Using Extreme Value Theory For the Prediction of Head-On Collisions During Passing Maneuvres

Carlos Lima Azevedo  
Singapore-MIT Alliance for Research and Technology  
1 Create Way, 138602 Singapore  
Singapore, Singapore  
cami@smart.mit.edu

Haneen Farah  
Department of Transport and Planning  
Delft University of Technology  
Delft, The Netherlands  
h.farah@tudelft.nl

Abstract— This paper tests the Generalized Extreme Value (GEV) distribution as an EV method using the minimum time-to-collision with the opposing vehicle during passing maneuvers. Detailed trajectory data of the passing, passed and opposite vehicles from a fixed-based driving simulator experiment were used in this study. One hundred experienced drivers from different demographic strata participated in this experiment on a voluntary base. Raw data were collected at a resolution of 0.1 s and included the longitudinal and lateral position, speed and acceleration of all vehicles in the scenario. From this raw data, the minimum time-to-collision with the opposing vehicle at the end of the passing, maneuver was calculated. GEV distribution based on the Block Maxima approach was tested for the estimation of head-on collision probabilities in passing maneuvers. The estimation results achieved good fit with respect to head-on collisions’ prediction indicating that this is a promising approach for safety evaluation.

Keywords— Road Safety; Probabilistic Model; Extreme Value; Driving Behavior; Minimum Time-to-Collision

I. INTRODUCTION

The literature has frequently addressed the advantages of using surrogate safety measures over crash data [1], especially nowadays when advanced sensing technologies which facilitate the collection of detailed data on vehicles' trajectories are becoming readily available [2]. Crash data suffer from underreporting and frequently poor quality. Furthermore, the use of crash data to infer conclusions about the safety level, is a reactive approach while using surrogate safety measures is a proactive and time-efficient approach [3]. The use of aggregate crash data to develop safety models does not provide insights on the crash causations or details on the driver crash avoidance behavior. Thus, the use of surrogate safety measures for modeling and estimating safety is considered as a promising approach to achieve those targets. The developed surrogate safety measures can also be implemented in traffic simulation packages. Crashes are also infrequent, the ratio between conflicts and actual crash frequencies, according to [4], is 20000 to 1. Thus, there is a clear advantage of using surrogate safety measures over crash data. Reference [5] indicates that the validity of a surrogate safety measure is usually determined by its correlation with crash frequency which is usually assessed using regression analysis. For example, [6] found a statistically significant relationship between crashes and conflicts with an R² in the range of 0.70 - 0.77 at signalized junctions. However, the regression analysis still incorporates the use of crash counts which are known to suffer from availability and quality issues, and thus this approach is limited. Besides, it is difficult to insure the stability of the crash-to-surrogate ratio and this relationship also hardly reflects the physical nature of crash occurrence [5]. Reference [7] concluded that comprehensive and generalized answer to the question “are near-crashes representative for crashes?” may be less useful. Instead careful separate analyses for different types of situations are needed. Recently [8] developed a new and more sophisticated approach based on the Extreme Value (EV) theory to estimate the frequency of crashes based on measured crash proximity. The field of EV theory was pioneered by [9]. It is a commonly applied theory in many fields, such as in meteorology, hydrology, and finance [5]. However, [8] indicates that its application in the field of transportation engineering is still limited. According to [1] the EV approach has three considerable advantages over the traffic conflict technique in the detailed analysis of safety: (1) The EV theory abandons the assumption of a fixed coefficient converting the surrogate event frequency into the crash frequency; (2) the risk of a crash given the surrogate event is estimated for any condition based on the observed variability of crash proximity without using crash data; (3) the crash proximity measure precisely defines the surrogate event. This method has the potential to estimate the probability of extreme events from relatively short period of observations and it proposes a single dimension to measure the severity of surrogate events and to identify crashes. 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 [8]. In the context of road safety, the more observable traffic events are used to predict the less frequent crashes, which are often unobservable in a short time period [5]. More recently, [8] used an EV approach to build up relationships between occurrence of right-angle crashes 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 crash occurrence to the frequency of conflicts estimated from observed variability of crash proximity, using a probabilistic framework and without using crash records. The generic formulation of the application of EV to road safety analysis...
was then proposed by [2] and it was only very recently applied to other crash types and data sets [5,7].

In this study the time-to-collision or TTC [10] will be used as a surrogate safety measure of the risk to be involved in a head-on collision with the opposite vehicle while passing on two-lane rural highways, using the EV approach. According to [11] in 2012 there were 3,561 fatalities in the US as a result of a head-on collision. Not many studies have focused on the detailed analysis of the link between passing maneuvers and head-on-collisions. The TTC was previously used by [12] to evaluate the risk of passing behavior on two-lane rural highways. The authors defined the minimum TTC, as the remaining gap between the passing vehicle and the opposing vehicle at the end of the passing process. This measure expresses the risk involved in the passing maneuver. The authors developed a Tobit regression model that explains the minimum TTC. Traffic related explanatory variables were found to have the most important effect on the minimum TTC, but also the road geometric design and driver characteristics were also found to have a significant contribution. Other studies used as well the TTC as a measure for head-on conflicts [13,14].

There are two families of EV distributions which follows 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 [15], where all values above some high level are used. In this paper the BM approach only is examined.

II. RESEARCH METHOD - BLOCK MAXIMA (BM) APPROACH

In the GEV distribution the extreme events are sampled based on the block maxima (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]:

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

(1)

where \(X_1, X_2, \ldots, X_n\) is a set of independently and identically distributed random observations with unknown distribution function \(F(x) = \Pr(X_i \leq x)\), the maximum \(M_n = \max\{X_1, X_2, \ldots, X_n\}\) 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, \(\xi\), is positive, then this would yield the Frechet cdf with a finite lower endpoint, \((\mu - \sigma / \xi)\), if \(\xi\) 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.

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. This is how the minimum TTC will be handled in this study.

For the BM approach, and in the case that most blocks have enough observations, the r-largest order statistics is recommended, it enables the incorporation of more than one extreme from each interval in order to increase the confidence of parameter estimates. It is usually recommended to have at least a sample of 30 maxima (or minima). The size of the chosen interval should be large enough so that there are enough observations from which a maxima is chosen in which it is truly an extreme value, and small enough to provide a sample larger than 30. In the case of passing maneuvers, there is only one extreme in each interval, which is the TTC at the end of the passing maneuvers which we refer to here as the minimum TTC.

The model’s parameters were estimated using the maximum likelihood method (ML) in R (v3.0.3) using the “extTemes” and “evd” packages [16]. Details on the statistical properties of the GEV can be found in [17] and on the theoretical background of its applicability for surrogate safety analysis in [2].

III. LABORATORY EXPERIMENT

A laboratory experiment using a driving simulator previously developed by [12] for modelling drivers’ passing behavior on two-lane highways was used in order to collect data on the time-to-collision with the opposing vehicle. The simulator used in this experiment, STISIM [18], 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. The simulator updates the images at a rate of 30 frames per second. The situations that participants encountered were defined by the vehicles shown in Fig. 1. The subject vehicle is passing an impeding vehicle (front vehicle) while another vehicle is approaching from the opposite direction. This paper focuses on the minimum TTC surrogate safety measure while passing on two-lane rural highways. Mathematically, the TTC is calculated by the division of the distance between the fronts of the subject vehicle and the opposing vehicle by the sum of their speeds. The minimum TTC is the TTC value at the end of a successful passing maneuver.

![Fig. 1. TTC with the Opposing Vehicle.](image-url)

To understand how various infrastructure and traffic factors affect the TTC when passing, a number of simulator scenarios were designed. Each scenario included 7.5 km of two-lane rural highway section with no intersections. The road sections were on a level terrain and with daytime and good weather conditions, which allowed good visibility. However, each scenario design varied according to four main factors of two
levels each. The choice of these factors was based on previous studies that showed their impact on passing decisions. Two levels were used for each factor. These factors are: the speed of the front vehicle (60 or 80 km/h); the speed of the opposite vehicle (65 or 85 km/h); opposite lane traffic volume (200 or 400 veh/h); and road curvature (300-400 m or 1500-2500 m). This produces 2^5 = 16 different scenarios. The partial confounding method [19] was used to allocate for each driver 4 scenarios out of the 16 scenarios. Detailed information on this experiment can be found in [12].

A. Participants

One hundred drivers (64 males and 36 females) with at least 5 years of driving experience participated in the driving simulator experiment on a voluntary base. The drivers' age ranged between 22 and 70 years old. Drivers were instructed to drive as they would normally do in the real world. An advertisement on the experiment was published at the Technion campus in Israel and drivers who were interested to participate contacted the researchers.

B. The data

The data set from the driving simulator experiment resulted in 1287 completed passing maneuvers, in which 9 ended with a front-front collision (these observations were removed from the estimation data sets). TABLE I presents a summary statistics of passing maneuvers related variables. Passing gaps were defined as the gap between two successive opposite vehicles at the time the lead vehicle on the opposite lane is at the same line with the subject vehicle. The passing duration is measured from the moment the subject vehicle left-front wheel crosses the center line (as shown in Fig. 1) until the passing maneuver ends when the rear-left wheel crosses the centerline. Vehicles’ speeds as summarized in TABLE I are measured at the beginning of the passing maneuvers. The following distance (m) from front vehicle when starting to pass is measured as the distance between the front of the subject vehicle and the end of the front vehicle as illustrated in Fig. 1. Finally, the minimum TTC and the gap from passed front vehicle are both measured at the end of the passing maneuver and reflect the risk to collide with the opposing vehicle, and the front vehicle, respectively.

IV. RESULTS AND ANALYSIS

This section presents the results of the analysis following the research method described above. First, the estimation results of the stationary BM using the GEV model is presented, followed by the non-stationary BM estimation results.

A. Stationary Block Maxima Approach (BM) Results

A Generalized Extreme Value (GEV) distribution is fitted to the 1287 passing maneuvers and the respective minimum TTC measurements. For the block intervals we use the annotated time that contain the entire passing maneuver. Both the chosen block interval and the resulting number of observations in each block are variable. In this case, the calculated probability represents the probability of a head-on collision for a single passing maneuver. Furthermore, past studies concluded that with minimum TTC smaller than a low limit (typically, 1 to 1.5 s) are useful as crash surrogates [7, 20]. As a first test, the filtered data according to this approach, and choosing a limit of 1.5 s, resulted in 463 maxima. Fig. 2 (top) presents the Cumulative Distribution Function (CDF) of the minimum TTC (min[TTC]) for the full data set, while Fig. 2 (bottom) presents the CDF of the min[TTC] for the filtered data. For the full data set, 50% of the observations were less than a TTC of about 2 s, while in the filtered data, 50% of the observations were less than a TTC of about 0.9 s.

<table>
<thead>
<tr>
<th>Variable</th>
<th>mean</th>
<th>median</th>
<th>15th percentile</th>
<th>85th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted passing gap (s)</td>
<td>21.47</td>
<td>20.75</td>
<td>17.39</td>
<td>28.79</td>
</tr>
<tr>
<td>Passing duration (s)</td>
<td>4.98</td>
<td>4.83</td>
<td>3.50</td>
<td>6.48</td>
</tr>
<tr>
<td>Passing vehicle speed (m/s)</td>
<td>22.21</td>
<td>21.29</td>
<td>17.27</td>
<td>27.39</td>
</tr>
<tr>
<td>Front vehicle speed (m/s)</td>
<td>66.20</td>
<td>60.00</td>
<td>60.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Opposing vehicle speed (m/s)</td>
<td>76.28</td>
<td>85.00</td>
<td>65.00</td>
<td>85.00</td>
</tr>
<tr>
<td>Following distance from front vehicle when starting to pass (m)</td>
<td>15.47</td>
<td>12.80</td>
<td>8.39</td>
<td>22.92</td>
</tr>
<tr>
<td>Minimum TTC (s)</td>
<td>2.37</td>
<td>1.98</td>
<td>0.76</td>
<td>4.10</td>
</tr>
<tr>
<td>Gap from passed front vehicle at end of the passing maneuver (s)</td>
<td>2.44</td>
<td>2.24</td>
<td>1.49</td>
<td>3.42</td>
</tr>
</tbody>
</table>

Fig. 2. CDF of minTTC (s) for the full dataset (top) and filtered data (bottom).
We estimate a stationary block maxima model for the maxima of the negated values instead of minima of the original values, i.e. max{-TTC}. The fitted distribution resulted in the following parameters of the GEV cumulative distribution function: \(\hat{\mu} = -0.993 \pm 0.0212\), \(\hat{\sigma} = 0.383 \pm 0.0163\) and \(\hat{\xi} = 0.236 \pm 0.0500\). Fig. 3. (top) presents the probability density function of the empirical and modeled negated TTC, and Fig. 3. (bottom) presents the simulated QQ plot. From these figures it can be concluded that the modeled GEV distribution has satisfactory fitting results to the empirical data since the points fall close to the 45° line in the simulated QQ plot.

With this stationary model using the fitted GEV distribution, the estimated probability of max{-TTC} ≥ 0 is 0.0179 with 95% confidence interval (0.0177, 0.0182). The confidence intervals of estimations were computed assuming the normal distribution under regularity conditions of the parameters, a simulation experiment size of 1×10^6 and its simulated distribution quantiles. Out of the 463 near head-on collisions in the driving simulator (using the threshold of 1.5 s), 9 maneuvers ended with actual collisions. In other words, the probability for a head-on collision assuming a near head-on collision in a passing maneuver is 9/463 = 0.0194, with a 95% binomial confidence interval (0.00893, 0.0366). This value is comparable to the estimate resulting from the fitted GEV distribution.

However, the process of passing maneuver may be affected by the detailed conditions of each specific passing, such as the relative gaps and speeds between the vehicles surrounding the subject vehicle. To account for the fact that the TTCs at the end of the passing maneuvers are non-stationary observations and are affected by several factors, we tested the inclusion of different covariates that were collected during the driving simulation experiment in the location parameter of the BM model.

### B. Non-Stationary Block Maxima Approach (BM) Results

Several linear combinations of these variables were tested during model estimation task. To test reduced model structures and the inclusion of variables, the likelihood ratio test was used [17]. The final model was also tested against the stationary one, resulting in a p-value (3.741×10^{-8}) significantly smaller than \(\alpha = 0.05\).

The results in TABLE III indicate that as the speed of the front vehicle increases the negated TTC increases, and the TTC decreases which is logical since it is easier for the subject vehicle to pass the front vehicle. This is also according to the conclusions reached in previous studies [12, 21]. Similarly, as the accepted passing gap is larger, the negated TTC decreases, and the TTC increases. On the other hand, as drivers start their passing maneuver from a larger gap from the front vehicle, the negated TTC increases and the TTC decreases. This is because it take drivers longer to pass the front vehicle, and during this time the opposing vehicle has become closer, resulting in a shorter TTC. The road design as well impacts the TTC. As the road curvature is higher, the negated TTC is lower, and the TTC is higher.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>passinggap</td>
<td>The time gap between two opposite vehicles at the time the subject meet the lead opposite vehicle (s)</td>
</tr>
<tr>
<td>speedopposing</td>
<td>The speed of the opposite vehicle at the moment of start passing (m/s)</td>
</tr>
<tr>
<td>speedfront</td>
<td>The speed of the front vehicle at the moment of start passing (m/s)</td>
</tr>
<tr>
<td>tailgatetp</td>
<td>The time gap between the subject vehicle and the front vehicle at the moment of start passing (s)</td>
</tr>
<tr>
<td>passduration</td>
<td>The passing duration (s)</td>
</tr>
<tr>
<td>curvature</td>
<td>The road curvature (1/m)</td>
</tr>
</tbody>
</table>

### TABLE III. ESTIMATION RESULTS OF BEST MODELS FOR BM APPROACH

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated value</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\mu})</td>
<td>-1.06</td>
<td>0.139</td>
</tr>
<tr>
<td>(\hat{\mu}_0)</td>
<td>0.0245</td>
<td>0.00644</td>
</tr>
<tr>
<td>(\hat{\mu}_1) (speedfront)</td>
<td>0.00274</td>
<td>0.00179</td>
</tr>
<tr>
<td>(\hat{\mu}_2) (tailgatetp)</td>
<td>-0.0212</td>
<td>0.00445</td>
</tr>
<tr>
<td>(\hat{\mu}_3) (passinggap)</td>
<td>-38.1</td>
<td>13.5</td>
</tr>
<tr>
<td>(\hat{\sigma}) (curvature)</td>
<td>0.369</td>
<td>0.0145</td>
</tr>
<tr>
<td>(\hat{\xi})</td>
<td>-0.225</td>
<td>0.0412</td>
</tr>
</tbody>
</table>

| N | 463 |
| Neg. loglikelihood | 215.54 |
This might indicate of an adaptation behavior by drivers who compensate for the difficulty of the passing maneuver on complex roads by increasing their safety margin. Previous results using the same dataset of this study by [22] found that on roads with sharper curvature, drivers accept larger critical gaps, which supports the results and conclusion of this study. The speed of the opposing vehicle was not found to be significant at the 95% confidence level, however, this variable is indirectly included through the passing gap which is measured in time.

Fig. 4 (top) presents the probability density function of the empirical and modeled standardized maximum negated TTC, and Fig. 4 (bottom) presents the simulated QQ plot for the non-stationary model. The results indicate a good fit between the modeled GEV distribution and the empirical data, and a better fit compared to the results of the stationary model presented in Fig. 3. Also, the negative log-likelihood has improved from 229.5 to 215.5, maintaining a $\xi > -0.5$ that assures the regular asymptotic properties of the maximum likelihood estimators [17].

To estimate the probability of a head-on-collision during the passing maneuver ($\max \{-\text{TTC}\} \geq 0$) for the non-stationary model, simulated covariates or directly location parameters have to be generated. From the estimated location parameters for the estimation dataset, a normal distribution was fitted with satisfactory results with mean of -0.996, standard deviation of 0.115 and a Kolmogorov-Smirnov test statistic of 0.0452. The simulated probability of $\max \{-\text{TTC}\} \geq 0$ is 0.0190 with 95% confidence interval (0.0188,0.0193), resulting in a better estimate than the stationary model.

V. SUMMARY AND CONCLUSIONS

In this on-going study an Extreme Value (EV) approach was applied for the estimation of the probability of head-on collisions that result from unsuccessful passing maneuvers on two-lane rural highways. The Block Maxima (BM) approach using the Generalized Extreme Value (GEV) distribution was tested using the minimum time-to-collision with the opposing vehicle during passing maneuvers.

The estimation showed that the BM approach yielded satisfying results and that the non-stationary BM model performed better than the stationary BM model. This is according to expectation since the introduced covariates significantly affect the TTC and were found to be important explanatory variables in previous studies [12, 21]. Furthermore, the predicted probability of head-on collisions based on the BM approach was sufficiently close to the probability of head-on collisions based on the empirical data from the driving simulator. This also indicates that for passing maneuvers the TTC is a good surrogate safety measure for near-crashes of head-on collisions. This is different from the conclusion reached by [7] who found severe discrepancy between the rear-striking near-crashes (using the TTC) and rear-striking crashes. However, this can be explained by the mechanism of crash occurrence and the state of the driver. In passing maneuvers drivers are completely aware and conscious of their actions and therefore head-on collisions usually result from an error in drivers’ judgment of the suitability of the passing gap. On the other hand, in rear-striking collisions, the state of the driver in these collisions might vary a lot. It can result, similarly to passing collisions, from drivers’ errors in judging their gap and speed from the front vehicle but can also result from the driver being distracted. In the first case, it is most likely to observe an evasive action of the driver to prevent the collision but in the second case no evasive action will be observed. This causes, as [7] indicate, a selection bias, and therefore, careful selection of near-crashes is a crucial issue in preventing this from occurring.

Despite these promising results, future research by the authors will attempt to expand this work in several possible directions as follows: (1) test and compare the results with the EV using the Peak-Over Threshold approach; (2) testing alternative surrogate measures of head-on collisions such as the Time Exposed Time to Collision or Time integrated Time to Collision [25]; (3) developing a more sophisticated measure of

\begin{equation}
\text{max} \{-\text{TTC}\} \geq 0 \text{ is } 0.0190 \text{ with 95% confidence interval (0.0188,0.0193), resulting in a better estimate than the stationary model.}
\end{equation}
risk which accounts for the complexity of the passing maneuver and considers the probability to collide not only with the opposite vehicle but also with the passed vehicle. One possibility is, similarly to [7], to use a bivariate GEV which is built from two components of the Block Maxima vectors and which considers the TTC and the headway between the passing and passed vehicle at the end of the passing maneuver; (4) extending the non-stationary models by including other covariates related to road design (this study accounted for only the road curvature) and drivers’ characteristics, such as socio-demographic, driving styles and cultural differences; (5) examining the transferability of such model and validation of the results with other datasets especially from field studies; (6) applying the developed models in traffic microscopic simulation environments for safety assessment [26].

REFERENCES