# TUDelft

Delft University of Technology

## Are collision and crossing course surrogate safety indicators transferable?
## A probability based approach using extreme value theory

Borsos, Attila; Farah, Haneen; Laureshyn, Aliaksei; Hagenzieker, Marjan

**Important note**
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

# Are collision and crossing course surrogate safety indicators transferable? A probability based approach using extreme value theory

Attila Borsos[a,*], Haneen Farah[b], Aliaksei Laureshyn[c], Marjan Hagenzieker[b]

[a] *University of Győr, Department of Transport Infrastructure and Water Resources Engineering, Egyetem tér 1, Győr 9026, Hungary*
[b] *Delft University of Technology, Department of Transport and Planning, Stevinweg 1, 2628 CN Delft, Netherlands*
[c] *Lund University, Department of Technology and Society, John Ericssons vag 1, 1 TR Lund, Sweden*

ABSTRACT

In order to overcome the shortcomings of crash data a number of surrogate measures of safety have been developed and proposed by various researchers. One of the most widely used temporal indicators is time-to-collision (TTC) which requires the road users to be on a collision course. Road users that are strictly speaking not on a collision course actually might behave and take evasive actions as if they were, thus indicating that such near-miss situations might also be relevant for safety analysis. Taking that into account, a more flexible indicator $T_2$, which does not require the two vehicles to be on a collision course, describes the expected time for the second road user to arrive at the conflict point.

Recently extreme value theory (EVT) offering two approaches, block maxima (BM) and Peak over Threshold (POT), has been applied in combination with surrogate indicators to estimate crash probabilities. Most of this research has focused on testing BM and POT as well as validating various surrogate safety indicators by comparing model estimates to actual crash frequencies. The comparison of collision course indicators with indicators including crossing course interactions and their performance using EVT has not been investigated yet. In this study we are seeking answers to under what conditions these indicators perform better and whether they are transferable. Using data gathered at a signalized intersection focusing on left-turning and straight moving vehicle interactions our analysis concluded that the two indicators are transferable with stricter threshold values for $T_2$ and that POT gives more reasonable results.

## 1. Introduction

In this section we provide a literature review on surrogate measures of safety focusing on two particular indicators time-to-collision (TTC) and $T_2$, give an introduction to extreme value theory as well as specify the research gap and questions.

### 1.1. Role of surrogate measures of safety

To improve traffic safety and to make sure that it is done in an efficient way, one has to be able to quantify safety in order to support evidence-based policy making. The most plausible way to evaluate safety is investigating the occurrence and severity of crashes using historical data. This approach however has a number of limitations (Tarko et al., 2009): accidents are rare events (Hauer, 1997) associated with the random variation inherent in small numbers (Svensson and

Hydén, 2006), at least 3 years of observations are needed (Nicholson, 1985) and thus safety analysis based on crash records is a reactive approach, accident records are prone to underreporting, and finally data quality is not always sufficient.

In order to overcome the above limitations the use of non-crash events have gained a lot of attention especially due to the rapid improvement of sensing technologies facilitating the collection of trajectory data. Already 30 years ago Hydén (1987) pointed out that the interaction between road users can be described as a continuum of safety related events. In Hyden's safety pyramid crashes as the rarest events are followed by conflicts of different levels of severity (serious, slight and potential). These are critical events that do not result in a crash but are very close to that and can be used as surrogate safety measures (Tarko et al., 2009). Below the conflicts the majority of events are undisturbed passages or normal traffic processes (Laureshyn et al., 2010). This pyramid also shows how few and exceptional accidents are

---

that we usually base our safety estimates on Svensson and Hydén (2006).

With regard to the shape of the severity hierarchy Svensson (1998) made an important suggestion that it is not necessarily a pyramid. She proposed a diamond-shape based on the frequency of pedestrian–vehicle conflicts observed at signalized and unsignalized intersections. The idea behind the diamond shape is that at a particular site the majority of the interactions will be of moderate severity. Tarko (2012) also noted that this was the first evidence proving that there is a heterogeneity in the frequency-severity relationship due to the type of road facility influencing traffic conflicts. Other conditions, such as vehicle type, road users, collision angle and speed (Laureshyn et al., 2010) but also weather may affect this relationship.

It has to be noted, however, that Svensson limited the events in the hierarchy only to interactions with a collision course (Svensson, 1998; Svensson and Hydén, 2006). A very important implication of this is that even low-severity interactions should be utilized because they may carry useful safety information (Tarko, 2012). This statement is highly relevant as interactions with severe conflicts usually come with low frequencies.

In an attempt to apply Svensson's reasoning later on a few researchers tried to adapt it by broadening the concept of traditional approaches. In Canada St-Aubin et al. (2015) for instance developed an approach called probabilistic surrogate measures of safety (PSMS) with a more general framework for safety analysis considering all possible paths that may lead two road users to collide. The novelty of this approach is relaxing the traffic conflict by allowing a non-zero risk of collision for road users who are not on a collision course (Tarko, 2012).

Following the same reasoning Laureshyn et al. (2010) suggested a new indicator called $T_2$ broadening the concept of the most common nearness-to-collision surrogate measure of safety, time-to-collision (TTC). TTC can be calculated for any moment as long as the road users are on a collision course and defined as "the time until a collision between the vehicles would occur if they continued on their present course at the present rates" (Hayward, 1972). The lowest TTC value during the interaction, abbreviated as $TTC_{min}$, is the most commonly used indicator. The supplementary indicator proposed by Laureshyn et al. (2010) measures the expected time that it takes for the second road user to arrive at the potential collision point, hence it is called $T_2$. The logic behind this indicator is that TTC assumes the two road users to be on a collision course, which however sets a limitation to the situations to be considered in safety analysis. Laureshyn et al. (2017) argued that encounters without a collision course might have crash potential as well due to the possibility of minor changes in the spatial or temporal relationship between road users.

$T_2$ tells more about safety since the arrival at the potential collision point is the very last necessary condition for a collision to occur and it provides a smooth transfer between the collision course and crossing course situations (Laureshyn et al., 2017). $T_2$ assumes unchanged speeds and planned trajectories. If the road users are on a collision course, $T_2$ equals TTC. In the event that the two road users pass the conflict point with a time margin, $T_2$ reflects the maximum time available to take evasive actions and alleviate the severity of the situation. $T_2$ is no longer calculated after the first road user has left the conflict zone (since the crash is no longer possible) (Laureshyn et al., 2017). $T_2$ is a similar indicator to TTC in the sense that it is also continuous, therefore can be calculated for any time instance. The last possible value is when the first road user leaves the potential conflict area (the same as post encroachment time – PET). An alternative value is $T_{2min}$ which shows the moment when the two vehicles are closest in time. These two values can be different in case of significant speed changes.

Researchers testing the validity of traffic conflicts have tried to link historical crash data with conflict frequencies. These analyses lead to inconclusive results, as some studies could confirm a relationship, some could not.

Tarek and Sany (1999) for instance arrived at the conclusion that there is statistically significant relationship between crashes and conflicts. They identified a determination coefficient ($R^2$) in the range of 0.70–0.77 at signalized intersections. Notwithstanding, this approach is still hampered by the fact that accident data are inaccurate, thus finding a good correlation has a limited power. Zheng et al. (2014) also emphasized that the application of regression models is limited due to three reasons:

- the incorporation of crash counts suffers from the same quality issues as traditional road safety analysis;
- the stability of crash-to-surrogate ratio is difficult to ensure especially when mixing surrogates of varied severity levels;
- the statistical relationship between counts of crashes and surrogates hardly reflects the physical nature of crash occurrence.

An alternative approach to the traditional regression analysis without using observed crash counts was first proposed by Songchitruksa and Tarko (2006) based on the extreme value theory (EVT). An important feature of the EVT is that it enables the researcher to model the stochastic behavior of unusually large or small processes. This extreme behavior is typically very rare and unobservable within a reasonable data collection time period. It often involves estimating the probability of extreme events over an extended period of time given very short and limited historical data (Songchitruksa and Tarko, 2006). The key assumption of EVT is that the underlying stochastic behavior of the process being modeled is sufficiently smooth to enable extrapolations to unobserved levels (Coles, 2001). A general introduction to EVT can be found in the next subsection.

### 1.2. Extreme value theory

Extreme value theory offers two approaches to sample extreme events, in this case near-crashes, the block maxima (BM) (or minima) using Generalized extreme value distribution (GEV) and the Peak over Threshold (POT) using Generalized Pareto distribution (GPD). In the former case the method divides the sample time into blocks of a certain length and samples the largest value (or $r$ largest values) in each block, whereas in the latter case all peak values are sampled and the values over a certain threshold are used to model the extremes.

EVT models based on the block maxima approach focus on the behavior of

$$M_n = \max\{X_1, \ ...,X_n\}, \tag{1}$$

where $X_1,...,X_n$ is a sequence of independent random variables having a common distribution function $F$, $M_n$ represents the maximum of the process over $n$ time units of observation. The distribution of $M_n$ can be derived as $Pr\{M_n \leqslant z\} = \{F(z)\}^n$. The function of $F$ is unknown and to look for $F^n$ a similar approach to the central limit theorem can be used, by allowing a linear renormalization of the variable $M_n$ (Eq. (2)):

$$M_n^* = \frac{M_n - b_n}{a_n}, \tag{2}$$

where $\{a_n > 0\}$ and $\{b_n\}$ are constants for which the appropriate values have to be found.

According to the extremal types theorem

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leqslant z\right\} \to G(z) \text{ as } n \to \infty \ , \tag{3}$$

where $G$ belongs to one of the three families: Gumbel, Frechet or Weibull.

The rescaled sample maxima $M_n^*$ converge to a variable having a distribution within one of the above three families. All the three types have both a location ($b$) and a scale ($a$) parameter. The Frechet and Weibull distributions also have a shape ($\alpha$) parameter. These distributions can be generalized into a single distribution function (Eq. (4)):

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

(4)

defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The three parameters that have been already mentioned before are the location parameter ($\mu$), the scale parameter ($\sigma$), and the shape parameter ($\xi$). The distribution function itself determines the value of the shape parameter and vice versa. If $\xi > 0$, the model corresponds to a Frechet distribution; if $\xi < 0$, a Weibull distribution; and if $\xi = 0$, a Gumbel distribution.

Block maxima is criticized to be a wasteful approach as only the maximum value is used from each block, thus not considering other, but possibly still extreme values. Possible solutions to solve this issue is using the so-called r largest order statistic model (e.g. using the largest 5 observations) or by modeling threshold excesses. The latter one is the Peak over Threshold approach, in which observations over a certain threshold are selected and treated as extremes.

Using the GEV distribution for large enough threshold $u$, the distribution function of $(X - u)$, conditional on $X > u$ (Eq. (5)), is approximately

$$H(x) = 1 - \left[1 + \xi\left(\frac{x - u}{\sigma_u}\right)\right]^{-1/\xi},$$

(5)

where $u$ is a high threshold, $x > u$, scale parameter $\sigma_u > 0$ (depending on threshold $u$), and shape parameter $-\infty < \xi < \infty$.

The distribution family given in Eq. (5) is called the generalized Pareto family, in other words, threshold excesses have a generalized Pareto distribution (GPD) with two parameters, the shape $\xi$ and the scale $\sigma$ parameters (using the same notation as in GEV). Just like with GEV, the shape parameter $\xi$ determines the behavior of the GPD. If $\xi < 0$ the distribution has an upper bound of $u - \sigma/\xi$; if $\xi > 0$ there is no upper limit. If $\xi = 0$, then Eq. (5) simplifies to an exponential distribution function.

### 1.3. Research gap and questions

Extreme value theory is a promising tool to evaluate safety using surrogate safety measures. Most of the research that has been done so far focused on testing the method and validating various surrogate safety indicators by comparing model estimates to actual crash frequencies (Songchitruksa and Tarko, 2006; Farah and Azevedo, 2015; Zheng et al., 2014; Jonasson and Rootzén, 2014; Cavadas et al., 2017; Åsljung et al., 2016; Wang et al., 2018). However, less or no attention was paid to the comparison of various conflict indicators and their performance using EVT, especially the comparison of collision course indicators with indicators including crossing course interactions as well. In this research we investigated under what conditions these indicators perform better and whether they are transferable. To that end two research questions were formulated as follows:

1. What difference is there between the two indicators $TTC_{min}$ and $T_{2min}$ when analyzing safety using EVT and are these indicators transferable?
2. Which EVT approach (BM or POT) under what circumstances performs better for $TTC_{min}$ and $T_{2min}$ (e.g. sensitivity to sample size)?

### 2. Data

In this section data collection is briefly described as well as descriptive statistics is provided.

### 2.1. Data collection

A regular signalized intersection with two-phases in Minsk (Belarus) was analyzed (53°54′39. 1″ N; 27°35′44. 4″ E). The intersection was

recorded for two days (from 6 AM till 9 PM). The video footages of two cameras set on rooftops were then analyzed in the software T-Analyst (2016) allowing the manual tracking of vehicles as well as the calculation of various surrogate measures of safety such as TTC, $T_2$, PET. The dataset was provided by Lund University and has already been used in other publications such as in Laureshyn et al. (2017).

Accident data were gathered for 11 years (1999–2009) before the video recordings were made. Altogether 32 accidents were recorded, out of which 5 were due to the collision of left turning and straight going vehicles. The severity of all the recorded accidents were property damage only. As this type of severity is the most heavily prone to underreporting, unfortunately this historical accident dataset can be used for validation with certain limitations. In the course of video recordings no accidents were observed.

### 2.2. Descriptive statistics

Altogether 2749 interactions were detected. A subset of situations involving a vehicle making a left turn in front of an oncoming vehicle from the opposite direction was created ($n = 792$). Whenever an indicator cannot be calculated the software indicates $-1$ as entry (e.g. if the two vehicles are not on a collision course there is no value for $TTC_{min}$). These entries were not considered when compiling the descriptive statistics. All the statistical analyses were done in R (R Core Team, 2013).

Table 1 shows the descriptive statistics for the two indicators $TTC_{min}$ and $T_{2min}$.

Cumulative distribution functions of $T_{2min}$ and $TTC_{min}$ are shown in Fig. 1 for values smaller than 20 s. This figure reveals that the cumulative distribution function for $TTC_{min}$ is less steeper than that of $T_{2min}$ showing that the observed $TTC_{min}$ values are more spread out and that the share of observations in the lower range (between 0 and 5 s) is smaller than for $T_{2min}$.

The underrepresentation of $TTC_{min}$ compared to $T_{2min}$ is due to the nature of these indicators, as $TTC_{min}$ can be measured only for collision course interactions, whereas $T_{2min}$ can be measured for both collision as well as crossing course interactions.

### 3. Models and results

In this study both EVT approaches are applied to the above presented dataset. As for the block maxima approach each interaction can be considered as a block in which the minimum values of $T_2$ and $TTC$ are used. In the Peak over Threshold approach a threshold has to be selected over which all the values are considered. In order to study the extreme events in both cases the negated values of observations are used (i.e. the minima instead of maxima).

### 3.1. Block maxima approach

Since the minimum values are determined per interaction for both indicators, they can also be high and therefore irrelevant occurrences (e.g. a $TTC_{min}$ value of 10 seconds cannot be considered as a near-crash, hence an extreme value). Therefore a preliminary step of selecting the

**Table 1**
Descriptive statistics.

| Indicator | $TTC_{min}$ | $T_{2min}$ |
|---|---|---|
| Sample size | 194 | 792 |
| Min | 0.79 | 0.06 |
| Max | 182.50 | 35.12 |
| Mean | 6.45 | 3.61 |
| Stdev | 14.25 | 2.22 |
| Skewness | 10.83 | 5.49 |
| Kurtosis | 123.76 | 61.25 |

## CDF of T2min (<20s)
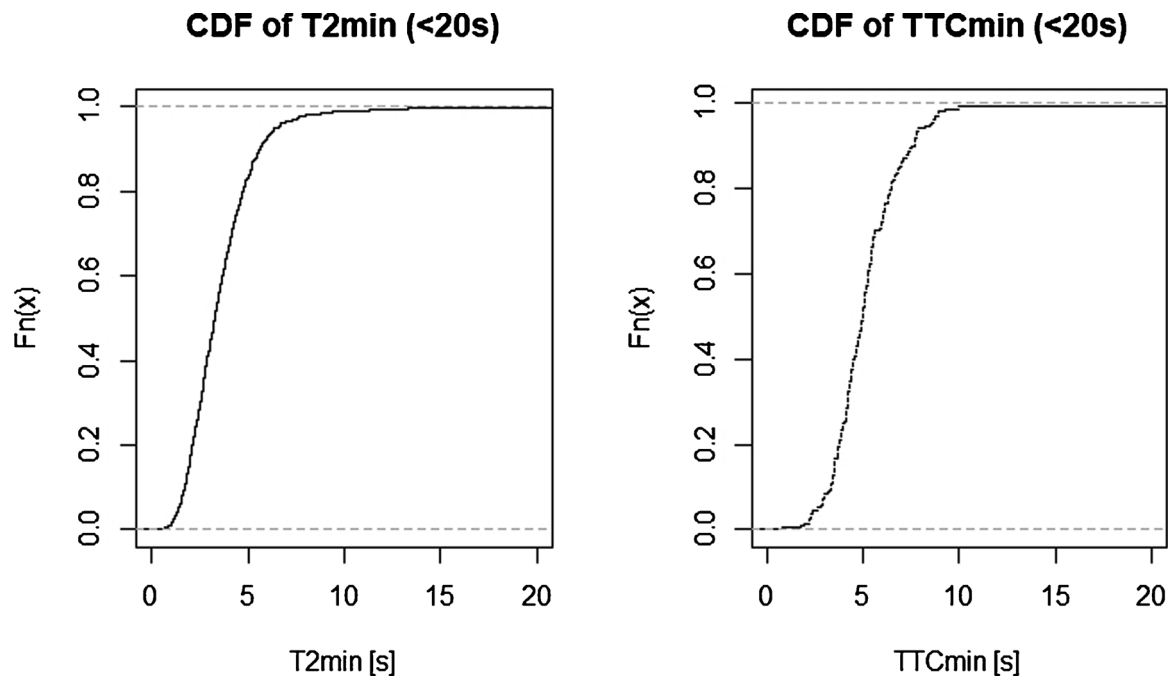


## CDF of TTCmin (<20s)



**Fig. 1.** Cumulative distribution functions of $T_{2min}$ and $TTC_{min}$ ($< 20\,s$).

near-crash events is needed, which can be considered as "subsampling of maxima" (Jonasson and Rootzén, 2014).

Mahmud et al. (2017) gave an overview of minimum and desirable TTC threshold values from a selection of studies for different conditions. As far as signalized intersections are concerned Mahmud et al. (2017) did not indicate any minimum values, however he cited two references (Huang et al., 2013; Sayed et al., 2013) where desired values of 1.6 and 3 s were given.

Taking 3 s as a threshold value for near-crashes would only result in 15 observations for $TTC_{min}$. Based on what the literature suggests and the observed values near-crashes were selected using a threshold value of 3.5 s for the first run. The above problem does not hold for $T_{2min}$ thanks to its bigger sample size, but for the sake of comparability the same threshold value was applied for the first run. The results of these two model runs were evaluated in detail and followed by several other runs using different threshold values for near-crash situations.

Table 2 gives a summary of the model results of the fitted GEV models.

As for a value of $TTC_{min} < 3.5\,s$ the 95% confidence intervals of the shape parameter does not include zero, thus we can accept the Frechet distribution as the shape parameter is greater than zero. Notwithstanding, a greater accuracy for the confidence intervals is usually attained by the profile likelihood, which yielded similar results.

Substituting the model estimates into the GEV function (Eq. (4)) using a given value for $z$ one can calculate its probability. We are interested in the probability of crash occurrence, that is, when $TTC_{min} < 0$ ($z = 0$ in Eq. (4)). As for a near-crash value of $TTC_{min} < 3.5\,s$ this calculation yields the probability of 0.0733 $(1 - G(z))$.

Using a given return level $z$ one can also obtain the return period, which is $1/p$. This means that the level $z$ is expected to be exceeded on average once every $1/p$. If each block corresponds to one year, then the

**Table 2**
Results of GEV for $TTC_{min}$ and $T_{2min}$ with different thresholds for near-crash situations.

| Indicator | $TTC_{min} < 3.5\,s$ | $TTC_{min} < 4\,s$ | $TTC_{min} < 4.5\,s$ | $TTC_{min} < 5\,s$ |
|---|---|---|---|---|
| Sample size | 31 | 48 | 76 | 100 |
| Location parameter $\mu$ | −3.336 | −3.552 | −3.977 | −4.277 |
| Scale parameter $\sigma$ | 0.230 | 0.378 | 0.466 | 0.605 |
| Shape parameter $\xi$ | 1.099 | 0.327 | 0.244 | 0.087 |
| Shape p. lower bound of conf. int. 95% | 0.193 | 0.010 | −0.024 | −0.109 |
| Shape p. upper bound of conf. int. 95% | 2.004 | 0.644 | 0.513 | 0.284 |
| Probability of crash $TTC_{min} < 0$ | 0.073 | 0.014 | 0.010 | 0.004 |
| Return period $TTC_{min} < 0$ | 13.65 | 73.92 | 101.04 | 246.57 |
| Kolmogorov–Smirnov test $p$-value | 0.826 | 1 | 0.974 | 0.994 |

| Indicator | $T_{2min} < 3.5\,s$ | $T_{2min} < 3\,s$ | $T_{2min} < 2.5\,s$ | $T_{2min} < 2\,s$ |
|---|---|---|---|---|
| Sample size | 443 | 341 | 232 | 130 |
| Location parameter $\mu$ | −2.674 | −2.382 | −2.050 | −1.712 |
| Scale parameter $\sigma$ | 0.615 | 0.473 | 0.352 | 0.246 |
| Shape parameter $\xi$ | −0.130 | −0.045 | 0.026 | 0.166 |
| Shape p. lower bound of conf. int. 95% | −0.211 | −0.150 | −0.098 | −0.031 |
| Shape p. upper bound of conf. int. 95% | −0.047 | 0.059 | 0.150 | 0.364 |
| Probability of crash $T_{2min} < 0$ | 0.002 | 0.003 | 0.004 | 0.010 |
| Return period $T_{2min} < 0$ | 596.02 | 302.40 | 226.10 | 101.96 |
| Kolmogorov–Smirnov test $p$-value | 0.589 | 0.537 | 0.982 | 0.992 |

## Probability Plot

## Quantile Plot

## Return Level Plot
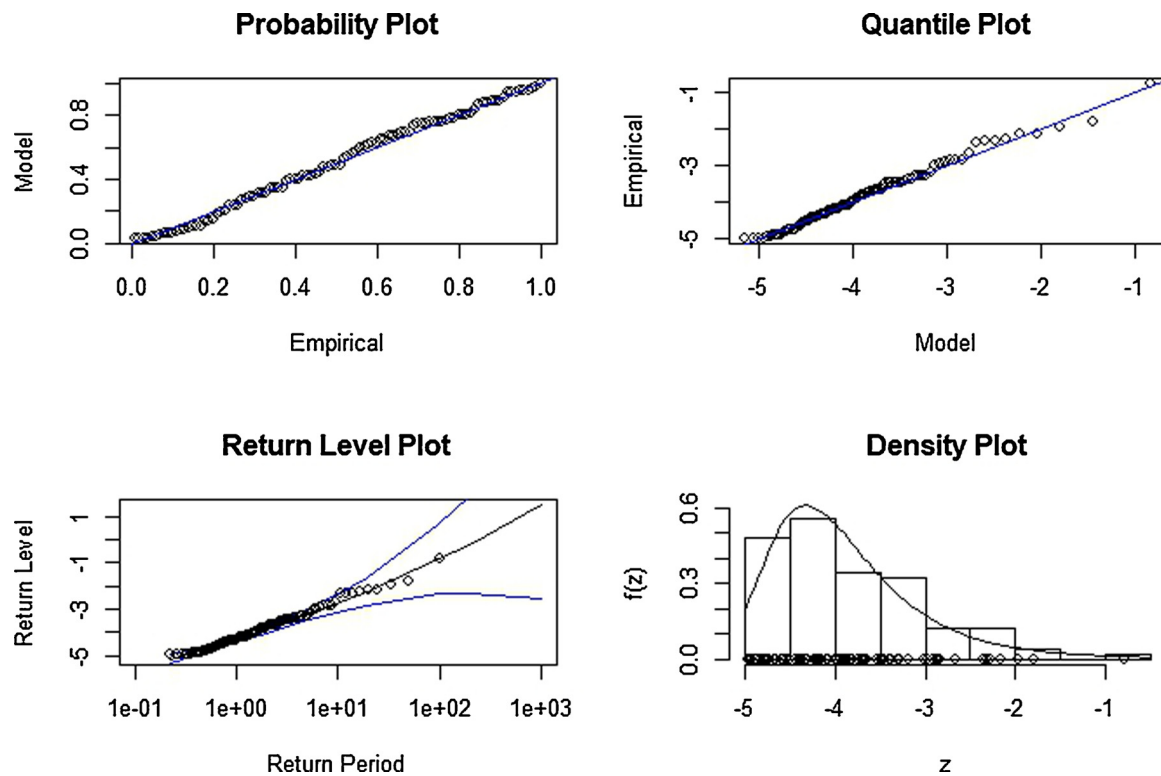
## Density Plot

**Fig. 2.** Diagnostic plots for GEV fit to $TTC_{min}$ (near-crash threshold < 5 s).

return period can be interpreted in years; in this particular case each block is an individual near-crash interaction. Using the previously calculated probability of crash occurrence (0.0733) one can calculate the return period, which is $1/0.0733 = 13.65$. In other words one out of every 14 near-crash interactions (with a $TTC_{min}$ smaller than 3.5 s) will result in a crash.

The analysis revealed that the 3.5 s as a threshold value for near-crash situations lead to unsatisfactory model results and irrationally high crash probability. This is due to the combined effect of practical as well as statistical reasons. The initially small sample size of $TTC_{min}$ is due to the fact that we are looking at left turning and straight moving vehicle interactions, where in many cases $TTC$ cannot be interpreted due to stopped left-turning vehicles waiting for straight moving ones to pass. From a statistical point of view the small sample size results in unreliable extrapolation and large variance.

For the above reasons several models were tested using different threshold levels for the pre-selection of near-crash situations. The threshold was gradually increased with a 0.5 s increment. Diagnostic plots revealed that by increasing the near-crash threshold the model fit improves gradually, Fig. 2 shows the results using 5 s as a threshold for near-crash situations.

With the help of these diagnostic plots – probability, quantile, return level and density plots – one can check the goodness of the model. The probability plot is a comparison of the empirical and fitted distribution functions, in the quantile plot their quantiles against each other are plotted. Both can be visually checked, if in both cases the points are sufficiently close to linearity, the model can be accepted. An example of a good fit is shown for $TTC_{min}$ (near-crash threshold < 5 s) in Fig. 2. The density plot is a comparison of the probability density function of the fitted model with the histogram of the data; in this case this plot also shows consistency. The return level plot (return periods vs. return levels) also shows observed values (dots) as well as model-based estimation (line) along with confidence bounds. What we are interested in is the return period associated with the return level when a temporal indicator is equal to zero. The shape of the curve also indicates the type of distribution, which is Frechet since $\xi > 0$.

In order to further investigate the probability plot and to compare the fitted and the empirical distributions, a Kolmogorov–Smirnov test was used, of which the null hypothesis is that the sample is drawn from the fitted distribution. As the *p*-values are greater than 0.05 we cannot reject the null hypothesis that our sample deviates from the GEV distribution.

From Table 2 it can also be seen that as the near-crash threshold increases (resulting in bigger sample size) the shape parameter converges to zero. With 4.5 s threshold the 95% confidence intervals include zero. Setting the shape parameter to zero the Gumbel distribution can be fitted and an analysis of deviance between the two models can reveal whether it is more appropriate (Penalva et al., 2013). The results obtained for 3.5, 4, and 4.5 s showed significant differences between the two models, however, for 5 s there was no significant difference so the Gumbel model with two parameters is a good choice for modeling these data.

As for $T_{2min}$ further steps in model checking are just the opposite as compared to those of $TTC_{min}$. As previously noted, for a critical value of near-crash situations the literature actually suggests a lower threshold than 3.5 s, as low as 1.5 s. Thus, it is interesting to check how the model fit and output values change as we gradually decrease the near-crash threshold level. In Table 2 the shape of the distribution changes from a Weibull ($\xi < 0$) to a Frechet ($\xi > 0$) as the near-crash threshold levels as well as the sample sizes decrease. Crash probability is gradually increasing by decreasing near-crash thresholds. At a near-crash threshold of 2 s a crash probability of 0.0098 is calculated associated with a return level of 101.96 meaning that one crash would happen out of 102 near-crash interactions. The model fit associated with the 2 s near-crash threshold still gave acceptable results. The Kolmogorov–Smirnov tests gave the same results as for $TTC_{min}$ for all the thresholds.

### 3.2. Peak over Threshold

The POT approach offers a different solution to modeling extreme events. It is necessary to choose a threshold over which extreme events are considered. "It is important to choose a sufficiently high threshold

in order that the theoretical justification applies thereby reducing bias. However, the higher the threshold, the fewer available data remain. Thus, it is important to choose the threshold low enough in order to reduce the variance of the estimates." (Gilleland and Katz, 2016).

There are basically two methods for selecting the appropriate threshold:

- Mean residual plot: this plot shows the mean of the excesses depending on the value of the chosen threshold level $u$. Above a certain value the GPD provides a valid approximation to the excess distribution (Coles, 2001). Here a threshold has to be selected where the graph is linear within uncertainty bounds. This is, however, not always straightforward, and based on a subjective choice.
- Model estimation: the model is estimated at a range of threshold values with the intention to find stable model parameters. Again, above a certain level of $u$ the GPD is valid, if estimates of the shape parameter $\xi$ are constant, while estimates of the scale parameter $\sigma$ is linear in u. This point can be read from the plot by checking linearity, in other words estimates will not change much within uncertainty bounds, as the threshold increases.

As for negated $TTC_{min}$ the lowest threshold where the mean residual plot becomes linear within uncertainty bounds is a value around $-4$ s (Fig. 3). The parameter estimates against thresholds also show relatively stable results for the selected value (Fig. 4). The GPD model results are given in Table 3. The shape parameter $\xi$ is below zero resulting in a convex return level plot with a finite upper bound. To check the goodness of the model the same diagnostic plots as for the block maxima approach (probability, quantile, return level and density plots) as well as the Kolmogorov–Smirnov tests were used.

The probability of crash occurrence, namely when $TTC_{min}$ is smaller than zero, can be calculated using Eq. (6) substituting the model parameters $\xi$ and $\sigma$, as well as the threshold $u = -4$ and $x = 0$. This calculation yields a crash probability of 0.00017. The return period associated with $TTC_{min} < 0$ is 5884.8 (1/0.00017), meaning that one out of 5885 near-crash interactions results in a crash:

$$\Pr\{X > x | X > u\} = \left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi} \tag{6}$$

As for negated $T_{2min}$ a different threshold was chosen, both the mean residual plot and the plots of parameter estimates against thresholds (these are not presented here) suggest a threshold of $-2$ s to be used. The crash probability associated with this model ($T_{2min} < 0$) is 0.00055 and the return period is 1807.3 (1/0.00055), meaning that one out of 1807 near-crash interactions results in a crash.

The GPD diagnostic plots suggested a reasonable model fit in both

cases, also the Kolmogorov–Smirnov tests were not significant and thus we could not reject that the samples deviate from the GPD distribution. However the return level plot showed that as the return period increases the return level confidence bounds tend to be wider for $TTC_{min}$ than for $T_{2min}$ meaning that the prediction of unobserved extreme values comes with less uncertainty for the latter.

## 4. Discussion and limitations

Modeling results are summarized in Table 4. Applying the POT approach seems to give more reasonable results in terms of crash probabilities and return periods, which were in the hundreds with BM but in the thousands with POT. If we accept a few assumptions we can attempt to validate these probabilities. These assumptions are as follows:

- the number of interactions used in the analysis (194 for $TTC_{min}$ and 792 for $T_{2min}$) was all the interactions observed in the 2-day period between 6 AM and 9 PM, and no interactions were left out;
- the observation period (6 AM–9 PM) is a good representation of the entire day and accidents did not happen outside this time period;
- accident data provided are accurate, namely 5 crashes due to the collision of left turning and straight going vehicles in a 11-year period, which is approximately 800 days/accident occurrence (one accident happened in 800 days on average).

Accepting the above assumptions and comparing the model results we can actually state that indeed the POT results are much closer to the actual crash frequency. The POT model for $TTC_{min}$ (245.20 days/accident) gives the best prediction, especially if we accept the assumption that property damage accidents are in general underreported. Validation results also show that $T_{2min}$ tends to overestimate crash frequencies, e.g. for POT $T_{2min}$ estimated one crash in every 28 days, which value was more realistic (245) for $TTC_{min}$.

The above results also illustrate that the near-crash threshold value affecting sample size is a critical issue, especially with the BM approach. Fig. 5 illustrates this by further refining near-crash threshold values. Here the block maxima approach was used and sub-samples were created using near-crash threshold values by using a 0.05 increment. As for $T_{2min}$ 75 models were fitted for near-crash values ranging from 1.3 to 5 s and for $TTC_{min}$ 31 models were fitted for near-crash values ranging from 3.5 to 5 s. Crash probabilities were calculated for all the models. As the sub-sampling near-crash threshold increases, sample sizes become bigger resulting in better model fits, even though with less pragmatic near-crash thresholds.

Fig. 5 also illustrates that for different near-crash thresholds different crash probabilities are predicted for the two indicators. As we are
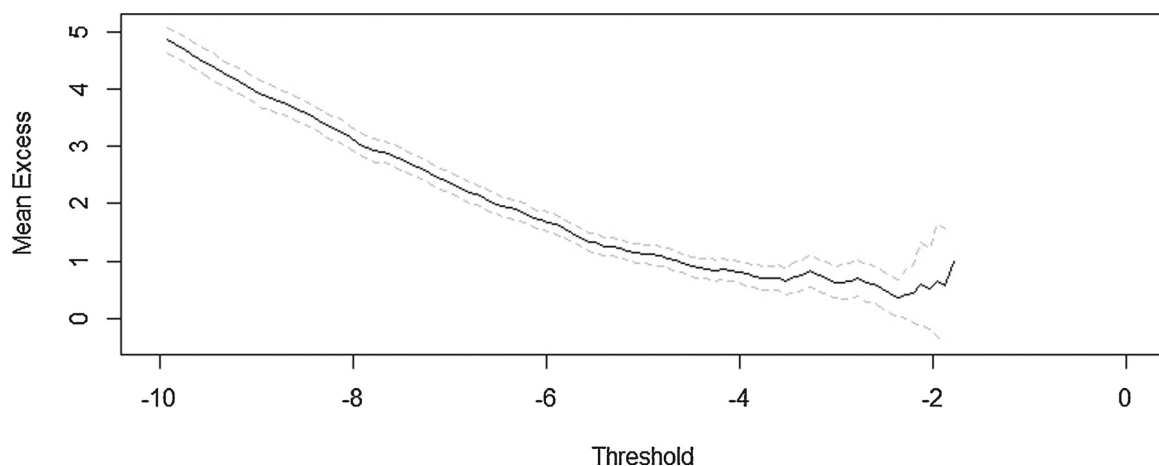


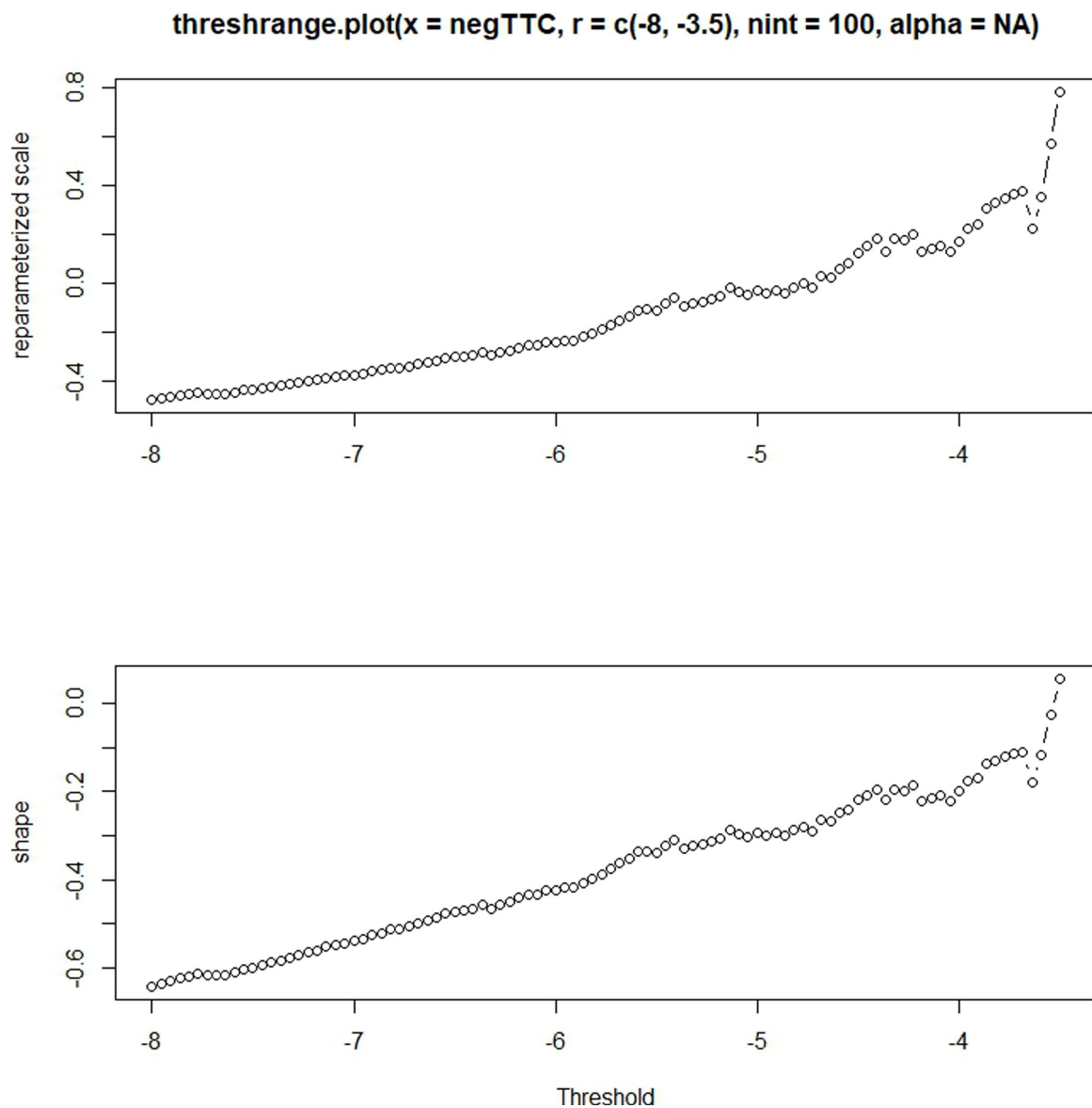**Fig. 3.** Mean residual plot for $TTC_{min}$.

## threshrange.plot(x = negTTC, r = c(-8, -3.5), nint = 100, alpha = NA)



<div align="center">Fig. 4. Parameter estimates against threshold for $TTC_{min}$.</div>

**Table 3**
Results of GPD for $TTC_{min}$ and $T_{2min}$.

| Indicator | | Scale $\sigma$ | Shape $\xi$ |
|---|---|---|---|
| $TTC_{min}$ | Sample size | | 48 |
| | Estimated parameter | 0.970 | −0.199 |
| | Standard error | 0.180 | 0.120 |
| | Lower bound of confidence interval 95% | 0.617 | −0.436 |
| | Upper bound of confidence interval 95% | 1.323 | 0.036 |
| | Log-likelihood | | 36.953 |
| | Kolmogorov–Smirnov test *p*-value | | 0.997 |
| $T_{2min}$ | Sample size | | 130 |
| | Estimated parameter | 0.585 | −0.246 |
| | Standard error | 0.060 | 0.058 |
| | Lower bound of confidence interval 95% | 0.467 | −0.360 |
| | Upper bound of confidence interval 95% | 0.703 | −0.132 |
| | Log-likelihood | | 28.244 |
| | Kolmogorov–Smirnov test *p*-value | | 0.992 |

analyzing two indicators of different nature (a collision course ($TTC_{min}$) and a crossing course ($T_{2min}$) indicator), it is worth investigating for what near-crash threshold values would they predict similar crash

probabilities, in other words whether there is transferability in between them. This would provide further insight into the applicability of collision and crossing course indicators. Based on the results shown in Fig. 5 threshold values were selected for which both indicators yielded the most similar crash probability (i.e. the difference between the predicted crash probabilities was marginal). This plot (Fig. 6) actually describes the relationship between these two indicators, saying under what near-crash thresholds we receive almost the same crash probability. There is some fluctuation in the graph, but the pattern clearly shows that for $T_{2min}$ lower near-crash thresholds than compared to $TTC_{min}$ would yield the same crash probability (e.g. 5 s for $TTC_{min}$ and 2.5 s for $T_{2min}$). A Pearson correlation test was used to determine how strong the relationship is. This test was highly significant (*p*-value = 8.96e−14) and indicated a strong correlation (0.93).

As every research, this one also comes with certain limitations, the most important of those are summarized below.

The basis of the analysis was a dataset collected in Minsk, Belarus. All the data came from one location, thus cross comparison in between different locations was not possible. Using several locations can actually give an added value, Zheng et al. (2018) for instance used 16 merging areas and could also provide a comparison of models across these

**Table 4**
Summary of results of GEV and GPD for $TTC_{min}$ and $T_{2min}$.

| Method | Indicator (threshold) | Probability | Probability of crash | Return period | Sample (2 days) | Days per accident |
|--------|----------------------|-------------|---------------------|---------------|-----------------|-------------------|
| BM | TTC (< 3.5 s) | Pr{TTC < 0\|TTC < 3.5} | 0.0733 | 13.65 | 31 | 0.88 |
| | TTC (< 4 s) | Pr{TTC < 0\|TTC < 4} | 0.0135 | 73.92 | 48 | 3.08 |
| | TTC (< 4.5 s) | Pr{TTC < 0\|TTC < 4.5} | 0.0098 | 101.04 | 76 | 2.66 |
| | TTC (< 5 s) | Pr{TTC < 0\|TTC < 5} | 0.0040 | 246.57 | 100 | 4.93 |
| | T2 (< 2 s) | Pr{T2 < 0\|T2 < 2} | 0.0098 | 101.96 | 130 | 1.57 |
| | T2 (< 2.5 s) | Pr{T2 < 0\|T2 < 2.5} | 0.0044 | 226.10 | 232 | 1.95 |
| | T2 (< 3 s) | Pr{T2 < 0\|T2 < 3} | 0.0033 | 302.40 | 341 | 1.77 |
| | T2 (< 3.5 s) | Pr{T2 < 0\|T2 < 3.5} | 0.0016 | 596.02 | 443 | 2.69 |
| POT | TTC (< 4 s) | Pr{TTC < 0\|TTC < 4} | 0.00017 | 5884.80 | 48 | 245.20 |
| | T2 (< 2 s) | Pr{T2 < 0\|T2 < 2} | 0.00055 | 1807.26 | 130 | 27.80 |

locations. The observation periods at these locations were quite short (56-88 minutes) and the authors admitted that this short time frame can hardly claimed to be representative for a five-year period for which accident data was gathered for validation. Nevertheless, similar to Zheng et al. (2018) the authors argue that "the estimated number of crashes can still provide some reference for model evaluation". In the current research one location with data for two days was used (6 AM–9 PM), which is in some sense more representative, even though estimated results were compared with accident data gathered for an eleven-year period.

Besides being representative in terms of the length of time period analyzed it can also be questioned whether the results based on observations in Minsk can be generalized or they are location specific. The transferability of these results to other locations depends on many factors, such as geometric features of intersections (e.g. size, channelization) and road user behavior (e.g. priority giving or surrendering attitude, gap acceptance).

The type of interaction analyzed was exclusively left-turning vs. straight moving vehicle to vehicle interactions. The results gained in this research are therefore restricted and applicable to this interaction type only.

Probably the most important limitation was the uncertainty of validation, which was possible along with certain assumptions. Even if accident data are available one has to be cautious about its accuracy, especially if mostly low severity or property damage accidents are present (underreporting).
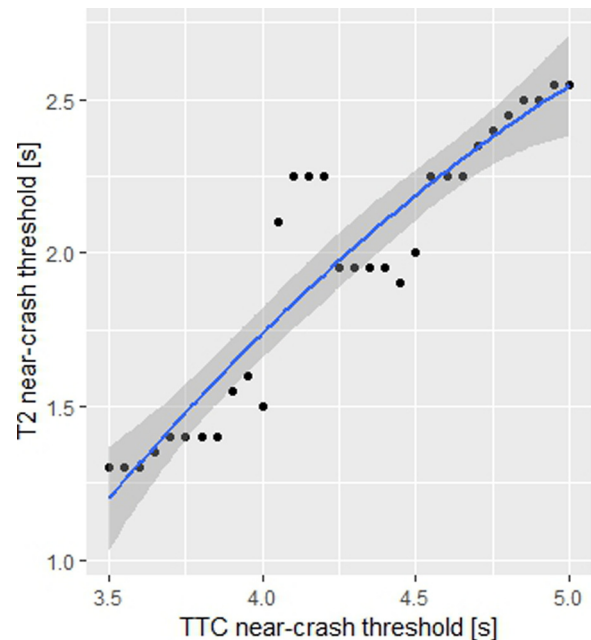


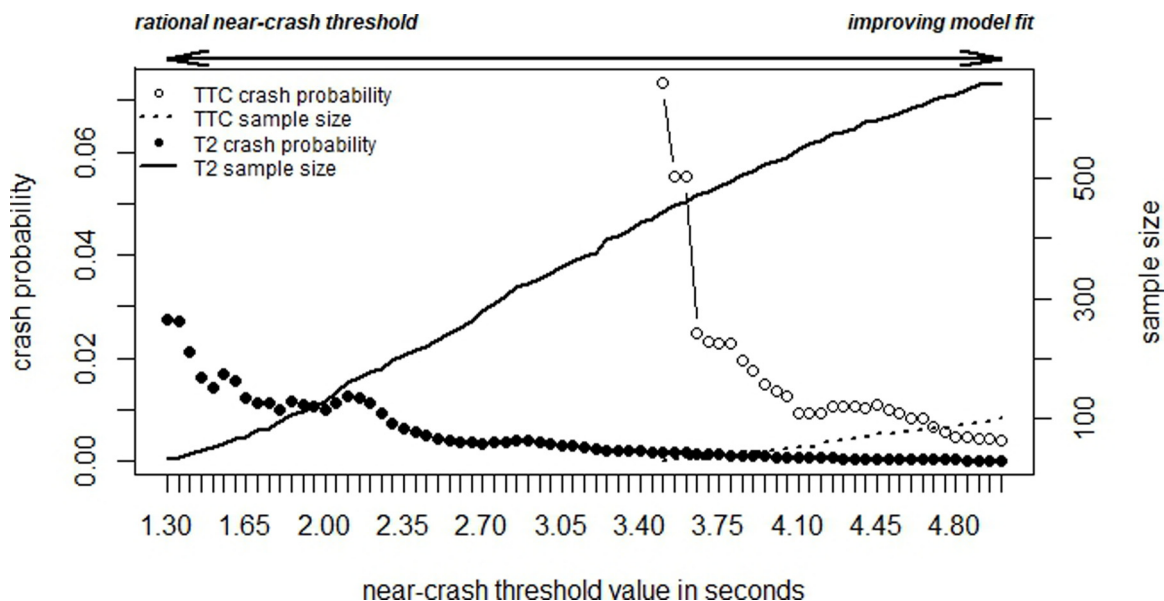**Fig. 6.** Temporal indicators with similar crash probabilities.



**Fig. 5.** Results of block maxima approach for a sequence of threshold values.

## 5. Conclusions

As for the two EVT approaches it can be concluded that overall applying the POT approach seemed to give more reliable and pragmatic results. When applying the block maxima approach the selection of near-crash situations as a sub-sampling step proved to be a critical issue. As for $TTC_{min}$ the question was to which level the near-crash threshold should be increased to have a reasonable model fit, whereas for $T_{2min}$ to which – from a traffic a safety point of view – more reasonable level can we decrease the near-crash threshold in such a way that we still have a good model fit.

In the former case with $TTC_{min}$ increasing the near-crash threshold resulted in better model fits, however from a traffic safety point of view these high thresholds cannot actually be considered as near-crash events. In the latter case with $T_{2min}$ the threshold value could be further decreased with the disadvantage of slightly less well performing models. Obviously there is a trade-off between a good model fit and reasonable threshold values.

Judging which indicator is better could be done by validation using a proper accident dataset which was unfortunately not at hand for this study, which is a limitation. Notwithstanding, it has to be noted that this is not an exceptional case, and even with available historical data its applicability can be questioned in general (for reasons associated with the drawbacks of accident records already outlined previously). A judgment can only be done by considering the rationalism of crash probabilities, return periods, goodness of the models and by using the available accident data for validation along with a few assumptions. Overall we found that using the POT approach $TTC_{min}$ yielded more pragmatic results with wider confidence intervals (more accurate but less precise), whereas $T_{2min}$ showed better model fits but overestimated crash probabilities (more precise but less accurate).

Having compared the estimated crash probabilities for different near-crash thresholds and by checking the correlation between them we can conclude that collision and crossing course indicators are transferable. Crash probabilities calculated using EVT showed that one has to be "stricter" against crossing course interactions, as compared to collision course interactions, lower near-crash values would yield similar probabilities. The analysis revealed that for straight moving and left-turning vehicle-vehicle interactions, in comparison with a crossing course indicator $T_{2min}$, the limitation in using the collision course indicator $TTC_{min}$ is due to its smaller sample size that can be gathered in a given time period.

A possible step to refine the models is using motion prediction. As the above investigated indicators both assume constant speeds and unchanged paths, which is not realistic, it is worthwhile considering a probabilistic approach to predict trajectories and speeds of interacting vehicles. This approach would result in different values with different probabilities for a single interaction, thus providing an increased sample size for both indicators. St-Aubin et al. (2015) developed an approach called Probabilistic Surrogate Measures of Safety (PSMS) considering all possible paths that may lead to two road users to collide. At the time of writing there are also initiatives at Lund University to apply a probabilistic framework.

Another aspect that of relevant interest is modeling the severity of conflicts using surrogate safety indicators. Even though the above investigated temporal indicators may be used on their own to capture the severity of an interaction for instance by applying a threshold value in case of TTC, they are not sufficient to fully describe the severity of consequences. As Laureshyn et al. (2010) stated there is a need for the time-based indicators to be complemented with some speed-related indicator. To that end a number of researchers (Zheng et al., 2018, 2019; Jonasson and Rootzén, 2014; Cavadas et al., 2017; Farah and Azevedo, 2017; Wang et al., 2019) have applied bivariate extreme value models. Currently the authors are also working on bivariate models trying to capture the severity dimension of interactions and intend to publish the results in a separate paper.

## Authors' contribution

Attila Borsos: conceptualization, methodology, validation, writing – original draft, writing – review & editing; Haneen Farah: methodology, writing – original draft, supervision; Aliaksei Laureshyn: data curation, validation, writing – review & editing; Marjan Hagenzieker: supervision.

## References

Åsljung, D., Nilsson, J., Fredriksson, J., 2016. Comparing collision threat measures for verification of autonomous vehicles using extreme value theory. IFAC-PapersOnLine 49 (15), 57–62.

Cavadas, J., Azevedo, C.L., Farah, H., 2017. Extreme value theory approach to analyze safety of passing maneuvers considering drivers' characteristics. IATSS Res. 41 (1), 12–21.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values.

Farah, H., Azevedo, C.L., 2015. Using extreme value theory for the prediction of head-on collisions during passing maneuvres. IEEE Conference on Intelligent Transportation Systems, Proceedings.

Farah, H., Azevedo, C.L., 2017. Safety analysis of passing maneuvers using extreme value theory. IATSS Res. 41 (1), 12–21.

Gilleland, E., Katz, R.W., 2016. extRemes 2.0: an extreme value analysis package in R. J. Stat. Softw. 72 (8).

Hauer, E., 1997. Observational Before-After Studies in Road Safety. Emerald Group Publishing Limited, Oxford, UK.

Hayward, J.C., 1972. Near-miss determination through use of a scale of danger. Highway Res. Rec. 384, 22–34.

Huang, F., Liu, P., Yu, H., Wang, W., 2013. Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections. Accid. Anal. Prev. 50, 1014–1024.

Hydén, C., 1987. The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique (Ph.D. thesis). Lund University, Sweden.

Jonasson, J.K., Rootzén, H., 2014. Internal validation of near-crashes in naturalistic driving studies: a continuous and multivariate approach. Accid. Anal. Prev. 62, 102–109.

Laureshyn, A., De Ceunynck, T., Karlsson, C., Svensson, A., Daniels, S., 2017. In search of the severity dimension of traffic events: extended Delta-V as a traffic conflict indicator. Accid. Anal. Prev. 98, 46–56.

Laureshyn, A., Svensson, A., Hydén, C., 2010. Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation. Accid. Anal. Prev. 42, 1637–1646.

Mahmud, S.S., Ferreira, L., Hoque, M.S., Tavassoli, A., 2017. Application of proximal surrogate indicators for safety evaluation: a review of recent developments and research needs. IATSS Res. 41, 153–163.

Nicholson, A.J., 1985. The variability of accident counts. Accid. Anal. Prev. 17 (1), 47–56.

Penalva, H., Neves, M., Nunes, S., 2013. Topics in data analysis using R in extreme value theory. Metodol. Zvezki.

R Core Team, 2013. R: A Language and Environment for statistical Computing.

Sayed, T., Zaki, M.H., Autey, J., 2013. Automated safety diagnosis of vehicle-bicycle interactions using computer vision analysis. Saf. Sci. 59, 163–172.

Songchitruksa, P., Tarko, A.P., 2006. The extreme value theory approach to safety estimation. Accid. Anal. Prev. 38, 811–822.

St-Aubin, P., Saunier, N., Miranda-Moreno, L., 2015. Large-scale automated proactive road safety analysis using video data. Transp. Res. Part C: Emerg. Technol. 58, 363–379.

Svensson, A., 1998. A Method for Analysing the Traffic Process in a Safety Perspective (Doctoral dissertation), (Ph.D. thesis). Lund University, Sweden.

Svensson, A., Hydén, C., 2006. Estimating the severity of safety related behaviour. Accid. Anal. Prev. 38, 379–385.

T-Analyst, 2016. Software for Semi-Automated Video Processing.

Tarek, S., Sany, Z., 1999. Traffic conflict standards for intersections. Transp. Plann. Technol. 22 (4), 309–323.

Tarko, A., Davis, G., Saunier, N., Sayed, T., Washington, S., 2009. White Paper SURROGATE MEASURES OF SAFETY ANB20(3) Subcommittee on Surrogate Measures of Safety ANB20 Committee on Safety Data Evaluation and Analysis.

Tarko, A.P., 2012. Use of crash surrogates and exceedance statistics to estimate road safety. Accid. Anal. Prev. 45, 230–240.

Wang, C., Xu, C., Dai, Y., 2019. A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. Accid. Anal. Prev. 123, 365–373.

Wang, C., Xu, C., Xia, J., Qian, Z., Lu, L., 2018. A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation. Transp. Res. Part C: Emerg. Technol. 90, 281–291.

Zheng, L., Ismail, K., Meng, X., 2014. Freeway safety estimation using extreme value theory approaches: a comparative study. Accid. Anal. Prev. 62, 32–41.

Zheng, L., Ismail, K., Sayed, T., Fatema, T., 2018. Bivariate extreme value modeling for road safety estimation. Accid. Anal. Prev. 120, 83–91.

Zheng, L., Sayed, T., Essa, M., 2019. Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. Accid. Anal. Prev. 123, 314–323.