Extreme value theory approach to analyze safety of passing maneuvers considering drivers' characteristics

Joana Cavadas, Carlos Lima Azevedo, Haneen Farah

CITTA, Department of Civil Engineering, University of Coimbra, Portugal, joana.cavadas@uc.pt

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, 2139 Cambridge, MA, United States, cami@mit.edu

Department Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, P.O. Box 5048, 2600 GA Delft, The Nether-lands, h.farah@tudelft.nl

Abstract

Observed accidents have been the main resource for road safety analysis over the past decades. Although such reliance seems quite straightforward, the rare nature of these events has made safety difficult to assess, especially for new and innovative traffic treatments. Surrogate measures of safety have allowed to step away from traditional safety performance functions and analyze safety performance without relying on accident records. In recent years, the use of surrogate measures to estimate accident probabilities with extreme value theory (EV) models has been an alternative approach to its use as aggregate accident frequency predictors. In this paper we extend existing efforts on EV for accident probability estimation with driver specific characteristics and joint distributions estimates using two dependent surrogate measures. Using detailed trajectory data from a driving simulator, we model the probability of head-on and rear-end collisions in passing maneuvers.

We show that accounting for driver specific variables and road infrastructure variables improve the head-on collision probability estimation. This is valuable since driver and road heterogeneity can now be considered in evaluating various safety interventions both for in-vehicle and infrastructure based solutions.. We also present an exploratory structure and results for combining surrogate measures that describe correlated events. Such feature is essential to keep up with the expectations from surrogate safety measures for the integrated analysis of accident phenomena.

Keywords

Road safety; Extreme value theory; Driving Behavior; Minimum time to collision; Minimum time head-way

1. Introduction

1.1. Motivation & Problem Definition

Prediction of accidents has been a major topic in traffic safety for the last couple of decades. Despite the huge efforts that researchers have put in developing accident prediction models [**1**], there is a great tendency in the last decades to develop new proactive methods for safety evaluation that are not based on accident records (*[2, 3]*. Evaluating conflicts and risky situations between road users has been the main alternative and multiple methodologies can be found in the literature: the Swedish traffic conflict technique [**4**], DOCTOR method [**3**], and the use of surrogate safety measures [**2**]. The main challenge is the link between these measures and the number of accidents. Zheng, Ismail and Meng [**5**] indicate that the validity of surrogate safety measures is usually determined by its correlation with accident frequency which is usually assessed using regression analysis. However, regression analysis still incorporates accident counts which are known to suffer from underreporting and quality issues, and thus this approach is limited. Besides, it is difficult to insure the stability of the accident-to-surrogate ratio and this relationship also hardly reflects the physical nature of accident occurrence [**5**]. Therefore, there is a need to develop an alternative approach to predict the number of accidents based on surrogate safety measures. Recently [**6**] developed a new and more sophisticated approach based on the Extreme Value (EV) theory to estimate the frequency of accidents based on measured accident proximity.

1.2. Extreme Value (EV) Approach

The EV approach has three considerable advantages over the traffic conflict technique: (a) it abandons the assumption of fixed ratio converting the surrogate event frequency into accident frequency; (b) accident risk given the surrogate event is estimated based on the observed variability of accident proximity without using accident data; (c) the accident proximity measure precisely defines the surrogate event.

The implicit assumption of the EV theory is that the stochastic behavior of the process being modeled is sufficiently smooth to enable extrapolation to unobserved levels [**6**]. In the context of road safety, the more observable traffic conflict events are used to predict the less frequent accidents, which are often unobservable in a short time period [**5**]. The field of EV theory, pioneered by Fisher and Tippett [**7**], is a commonly applied theory in many fields, such as in meteorology, hydrology, finance [**5**] and very recently, road safety analysis Songchitruksa and Tarko [**6**]. Songchitruksa and Tarko [**6**] used an EV approach to build up relationships between occurrences of right-angle accidents at urban intersections and frequency of traffic conflicts measured by using post-encroachment time. A major improvement of this study is that it links the probability of accident occurrence to the frequency of conflicts estimated from observed variability of accident proximity, using a probabilistic framework and without using accident records. The generic formulation of the application of EV to road safety analysis was then proposed by Tarko [**8**] and it was only very recently applied to other accident types and data sets [**5, 9**]. Despite these recent efforts the extent in which EV can accurately be used in accident frequency prediction is yet to be tested.

1.3. Risk of Passing Maneuvers

Passing maneuvers on two-lane roads (one lane per travel direction) carries several types of risks. The process of passing involves, synchronizing the vehicle's speed with that of the front vehicle, estimating the available gap on the opposite direction and evaluating its suitability to successfully perform the passing maneuver, and finally return to the main driving lane while keeping a sufficient safe gap from the passed vehicle, as well as, from the vehicle on the opposite direction. The gap from the passed vehicle is termed 'THW' for time headway between the front of the passed vehicle and the rear of the passing vehicle – a measure for rear- and side-collisions with the passed vehicle. The gap from the vehicle on the opposite direction is termed 'TTC' for time-to-collision between the passing and the opposite vehicle – a measure for head-on collisions. Both of these gaps are calculated at the end of the passing maneuver. In this study both measures will be used: the THW was assumed as the remaining distance between the passing and passed vehicle divided by the driving speed of the passed vehicle, while the TTC was calculated as the remaining distance between the passing and opposing vehicle divided by the sum of their speeds.

1.4. Drivers' Characteristics

Several studies have shown that there are significant differences in passing behaviors among different drivers. Farah [**10**] using a driving simulator found that gender and age have a significant impact on the passing behavior. She found that the passing frequency of male drivers is higher than female drivers, male drivers also maintain smaller following time gaps from the front vehicle before initiating a passing maneuver, and have smaller critical gaps. Younger drivers, compared to older drivers, have significantly lower critical gaps and higher desired driving speeds, and keep smaller gaps from the front vehicle at the end of the passing maneuvers. These behaviors increase the risk of accidents. Llorca, Garcia, Moreno and Perez-Zuriaga [**11**] reached similar conclusions using an instrumented vehicle. The authors found that young male passing drivers have shown a more aggressive behavior. Passing times were around 1s lower than other drivers, while average speed difference was 4 km/h higher. Farah, Polus, Bekhor and Toledo [**12**] tested the significance of including driving styles in the passing behavior model, and found that drivers who are characterized by an anxious driving style and/or patient and careful driving style have larger critical gaps. Vlahogianni and Golias [**13**] emphasize that the behaviors of young male and female drivers during passing maneuvers are different, because of the differences in the process of scanning and evaluating available opportunities for passing.

To summarize, the integration of drivers' characteristics and driving styles in accident prediction is valuable and can contribute to understanding of the accident causation contributing factors and contribute to the improvement of the model prediction accuracy. Previous EV models did not account for these factors.

2. Research Method

The aim of this study is to develop and test two different methods to estimate accident probability in passing maneuvers. The first approach analyzes the risk of individual types of accidents during passing maneuvers: (1) head-on collisions using the proximity measure of the minimum TTC to the vehicle in the opposite direction at the end of a passing maneuver; (2) rear-end collisions with the passed vehicle using the proximity measure of the minimum THW from the front of the passed vehicle to the rear-end of the passing vehicle at the end of the passing maneuver. The second approach aims to analyze the joint risk of colliding with the opposite vehicle or with the passed vehicle during passing maneuvers using the two surrogate safety measures (THW and TTC).

2.1. Modeling Approach

There are two families of EV distributions which follow two different approaches to sample extreme events: (1) the Generalized Extreme Value (GEV) distribution which is used in the block maxima or minima (BM) approach, in which maxima over blocks of time (or space) are considered; (2) the Generalized Pareto Distribution (GPD) which is used in the peak over threshold approach [**14**], where all values above some high level are used. In this paper only the BM approach is examined following the conclusion reached by Azevedo and Farah [**15**] and Farah and Azevedo [**16**] that the peak over threshold approach was found to be less efficient for the analysis of the risk of passing maneuvers using the same dataset as in this study. Therefore, non-stationary Block Maxima approach was applied for estimating the risk of a single type of accident (head-on or rear-end collision), while for estimating the risk of both types of collisions jointly the bivariate distribution with copula approach was considered. These two methods are described in detail below.

2.2. Non-Stationary Block Maxima Approach

In the GEV distribution the extreme events are sampled based on the BM approach. Following this approach the observations are aggregated into fixed intervals over time and space, and then the extremes are extracted from each block by identifying the maxima in each single block. Mathematically, the standard GEV function is as follows [**5**]:

$$G(x) = \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right) \qquad \text{(eq. 1)}$$

where, $\{X_1, X_2, \dots, X_n\}$ is a set of independently and identically distributed random observations with unknown distribution function $F(x) = Pr(X_i \leq x)$, the maximum $\mathrm{M_n} = \max\{\mathrm{X_1}, \mathrm{X_2}, \dots, \mathrm{X_n}\}$ will converge to a GEV distribution when $n \to \infty$. Three parameters identify this distribution: the location parameter, $-\infty < \mu < \infty$; the scale parameter, $\sigma > 0$; and the shape parameter, $-\infty < \xi < \infty$. If the shape parameter, ξ, is positive, then his would yield the Frechet Cumulative Distribution Function (CDF) with a finite lower endpoint, $(\mu - \sigma/\xi)$, if ξ is negative, this will yield the (reversed) Weibull CDF with finite upper endpoint $(\mu + \sigma/|\xi|)$, and if $\xi = 0$ this yields the Gumbel CDF.

In a non-stationary BM model several impacting factors can be included typically in the location parameter to account for their impact on the probability of the extreme events. The BM method can also be used to study minima by considering the maxima of the negated values instead of minima of the original values. Such property is extremely useful for the analysis of gap and distanced based variables, typically used in surrogate measures. More details on the GEV properties can be found in [**8**].

In some applications the study of probability using multivariate distributions is of interest. Traditionally, single surrogate safety measures are used to estimate a single type of events. However, it is expected that in some of the complex accident phenomena, multiple pre-accident events can play an important role in a potential accident. Passing maneuvers are a typical case where both the opposite and passed vehicles are key stimulus during driver's decision making.

2.3. Bivariate distributions with Copula method to model dependence

We focus our attention on the bivariate distribution with copula method due to the expectable existence of dependence (not necessary linear) between the two surrogate safety measures. A copula is a multivariate distribution whose margins are all uniform over (0,1), following as shown in equation (2) for a p-dimensional random vector U on the unit cube:

$$C(u_1, \dots, u_p) = \Pr(U_1 \leq u_1, \dots, U_p \leq u_p). \qquad \text{(eq. 2)}$$

The copula not only provides a structure for the dependence between the variables but also reveals itself to be invariant under strictly monotone transformations. The Sklar's Theorem (1959) ensures that it is possible to estimate a multivariate distribution by separately estimating the marginal distributions and the copula function C. In this sense, let F be the p-dimensional distribution function of the random vector U with margins $F_1, \dots, F_p$, then the copula C is such that for all vector x the equality $F(x) = C(F_1(x1), \dots, F_p(x_p))$ holds, where C is unique if the marginal distributions are continuous.

The two most frequently used copula families are elliptical copulas and Archimedean copulas. More details on these copula families can be found in Fang, Kotz and Ng [**17**], Nelsen [**18**], and Genest and Rivest [**19**]. To assess if a given copula is well fitted to the data under analysis, goodness-of-fit are performed based on statistics such as the rank-based versions of the Cramer-von Mises or the Kolmogorov-Smirnov. An example of goodness-of-fit testing overview are given in Berg [**20**].

2.4. Data Collection

Data on the time-to-collision (TTC) with the opposing vehicle and the time headway (THW) between the passed and passing vehicles at the end of passing maneuvers were obtained from a driving simulator experiment previously developed by Farah, Bekhor and Polus [**21**] for modelling drivers' passing behavior on two-lane rural highways. In this experiment the STISIM [**22**] driving simulator was used. STISIM is a fixed-base interactive driving simulator, which has a 60° horizontal and 40° vertical display. The driving scene was projected onto a screen in front of the driver with a rate of 30 frames per second.

In this experiment a total of 16 simulator scenarios were designed in order to have a better understating on how different infrastructure and traffic related factors affect drivers' passing behavior. The 16 different scenarios are the result of an experimental design that included 4 factors in 2 levels, which are: the speed of the front vehicle (60 or 80 km/h), the speed of the opposite vehicle (65 or 85 km/h), the opposite lane traffic volume (200 or 400 veh/h), or the road curve radius (300-400 or 1500-2500 m). However, all the scenarios included 7.5 km of two-lane rural highway section with no intersections, and good weather conditions. Each driver drove 4 scenarios out of the 16 scenarios which were selected following a partial confounding method that was adopted [**23**]. A more detailed information about this experiment can be found in [**21**].

A total of 100 drivers (64 males and 36 females) with at least 5 years of driving experience participated in the driving simulator experiment on a voluntary base. 67 drivers have an age between 22 and 34 years old, 20 drivers with an age between 35 and 49 years old, and the remaining 12 with an age between 50 and 70 years old.

Each driver, before participating in the driving simulator experiment, filled a questionnaire composed of two parts: the first part included questions on the driver personal characteristics (including questions such as: gender, age, and driving experience), while the second part included the multidimensional driving style inventory (MDSI) developed by Taubman Ben-Ari et al. [**24**]. The MDSI is a 6-point scale, which consists of 44 items that are used to characterize four factors that represent different driving styles: (1) *Reckless and careless driving style*, which refers to deliberate violations of safe driving norms, and the seeking of sensations and thrill while driving. It characterizes persons who drive at high speeds, race in cars, pass other cars in no-passing zones, and drive while intoxicated, probably endangering themselves and others; (2) *Anxious driving style*, which reflects feelings of alertness and tension as well as ineffective engagement in relaxing activities during driving; (3) *Angry and hostile driving style*, which refers to expressions of irritation, rage, and hostile attitudes and acts while driving, and reflects a tendency to act aggressively on the road, curse, blow horn, or "flash" to other drivers, and (4) *Patient and careful driving style*, which refers to planning ahead, attention, patience, politeness, and calmness while driving, as well as obedience to traffic rules. Factor scores were calculated for each respondent on each of these four driving styles.

3. Results and Analysis

The data set from the driving simulator experiment resulted in a total of 1287 completed passing maneuvers, 9 head-on collisions and 2 rear-end collisions.

3.1. Univariate Model

To fit the 1287 passing maneuvers and their respective minimum TTC and THW measurements, a Generalized Extreme Value (GEV) distribution was considered to measure the risk of each type of accident. In this approach, a set of block intervals was defined according to the annotated time that contains the entire passing maneuver, having as a result a variable number of observations for each block as well as the number of blocks.

3.1.1 Head-on collisions

Aiming at estimating the probability of a head-on collision for a single passing maneuver, the minimum TTC was considered as a head-on accident surrogate measure. The data was then filtered to account only for values smaller than 1.5s [**4, 9, 25**], leading to a total of 463 observations. Knowing that 9 maneuvers ended with actual head-on collisions, the probability of a head-on collision in a passing maneuver is 9/472=0.0191, with 95% binomial confidence interval (0.0089, 0.0366).

The estimation of the stationary BM model was developed by Farah and Azevedo (18) for the negated values instead of minima of the original values, i.e., $min\{TTC\}$. The authors estimated that the parameters of the univariate GEV cumulative function are $\hat{\mu} = -0.993\,(0.0212)$, $\hat{\sigma} = 0.0383\,(0.0163)$ and $\hat{\hat{\xi}} = -0.236\,(0.0500)$. Figure 1. presents the probability density function of the empirical and modeled negated TTC (*upper left*) and the simulated QQ plot (*upper right*). This model was then improved to a non-stationary BM model by the authors. They concluded that the covariates '*passinggap*', '*tailgatetp*', '*speedfront*', '*curvature*', related to the infrastructure and traffic, were found to significantly contribute to the prediction of the probability of a head-on-collision during a passing maneuver Farah and Azevedo [**26**]. While the covariate '*speedpv*' was not found to be significant. Variables related to drivers' personal characteristic (gender, age, and driving style) were not tested. In this study, we will include the traffic and road variables that were found to be significant, and we will test whether drivers' personal characteristic significantly contribute to the model. The variables are defined as following:

- *passinggap*: The time gap between two opposite vehicles at the time the subject meet the lead opposite vehicle ;
- *tailgatetp:* The time gap between the subject vehicle and the front vehicle at the moment of start passing (s);
- *speedfront*: The speed of the front vehicle at the moment of start passing (m/s);
- *curvature*: The road curvature (1/m);
- *speedpv*: The speed of the passing vehicle (m/s);

- *gender*: The gender of the driver (1-male; 0-female);
- *age*: Categorical variable, with ranges 22-34; 35-49 and 50-70 ;
- *drivingstyle*: Angry & Hostile; Anxious; Reckless & careless; Patient & careful [**24**].

A set of a non-stationary models considering different covariates were estimated. After testing several linear combinations, the four best non-stationary models are presented in Table 1 with a range of likelihood ratio p-value between 2.773×10^{-9} and 1.931×10^{-8} (Model #1 to #4). The estimated likelihood ratio tests are shown in Table 2. The previously estimated stationary model and non-stationary model without driver-specific variables (Model #0) from [**26**] are used as a benchmark in the remaining of this section.

Table 1: Estimation Results of the 5 Best Models for the BM Approach

Non-stationary model	#0 [**26**]	#1	#2	#3	#4
	Estimated v. (Std. Error)	Estimated v. (Std. Error)	Estimated v. (Std. Error)	Estimated v. (Std. Error)	Estimated v. (Std. Error)
$\hat{\mu}_0$	-1.0451 (0.1377)	-0.9838 (0.1394)	-0.9273 (0.1457)	-0.9528 (0.1456)	-1.1071 (0.139)
$\hat{\mu}_1$ (speedFront)	0.0245 (0. 0064)	0.0265 (0.0064)	0.027 (0.0064)	0.0257 (0.0064)	0.0273 (0.0064)
$\hat{\mu}_2$ (tailgatetp)	0.0026 (0.0018)	0.0028 (0.0018)	0.0028 (0.0018)	0.0026 (0.0018)	0.0027 (0.0018)
$\hat{\mu}_3$ (passinggap)	-0.022 (0.0044)	-0.0232 (0.0044)	-0.0235 (0.0044)	-0.0226 (0.0044)	-0.0227 (0.0044)
$\hat{\mu}_4$ (curvature)	-33.6534 (13.5196)	-34.3039 (13.4199)	-34.0688 (13.4034)	-34.0902 (13.4888)	-34.1393 (13.3976)
$\hat{\mu}_5$ (Gender)	-	-0.0967 (0.0421)	-0.0804 (0.0438)	-	-
$\hat{\mu}_6$ (Angry&Hostile)	-	-	-0.0211 (00162)	-0.0294 (0.0156)	-
$\hat{\mu}_7$ (F2234)	-	-	-	-	0.1166 (0.0442)
$\hat{\sigma}$	0.3639 (0.0145)	0.3616 (0.0143)	0.3607 (0.0142)	0.362 (0.0143)	0.3611 (0.0142)
$\hat{\varepsilon}$	-0.2196 (0.042)	-0.2176 (0.0413)	-0.216 (0.041)	-0.2163 (0.0414)	-0.2175 (0.0405)
Negative Loglikelihood	208.6541	206.0598	205.2183	206.8908	205.2379

Analyzing the results presented in Table 1, it is concluded that the introduction of *Gender* improves the accuracy of the model when compared to the non-stationary model (#0) developed by Farah and Azevedo [**26**]. The significance of this variable is given by the p-value of the likelihood ratio test, which is equal to 0.023, with 95% confidence level.

Table 2: Direct Value of the Likelihood Ratio Test (p-value) for the 5 Best Models for the BM Approach

Non-stationary model	0 [**26**]	#1	#2	#3	#4
#0 [**26**]	-				
#1	5.1887 (0.023)	-			
#2	6.8717 (0.032)	1.6831 (0.194)	-		
#3	3.527 (0.060)	-1.662 (1)	3.3451 (0.0674)	-	
#4	6.832 (0.008)	1.644 (2.2E-16)	0.0393 (0.843)	3.3057 (2.2E-16)	-

The contribution of the variables representing driving styles (*Angry&Hostile*, *Anxious*, *Reckless&Careless* and *Patient&Careful*) was tested considering all the possible combinations of these variables besides the ones included in Model 1. Comparisons between the different models were based on the likelihood ratio test. This procedure resulted in the inclusion of one driving style, *Angry&Hostile*, as presented in model #2. Analyzing the correlation between the different driving styles and the remaining variables, a small correlation of 0.29 was found between *Angry&Hostile* and *Gender*. Taking this into account, including the variable *Angry&Hostile* in a model with the variable *Gender* (model #2), may not contribute to improve our estimation. In order to test which variable among the two has a larger influence, the variable *Gender* was excluded from model #2, creating model #3. Comparing the results of models #1 to #3, it is concluded that the model that only includes *Gender* (model #1) has a better fit based on the p-value of the likelihood ratio test. Another reason to prefer this model is the simplicity of collecting data on driver gender compared to the driving styles, which require drivers to complete the MDSI survey.

Aligned with the conclusions achieved by Farah [**10**] and Llorca, Garcia, Moreno and Perez-Zuriaga [**11**] regarding the impact of age, this variable (*age*) was found to improve the accuracy of the model when compared to the stationary model but turned out to have a non-significant contribution if gender is also included. After several attempts, we included the interaction variable between gender (female drivers) with age (range 22-34), *F2234*, and the final model (model #4) is shown in Table 1. This model considers a new variable that takes 1 if the driver is a female with age range between 22 and 34, and zero otherwise.

To estimate the probability of a head-on collision along with the conclusion about which model is the one with the better fit (models #1 and #4), two different approaches were considered. The first approach considers that the location parameter value is calculated using the covariates from the data, achieving the estimated probabilities of 0.0195 and 0.0198 for models 1 and 4, respectively, with 95% confidence level (0.0192; 0.0198) and (0.0195; 0.0201), respectively. These confidence intervals of estimation were computed assuming a normal distribution under regular parameters' conditions, a simulation experiment size of 1×10^6 and its simulated distribution quantiles. The second approach considers the estimation of the location parameters based on the estimation dataset, where normal distributions with means (standard deviations) of -0.989 (0.123) and -0.988 (0.125), for models #1 and #4, respectively were considered. The Kolmogorov-Smirnov test statistic of 0.0444 and 0.0479, respectively was achieved. This procedure simulates the values 0.0197 and 0.0202 for the probabilities of head-on collisions of models 1 and 4, respectively, with 95% confidence interval of (0.01939, 0.0199) and (0.0199, 0.0205). Comparing the probabilities of these two methods with the probability for a head-on collision assuming a near head-on collision in a passing maneuver of 0.0194, results in model 1 giving better estimation compared to model #4. We can conclude that model 1 is better than model #4, for estimating the probability of head-on collisions for single passing maneuvers.

According to the results of model #1 presented in Table 1, if the speed of the front vehicle (*speedfront*) increases, or if drivers start their passing maneuver from a larger gap from the front vehicle (*tailgatetp*), the negated TTC increases (corresponding to a decrease in the TTC). These are logical results since it is more difficult to end the passing maneuver if the front vehicle has a higher driving speed. Similarly, starting the passing maneuver from a larger gap from the front vehicle results in a longer time to finish the maneuver and consequently smaller TTC. If the passing gap (*passinggap*) that is accepted is larger or the curvature of the road (*curvature*) is higher, the negated TTC is lower and the TTC is higher. This shows that drivers adapt their behavior in a passing maneuver if the road is too complex. Finally, male drivers have smaller TTC. This result is supported by previous studies [**12, 13**], where it was found that male drivers usually drive faster, have shorter passing gaps, and conduct a higher number of passing manoeuvers when compared to females.

The probability density function of the empirical and modeled standardized[1] maximum negated TTC and the simulated QQ-plot for the best non-stationary BM model (model #1) are shown in Figure 1.

[1] For non-stationary models, it is common practice to transform the data to a density function that does not depend on the covariates, using the following function $Z_i = -\log\left(1 + \frac{\epsilon}{\sigma} \times (X_i - \mu_i)\right)^{-\frac{1}{\epsilon}}$ 27. E. Gilleland and R. W. Katz, *New software to analyze how extremes change over time*, Eos, Transactions American Geophysical Union **92** (2011), no. 2, 13-14.

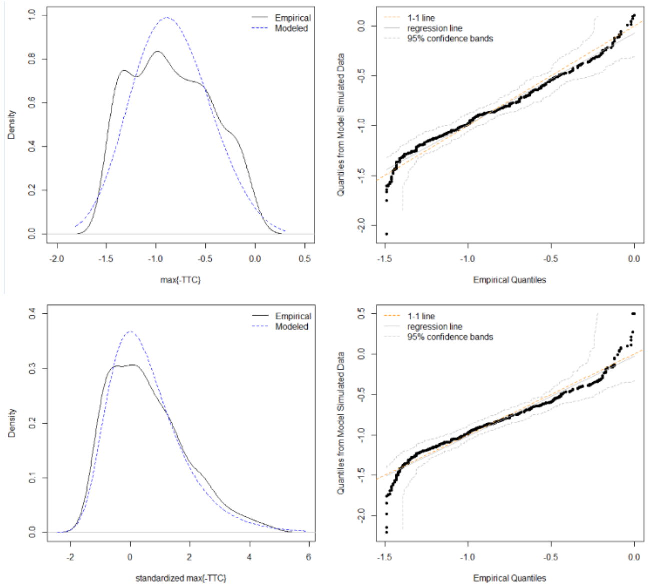

Figure 1: Probability Density plot (upper-left) and simulated QQ-plot (upper-right) for the stationary BM model (*18*) and standardized Probability Density plot (lower-left) and simulated QQ-plot (lower-right) for the best non-stationary BM model (Model #1).

3.1.2 Rear-end collisions

In order to estimate the probability of rear-end collisions for a single passing maneuver, the headway between the front passed vehicle and the passing vehicle at the end of the passing maneuver (THW) is used as accident surrogate. Similar to what was developed to calculate the probability of a head-on collision for a single passing maneuver, the minimum THW should be smaller than a limit to be useful as an accident surrogate. Based on the literature, this value varies between <0.6 [**25**] and <2 [**25, 28**]. Considering these thresholds, several BM stationary models were developed. Based on the estimated probabilities for rear-end collisions, it was concluded that 1.2s is the threshold with the best fit.

With a total of 81 observations (excluding the observations that ended with actual collisions) that have a minimum THW smaller than 1.2s and knowing that 2 rear-end collisions occurred, the theoretical probability of a rear-end collision was calculated as 2/83= 0.024, with a 95% binomial confidence interval (0.00293, 0.0843).

The estimation of the stationary BM for the model of the negated values of the THW, resulted in the parameters $\hat{\mu} = -1.11\ (0.0137)$, $\hat{\sigma} = 0.0956\ (0.0132)$ and $\hat{\xi} = 0.536\ (0.1666)$ for the GEV cumulative distribution function. The density function of the empirical and modeled negated THW and the simulated QQ plot are shown in Figure 2. Taking these plots into consideration, it is concluded that the fitting results of the modeled GEV distribution are satisfactory because the empirical data points fall close to the 45º line in the simulated QQ plot.

Using the fitted GEV distribution, the estimated probability of this stationary model of max{-THW} >=0 is 0.0246 with 95% confidence interval (0.0243, 0.0249). This interval was computed assuming a normal distribution under regularity conditions of the parameters, simulating an experiment with a size of $1x10^6$ and its simulated distribution quantiles. This estimated probability is comparable with the empirical probability of 0.02409. Concluding that the stationary BM model is a good approach to estimate the probability of a rear-end collision.

Notwithstanding, the passing maneuver may be affected by specific passing conditions, such as speeds of the vehicles surrounding the subject vehicle. Therefore, several linear combinations of covariates were tested according to a non-stationary BM model approach. This process was conducted in a similar way to the model developed to estimate the probability of a head-on-collision. Taking this into account, the final non-stationary BM model includes the covariate related with the passing vehicle speed (*speedpv*), where the location parameter takes the values $\hat{\mu}_0 = -1.169$ and the parameter for the covariate *speedpv* $\hat{\mu}_1 = 0.00266$, while $\hat{\sigma} = 0.0971$ and $\hat{\xi} = 0.453$. Testing this non-stationary model against the stationary one through the likelihood ratio test, a p-value of 0.0421 is achieved with a direct value of 4.128. The probability density plot as well as the QQ-plot for this model is also shown in Figure 2.

To estimate the probability of $max\{THW\} \geq 0$ with this non-stationary model, a normal distribution was fitted taking into account the estimated location parameter. This distribution with a mean of -1.102, a standard deviation of 0.0179 and a Kolmogorov-Smirnorv test statistic of 0.118, lead to a simulated probability of 0.0181 for the $max\{THW\} \geq 0$ with 95% confidence interval (0.0179, 0.0184), resulting in a worst estimation than the stationary model. Therefore, the model with the better fit is the stationary BM model (Figure 2.).

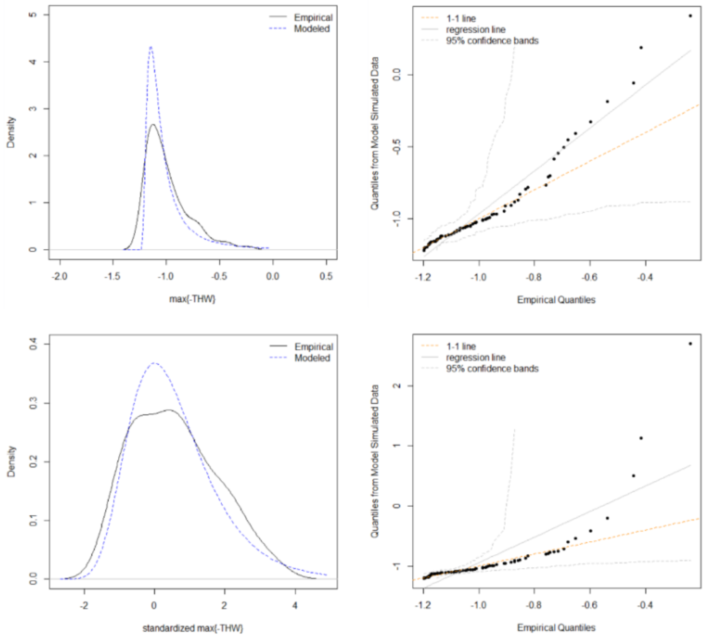

Figure 2. Probability Density plot (upper-left) and simulated QQ-plot (upper-right) for the stationary BM model and standardized Probability Density plot (lower-left) and simulated QQ-plot (lower-right) for the non-stationary BM model (with *speedpv* as covariate).

3.2. Bivariate Model

It is aimed to estimate a measure of risk that not only takes into account the possibility to collide with the opposite vehicle but also with the passed vehicle. Taking into account that performing a passing maneuver requires a split of attention by the driver regarding its location relative to the surrounding vehicle, it is assumed that the dependence between the TTC and the THW is unknown. Furthermore, an integrated analysis is possible to be developed if, and only if, a relationship of dependence can be found between TTC and THW. When examining the correlation between these two variables with using the whole dataset, a Pearson-correlation value of 0.186 was found. This value shows the lack of linear correlation between the time-to-collision measure and the time head-way measure. However, this does not mean that TTC and THW are independent [**29**]. To further examine potential correlation, the Kendall's rank correlation tau was computed and found to be significantly greater than zero, indicating the existence of dependence between TTC and THW ($\tau = 0.192, p - value = 2.2e - 16$). This statement is corroborated by the independence test Global Cramer-von Mises, where a significant p-value close to zero (p-value=0.000499) gives strong evidence against the null hypothesis of independence.

Based on these results, different copula families were estimated to model the dependence between the negated values of TTC and of THW. This analysis concluded that the copula with the better fit is the Joe-Frank copula [**30**], with parameters 1.631 and 0.929. This result was confirmed by performing the goodness-of-fit test based on Kendall's process (0.47 and 0.26 for the p-values of Cramer-von Mises statistic and Kolmogorov-Smirnov statistics, respectively).

Due to copula estimation' property of allowing separate estimations for the copula and the marginal distributions, we now proceed with the informed estimation of the latter, using insights from the previous univariate analysis. For this exploratory analysis, we estimate the bivariate distribution considering the stationary univariate BM distributions for each variable as the upper tail margins distributions. This integration estimates the probability of an accident if both TTC and THW are below their thresholds (1.5s and 1.2s, respectively). To perform the estimation for the remaining TTC and THW values, other distributions for the margins should be analyzed and fitted (e.g., gamma distributions or Gumbel distribution).

Simulating elements for the bivariate distribution analyzed in this exploratory approach, with a Joe-Frank copula and GEV distributions for the margins, a maximum loglikelihood of 50.73 and a Kendall's tau of 0.184 are achieved. The Kendall's tau is comparable to the one previously achieved for the original dataset (0.192), revealing the suitability of our approach to estimate the joint probability of an accident based on the two surrogate measures (TTC and THW).

Using this fitted distribution, the estimated probability of having an accident, if both surrogate measures are below their threshold, is 0.04. The empirical probability was calculated by knowing that a maximum of 5 out of the total 11 collisions had the other surrogate safety measure value below its threshold (i.e., if head-on collision, then THW was below 1.2 s, and if rear-end collision, then the TTC was below 1.5 s), and that the sample of size is given by the number of collisions plus the 60 observations where both TTC and THW were below 1.5 and 1.2, respectively. Therefore, the empirical probability takes a value within the range 0.0164 and 0.0769, which is in concordance with the estimated probability.

The probability density function of this bivariate distribution with GEV margins is displayed in Figure 3. This contour plot provides the confidence of the regions for the empirical data points (red dots), showing the suitability of this model to estimate the joint probability of colliding with the opposite vehicle or with the passed vehicle. As previously mentioned, further analysis of this approach should be performed where approaches such as extreme copulas [**31**] and the inclusion of other distributions for the margins should be explored.

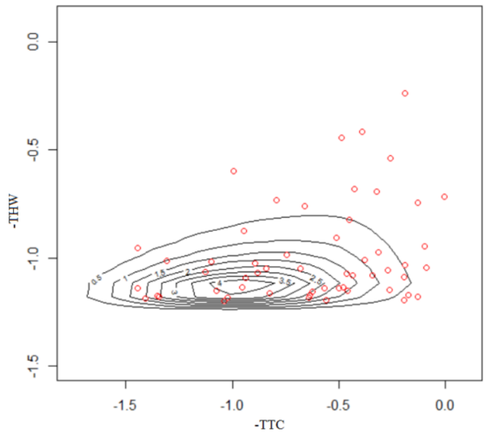

Figure 3. Probability Density contour plot for the Bivariate Distribution.

4. Conclusions

This paper analyzed the individual and joint probabilities of head-on collisions and rear-end collisions through the Block Maxima approach using the Univariate and Bivariate distributions with copula methods to model dependence between the two surrogate measures that capture those types of collisions during passing maneuvers.

The univariate non-stationary estimation allowed to conclude that aspects linked to drivers' characteristics, namely the gender, have a significant impact on the prediction of head-on collisions. However, these variables were not found to improve the prediction of rear-end collisions, where the stationary model seem to have a better fit. The bivariate model approach integrated the two different surrogate measures, TTC and THW, in order to estimate the risk of colliding with the opposite or with the passed vehicle in a single passing maneuver. Although the linear correlation between the two surrogate measures has proved to be weak, a bivariate distribution was estimated taking into account the existence dependence between these two surrogate measures by way of copula. This exploratory analysis is the first attempt to explain how two different surrogate measures are linked, providing guidelines to estimate the probability of colliding with the opposite or passed vehicles even in the presence of weak correlation.

To sustain the preliminary conclusions that both TTC and THW are good surrogate safety measures for near-accidents, head-on collisions and/or rear-end collisions, further analysis should be developed in order to validate through simulated data and/or data from other experimental scenarios the conclusions drawn by these models. This is the following work of the authors, together with the integration of these probabilities into a traffic microscopic simulation framework for safety assessment [**32, 33**].

References

1. J. A. Bonneson, *Highway safety manual*, Washington, DC: American Association of State Highway and Transportation Officials (2010).
2. J. Archer, *Methods for the assessment and prediction of traffic safety at urban intersections and their application in micro-simulation modelling*, Royal Institute of Technology (2004).
3. J. H. Kraay and A. Van der Horst, *The trautenfels study: A diagnosis of road safety using the dutch conflict observation technique doctor*, (1985).
4. C. Hydén, *The development of a method for traffic safety evaluation: The swedish traffic conflicts technique*, Bulletin Lund Institute of Technology, Department (1987), no. 70.

5. L. Zheng, K. Ismail and X. Meng, *Freeway safety estimation using extreme value theory approaches: A comparative study*, Accid. Anal. Prev. **62** (2014), 32-41.
6. P. Songchitruksa and A. P. Tarko, *The extreme value theory approach to safety estimation*, Accid. Anal. Prev. **38** (2006), no. 4, 811-822.
7. R. A. Fisher and L. H. C. Tippett, *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge Univ Press, 1928, pp. 180-190.
8. A. P. Tarko, *Use of crash surrogates and exceedance statistics to estimate road safety*, Accid. Anal. Prev. **45** (2012), 230-240.
9. J. K. Jonasson and H. Rootzén, *Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach*, Accid. Anal. Prev. **62** (2014), 102-109.
10. H. Farah, *Age and gender differences in overtaking maneuvers on two-lane rural highways*, Transportation Research Record: Journal of the Transportation Research Board (2011), no. 2248, 30-36.
11. C. Llorca, A. Garcia, A. T. Moreno and A. M. Perez-Zuriaga, *Influence of age, gender and delay on overtaking dynamics*, IET Intelligent Transport Systems **7** (2013), no. 2, 174-183.
12. H. Farah, A. Polus, S. Bekhor and T. Toledo, *Study of passing gap acceptance behavior using a driving simulator*, Advances in Transportation Studies an International Journal (2007), 9-16.
13. E. I. Vlahogianni and J. C. Golias, *Bayesian modeling of the microscopic traffic characteristics of overtaking in two-lane highways*, Transportation research part F: traffic psychology and behaviour **15** (2012), no. 3, 348-357.
14. S. Coles, J. Bawa, L. Trenner and P. Dorazio, *An introduction to statistical modeling of extreme values*, vol. 208, Springer2001.
15. C. L. Azevedo and H. Farah, *Using extreme value theory for the prediction of head-on collisions during passing maneuvres*, 2015 IEEE 18th International Conference on Intelligent Transportation Systems, IEEE, 2015, pp. 268-273.
16. H. Farah and C. L. Azevedo, *Safety analysis of passing maneuvers using extreme value theory*, IATSS Research (2016).
17. K.-T. Fang, S. Kotz and K. W. Ng, *Symmetric multivariate and related distributions*, Chapman and Hall1990.
18. R. B. Nelsen, *An introduction to copulas*, Springer Science & Business Media2007.
19. C. Genest and L.-P. Rivest, *Statistical inference procedures for bivariate archimedean copulas*, Journal of the American statistical Association **88** (1993), no. 423, 1034-1043.
20. D. Berg, *Copula goodness-of-fit testing: An overview and power comparison*, The European Journal of Finance **15** (2009), no. 7-8, 675-701.
21. H. Farah, S. Bekhor and A. Polus, *Risk evaluation by modeling of passing behavior on two-lane rural highways*, Accid. Anal. Prev. **41** (2009), no. 4, 887-894.
22. T. Rosenthal, *Stisim drive user's manual*, Systems Technology Inc., Hawthorne, CA (1999).
23. C. R. Hicks and K. Turner, "Fundamental concepts in the design of experiments. 1999," Oxford University Press, New York.
24. O. Taubman-Ben-Ari, M. Mikulincer and O. Gillath, *The multidimensional driving style inventory—scale construct and validation*, Accid. Anal. Prev. **36** (2004), no. 3, 323-332.
25. K. Vogel, *A comparison of headway and time to collision as safety indicators*, Accid. Anal. Prev. **35** (2003), no. 3, 427-433.
26. H. Farah and C. L. Azevedo, *Safety analysis of passing maneuvers using extreme value theory*, IATSS Research.
27. E. Gilleland and R. W. Katz, *New software to analyze how extremes change over time*, Eos, Transactions American Geophysical Union **92** (2011), no. 2, 13-14.
28. L. Evans and P. Wasielewski, *Do accident-involved drivers exhibit riskier everyday driving behavior?*, Accid. Anal. Prev. **14** (1982), no. 1, 57-64.
29. P. Embrechts, A. McNeil and D. Straumann, *Correlation and dependence in risk management: Properties and pitfalls*, Risk management: value at risk and beyond (2002), 176-223.
30. E. C. Brechmann and U. Schepsmeier, *Modeling dependence with c-and d-vine copulas: The r-package cdvine*, Journal of Statistical Software **52** (2013), no. 3, 1-27.
31. P. Capéraà, A.-L. Fougères and C. Genest, *A nonparametric estimation procedure for bivariate extreme value copulas*, Biometrika (1997), 567-577.
32. C. L. Azevedo, I. Cardoso and M. Ben-Akiva, *Probabilistic safety analysis using traffic microscopic simulation*, 94th Annual Meeting of the Transportation Research Board2014.
33. D. Gettman and L. Head, *Surrogate safety measures from traffic simulation models*, Transportation Research Record: Journal of the Transportation Research Board (2003), no. 1840, 104-115.