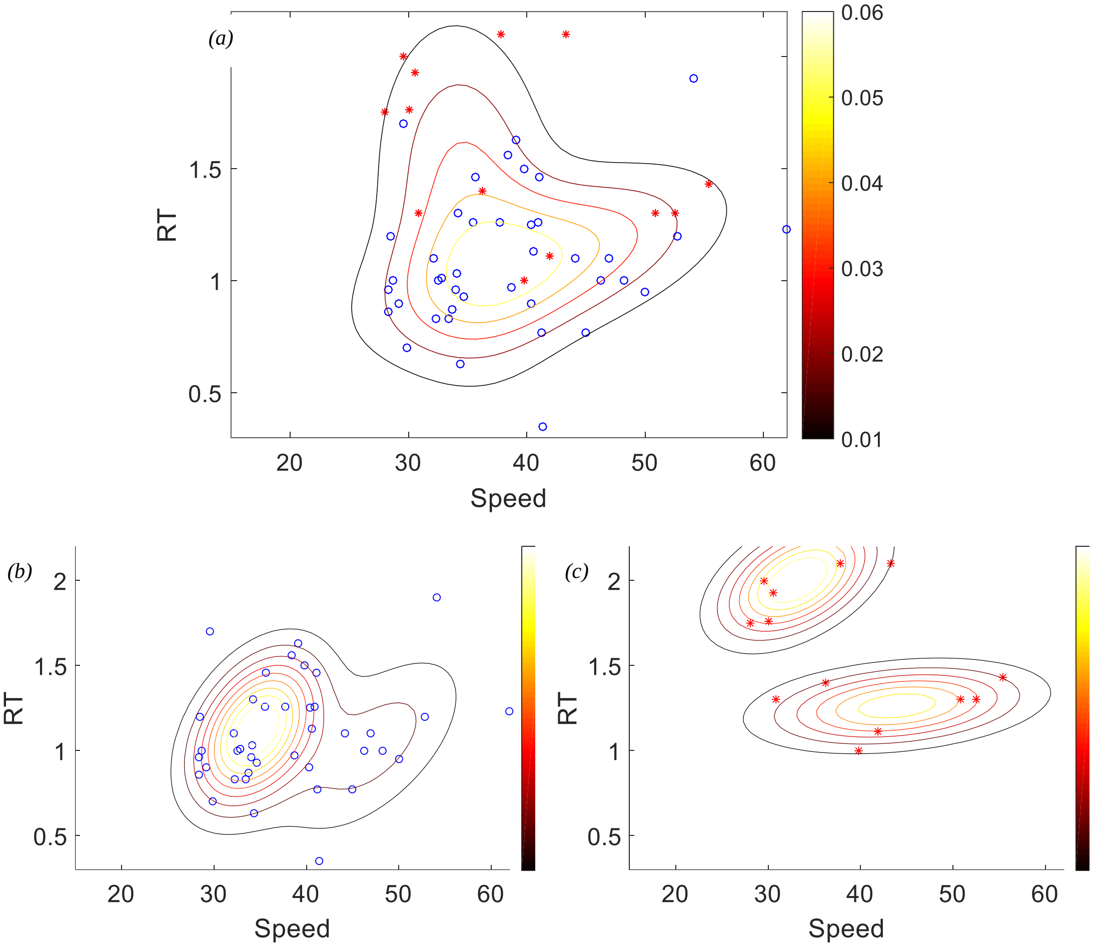


Figure 8: Rural Environment and Night time (a) All observations; (b) No collisions; (c) Collisions

4. Discussion and Conclusions

The aim of the presented study was to investigate the operational patterns of the driving behavior of young drivers when engaged in the secondary task of texting. The effects of texting were examined in combination with traffic conditions (urban-rural environment and normal-increased traffic conditions), as well as environmental conditions (good weather, rainy weather, nighttime).The analysis of the data consisted of two steps. In the first step a copula analysis was employed and in the second step the data was analyzed through clustering using Gaussian Mixture Modeling. A copula analysis is used in model dependence, and essentially describes the interrelation of several random variables. To take the analysis even further the use of GMMs provides the ability to regress multiple distributions which result in robust models.

In the current study, the use of mobile phone while driving was associated with poorer driving behavior in a simulated driving task. This research also explored the simultaneous actions of the drivers, i.e. the driver changing speed and at the same time the vehicle is off tracked, presenting consistent driving behavior patterns associated with the use of mobile phones while driving. In particular, in Scenario A, under all different conditions and regardless of whether the driver was engaged in a secondary task, speed and lane excursions present the highest correlations. This finding indicates that change of vehicle speed and vehicle off tracking occur simultaneously.

When examining Scenario B, the results indicate differences between the tasks that the driver is engaged into. In particular, when composing a text message, a correlation between speed and lane excursions is more frequent than when reading a message, where speed and reaction time are presented more frequently. This result shows the difference between driving behaviors when reading a text message and when composing a text message, indicating that the two tasks demand different levels of mental awareness. The highest correlations between two simultaneous actions were presented in the rural environment while it was raining suggesting that in this case, drivers change their driving style often. GMMs presented driving patterns of drivers associated with riskier driving behaviors when engaged in texting while driving. In general it was shown that the drivers who were involved in a collision while texting presented a different driving behavior compared to the drivers who were texting but were not involved in a collision. The current findings are not perhaps surprising, as they indicate that simultaneous task completions, in this case texting while driving, introduce distractions and as such contribute to poorer driving performance.

Such findings highlight the importance of future work in this area, since literature shows that drivers continue to use their mobile phones, despite the associated risk. The safety concerns that relate to texting while driving have attracted the attention of safety researchers, car manufacturers and policymakers. Being able to understand how texting impacts driving behavior can provide valuable insights that can be used in various manners including social awareness campaigns especially in young ages, illegalization of texting while driving worldwide or even technological solutions such as mobile applications that mute incoming calls and text messages while driving.

The results of this study can be used as a basis for further research, expanding the conditions examined here by investigating different traffic conditions (e.g. unfamiliar environment) or environmental conditions (e.g. snow). Furthermore, the investigation of other joined tasks as drinking or eating could allow the classification of risky behaviors while driving. Finally, given the findings of this research and the proposed approach in analyzing data, research can be expanded in models of statistical analysis or artificial intelligence.

References

Alosco, M., Spitznagel, M., Fischer, K., Miller, L., Pillai, V., Hughes, J., and Gunstad, J. Both Texting and Eating Are Associated With Impaired Simulated Driving Performance. Traffic Injury Prevention, Vol.13, 2012, pp. 468-475.

Atchley, P., Atwood, S., and Boulton, A. The choice to text and drive in younger drivers: Behavior may shape attitude. Accident Analysis and Prevention, Vol. 43, 2011, pp. 134-142.

Bhat, C., and Eluru, N. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transportation Research Part B: Methodological, Vol.43, 2009, pp. 749-765.

Christoforou, C. (2012). Investigation of the impact of texting to driver behaviour and safety with the use of driving simulator, Diploma Thesis. Greece: Department of Transportation Planning and Engineering, School of Civil Engineering, National TechnicalUniversity of Athens.

Cooper J., Yager C., and Chrysler T. An Investigation of the Effects of Reading and Writing Text Based Messages While Driving. No. SWUTC/11/476660-00024-1 U.S. Department of Transportation, 2011.

Cooper J., Yager C., and Chrysler T. An Investigation of the Effects of Reading and Writing Text Based Messages While Driving. No. SWUTC/11/476660-00024-1 U.S. Department of Transportation, 2011.

Crandall, J., and Chaparro, A. Driver Distraction: Effects of Text Entry Methods on Driving Performance. Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting. 2012, pp. 1693-1697

Darsow, F., W., Nguyen, B., and Olsen, N., E.Copulas and MarcovProcesses. Illinois Journal of Mathematics, Vol. 36, 1992, pp. 600-642.

Drews, F., Yazdani, H., Godfrey, C., Cooper, J., and Strayer, D. Text Messaging During Simulated Driving. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2009, pp. 762-770.

Demarta, S., andMcneil, A. The t Copula and Related Copulas. International Statistical Review, Vol. 73, 2005, pp. 111-129.

Dingus, T. A., Klauer, S.G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R.R. The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. No. H-810 593. U.S. Department of Transportation, National Highway Traffic Safety Administrator, 2006.

Farmer, C., Klauer, S., Mcclafferty, J., and Guo, F. Relationship of Near-Crash/Crash Risk to Time Spent on a Cell Phone While Driving. Insurance Institute for Highway Safety, 2014, pp. 1-9.

Fraley, C., andRaftery, A. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. Journal of the American Statistical Association,2002, pp. 611-631.

Gartzonikas, C. (2012). Impact of Texting on Young Drivers' Traffic and Safety on Motorways by the use of a Driving Simulator, Diploma Thesis. Greece: Department of Transportation Planning

He, J., Chaparro, A., Wu, X., Crandall, J., and Ellis, J. Mutual interferences of driving and texting performance. Computers in Human Behavior, Vol.52, 2015, pp. 115-123.

Hosking, S., Young, K., and Regan, M. The Effects of Text Messaging on Young Drivers. Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 51, No.4, 2009, pp. 582-592.

Jin, S., Wang, D., Xu, C., and Ma, D. (2013). Short-term traffic safety forecasting using Gaussian mixture model and Kalman filter. Journal of Zhejiang University SCIENCE A .(Applied Physics and Engineering), Vol.14, pp. 231-243.

Kalbfleisch, J. Probability and statistical inference. Springer-Verlag, New York, 1979.

Klauer, S.G., Dingus, T. A., Neale, V. L., Sudweeks, J.D., and Ramsey, D.J. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis using the 100-Car Naturalistic Driving Study Data. No. HS-810 594. U.S. Department of Transportation, National Highway Traffic Safety Administrator, 2006.

Lord, D., and Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, Vol. 44, 2010, pp. 291-305.

Mckeever, J., Schultheis, M., Padmanaban, V., and Blasco, A. Driver Performance While Texting: Even a Little is Too Much. Traffic Injury Prevention, Vol. 14, No.2, pp. 132-137.

Metz, B., Landau, A., and Just, M. Frequency of secondary tasks in driving – Results from naturalistic driving data. Safety Science, Vol. 68, 2014, pp. 195-203.

Olson L.R., Hanowski J.R., Hickman S.J., and Bocanegra J. Driver Distraction in Commercial Vehicle Operations. No FMCSA-RRR-09-042. U.S Department of Transportation. 2009.

Reynolds, D. Gaussian mixture models. Encyclopedia of Biometrics, pp. 659-663. Springer US, 2009.

Rumschlag, G., Palumbo, T., Martin, A., Head, D., George, R., and Commissaris, R. The effects of texting on driving performance in a driving simulator: The influence of driver age. Accident Analysis and Prevention, Vol.74, 2015, pp. 145-149.

Schmidt, T. Coping with Copulas. From Theory to Application in finance., 2007, pp. 3-34

Yannis, G., Laiou, A., Papantoniou, P., and Christoforou, C. Impact of texting on young drivers' behavior and safety on urban and rural roads through a simulation experiment. Journal of Safety Research,Vol. 49,2014, pp. 25-31.