UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Civil Environmental and Architectural Engineering

Ph.D. COURSE IN: SCIENZE DELL'INGEGNERIA CIVILE, AMBIENTALE E DELL'ARCHITETTURA
CURRICULUM: Materiali, strutture, sistemi complessi e architettura
SERIES 34th

A CONFLICT-BASED APPROACH FOR ROAD SAFETY ANALYSIS

Coordinator: Prof. Carmelo Maiorana
Supervisor: Prof. Riccardo Rossi
Co-Supervisor: Prof. Massimiliano Gastaldi

Ph.D. student: Federico Orsini

Table of contents

Acknowledgements

Working on my doctoral dissertation has been an exciting and rewarding time and an important step in my life.

This work would not have been possible without the support of my supervisors, Prof. Riccardo Rossi and Prof. Massimiliano Gastaldi, who provided me with several useful suggestions, insights, and guidance during my years as a doctoral student.

A special thanks goes also to Prof. Bernhard Friedrich, who welcomed me in his research group at the Technische Universität Braunschweig for some months, giving me the opportunity to conclude my doctoral path in the best possible way.

I thank the Department of Civil Environmental and Architectural Engineering of the University of Padua, the Ermenegildo Zegna Foundation and the Ing. Aldo Gini Foundation for the financial support that made this work possible.

Finally, a big thank you goes to my lovely fiancée Debora, my family, and my friends who always believe in me and make my journey in life special.

1. Introduction

Traditionally, road safety analysis relies on the use of crash data. However, several issues may affect these data: lack of availability, lack of spatial and/or temporal precision, under-reporting, misclassification; moreover, crashes are relatively rare events, so data must be collected for several years and/or in several different locations to obtain enough data. An alternative approach consists of analyzing traffic conflicts, which can be defined, intuitively, as "near-crashes" (see Section 1.1).

Despite an ever-growing interest in traffic conflicts in transportation research, there are still several open questions that this research aims to answer, for both long-term and real-time road safety applications.

1) How to model the probabilistic relationship between traffic conflicts and crashes?
2) Is it possible to predict crashes in real-time with a conflict-based approach?

The first part of this dissertation (Chapter 2) aims to provide further insight on the probabilistic relationship between traffic conflicts and crashes, by applying univariate and bivariate extreme value theory, for long-term road safety analysis (i.e., prediction of annual crash rates in selected infrastructures). A real-world case study using data collected with radar sensors at several cross-sections of an Italian highway illustrates an operational-oriented procedure which aims to identify surrogate measures of safety threshold separating normal traffic interactions and conflicts, and to translate raw conflict data into annual crash rates. The results of the analysis are validated using historical crash rates: in about 90% of the highway cross-sections analyzed, the observed average annual number of crashes falls within the 95% confidence interval of the predicted number of crashes.

The second part (Chapter 3) focuses on developing a conflict-based approach for real-time road safety analysis. Using data from the previous case study, conflicts are identified using the extreme value theory thresholds. Then machine learning-based methods are used to link real-time traffic variables to traffic conflicts that occur in the immediate future. The objective is using traffic conflicts to train an AI predictor able to forecast whether there will be an increased risk of crashes in the immediate future, based on the current traffic variable collected at several highway cross-sections. The results are compared to those of a traditional real-time crash prediction model (estimated with crash data), showing a significant improvement in terms of accuracy, sensitivity, and specificity.

The application of conflict-based approaches has the potential to provide a positive broader impact on road safety. In particular, the possibility to avoid the use of crash data in practical applications can (i) allow to apply statistical methods to new scenarios in which crash data are unavailable or unreliable (e.g., rural roads, third-world countries, new infrastructures); (ii) proactively analyze safety, avoiding the ethical dilemma of crash-based approaches, in which injuries and fatalities are needed in order to correctly identify unsafe

behaviors and locations; (iii) provide faster road-safety evaluations, since traffic conflicts are more frequent than crashes; for the same reason, provide more flexible and resilient road safety models, which can be based on more recent data.

This dissertation combined the use of existing tools (traffic conflict theory, extreme value theory, machine learning) in order to answer in an innovative way to the research question expressed above. In particular, the application of extreme value theory to such a large-scale scenario is unique in the literature, whereas the combination of extreme value theory with real-time road safety is the most innovative contribution of the present research. The procedures described in the dissertation were developed with practical operations in mind and can be realistically applied in the real world.

1.1 Traffic conflict theory

1.1.1 Issues of dealing with crash data

Traditionally, road safety analysis is carried out with the use of crash data. However, several issues may affect these data: lack of availability, lack of spatial and/or temporal precision, under-reporting, and misclassification. The reliability of a model calibrated with crash data is strongly dependent on the reliability of such data. Furthermore, crashes are relatively rare events, so data must be collected for several years and / or at several different locations to obtain enough data to apply certain statistical models (Tarko, 2018a). Drivers involved in crashes may not be a representative sample of the driving population, as riskier drivers are likely overrepresented in crash data, potentially leading to biased interpretation of model parameters (Mannering, 2018; Mannering et al., 2020). Finally, since the crash-based approach is reactive, it faces the ethical dilemma that injuries and fatalities are needed to correctly identify unsafe behaviors and locations (L. Zheng et al., 2021).

1.1.2 An alternative approach: traffic conflicts

To avoid the use of crash data, in road safety analysis there is an alternative approach, which consists in analyzing conflicts. The use of the term "traffic conflict" to describe a near-crash situation dates back to the 1960s and 1970s (Hayward J.C., 1972; Perkins & Harris, 1968). A formal definition was proposed by the Association for International Cooperation on Traffic Conflict Techniques: "A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged" (Güttinger, 1984). As underlined by Davis et al. (2011), this definition implies that an event can be considered a conflict only if it satisfies a counterfactual analysis: if the users had not modified their movement, then a collision probably would have occurred. Tarko (2019) further elaborated on this concept, explicitly affirming that conflicts are "precursors of crashes and not alternative outcomes." This is visually represented in Figure 1: a set of varied causes may induce a traffic conflict between road users; if there is no intervention, this results in a crash, whereas an effective intervention can prevent it. The presence of some kind of system "failure" (e.g., high traffic volume, driver distraction, etc.) is a necessary condition for a conflict to occur, which in turn is necessary for a collision to occur.

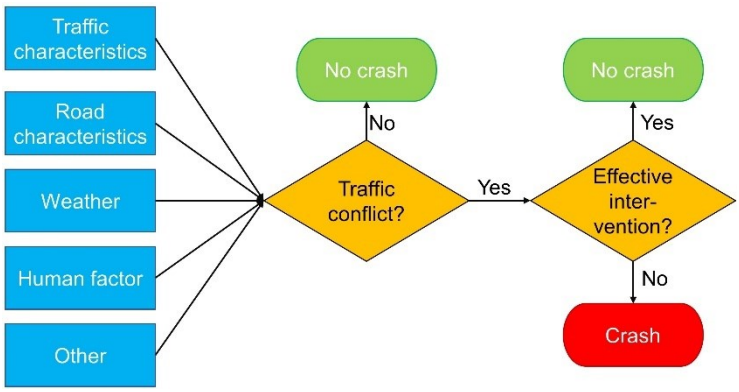

Figure 1. Causality model of traffic conflicts and crashes (adapted from Tarko, 2019)

1.1.3 Quantifying traffic conflicts: surrogate measures of safety

Several conflict-based studies identified and used traffic conflicts qualitatively, e.g., considering only those situations in which an evasive maneuver occurred (Caleffi et al., 2017; Williams, 1981). Although an evasive action-based approach has shown some potential in studying vehicle-pedestrian interactions (Tageldin et al., 2017), it may have some drawbacks in vehicle-to-vehicle applications, because it may generate a large number of false positives, as there may be substantial similarity between drivers' responses to hazardous and (misperceived) nonhazardous situations (Tarko, 2019). Furthermore, some road conflicts can occur even in the absence of observable evasive maneuvers, due to the lack of awareness of drivers, and it has been observed that a significant number of crashes are not preceded by any attempt to avoid them (Ueyama, 1997).

Consequently, most conflict-based studies adopted a quantitative approach by considering either the temporal/spatial closeness to a collision or the intensity of the evasive maneuver. In this way, it is possible to determine the conflict severity and/or identify a threshold to separate traffic conflicts from normal interactions between road users. Proximity indicators used to quantify traffic conflicts are commonly referred to as "surrogate measures of safety" or "surrogate safety measures".

As regards quantifying the closeness to a collision, the most commonly used are time-to-collision (TTC) and post-encroachment time (PET). TTC is a continuous measure, defined as the time remaining to a collision if the road users involved maintain constant speed/trajectories (Hayward J.C., 1972). PET is the time difference between the instant in which the first vehicle leaves the conflict point and that in which the second vehicle reaches it (Allen et al., 1978). Since traffic conflicts would end up as crashes if the separation between vehicles reaches zero, lower values of such measures indicate a higher severity of the conflict.

As regards quantifying the intensity of the evasive maneuver, it is worth mentioning the deceleration rate to avoid a crash (DRAC), introduced by Cooper & Ferguson (1976). Here, the higher the value of the measure, the higher the severity of the conflict.

Recently, extensive reviews of surrogate measures of safety have put some order in the vast literature on them. The reader is referred to the works of (Arun et al., 2021; Johnsson et al., 2018; Chen Wang et al., 2021; L. Zheng et al., 2021) for further details and alternative surrogate measures of safety. In this dissertation, the measures used in each case study will be properly introduced and described in the respective sections.

1.1.4 The crash-conflict relationship

Since early studies in traffic conflict theory, researchers have tried to connect the frequency of conflicts with that of crashes. Tarko (2018a) has proposed a classification of different approaches to model the relationship between conflict and crashes (see Table 1).

Table 1. Classification of approaches to the crash-conflict relationship (adapted from Tarko, 2018a)

Type	Model form
1	Crashes $(T,Y)=K$*Conflicts(t,X)
2	Crashes $(T,Y)=K(Y,X)$*Conflicts(t,X)
3	Crashes $(t,X)=R(X)$*Conflicts(t,X)

In the earliest works (type 1 approach in Table 1), a single and fixed conflict ratio K was sought (Hauer, 1982; Hydén, 1987). The idea was to compare crash and conflict data collected at various locations and to identify a single coefficient to link them. Since collisions are rare events, crash data were collected for long periods of time T (e.g., months or years); conversely, as the collection of conflicts is expensive and troublesome, conflict data were recorded in a much shorter time interval t (e.g., hours or days). Consequently, there was a strong heterogeneity between conditions X (i.e., traffic, weather, etc.) in which conflicts were observed and conditions Y for historical crash data; also, the latter conditions were mostly aggregated or unknown. This led to obtaining mixed results and highlighted the need to adopt more advanced approaches.

The type 2 approach (Table 1) recognizes the need to account for this heterogeneity between conditions X and Y and therefore defines K not as a fixed ratio, but rather as a function of such conditions (El-Basyouny & Sayed, 2013; Guo et al., 2010; Sayed & Zein, 1999). This function can be estimated with count models, including conflict rates among explanatory variables.

The third approach is based on a rather different premise, according to which conflicts and crashes are the same types of events; in addition, the continuity of conflict severity is

assumed, similarly to what was originally proposed by Glauz & Migletz (1980) (see Figure 2). From this, it is possible to hypothesize a probability distribution for the collision nearness of traffic events and, from that distribution, to extract a crash probability. Campbell et al. (1996) and Songchitruksa & Tarko (2006) proposed to use extreme value distributions in order to obtain the probability of a crash to happen, given the observed conflicts. The most interesting aspect of this approach is the possibility to estimate this probability without the use of crash data, which are instead used only for model validation.

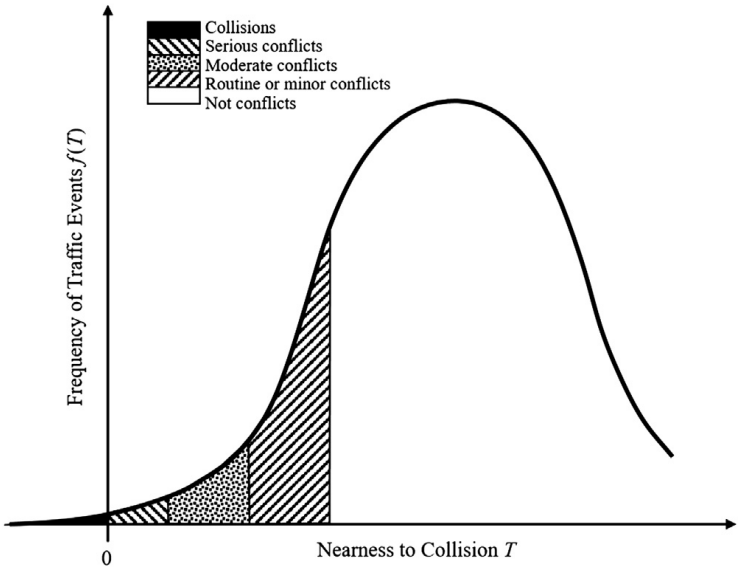

Figure 2. Illustration of the concept of continuous distribution of crash nearness. Source: Tarko (2019), adapted from Glauz & Migletz (1980).

1.2 Dissertation outline and objectives

As mentioned at the beginning of the Introduction, this dissertation is divided into two main parts: a long-term road safety analysis based on extreme value theory and a real-time prediction model based on AI methods. Figure 3 shows an outline of the dissertation. Traffic conflict theory (Section 1.1) is a fundamental starting point for this work, in order to carry out both long-term and real-time road safety evaluations.

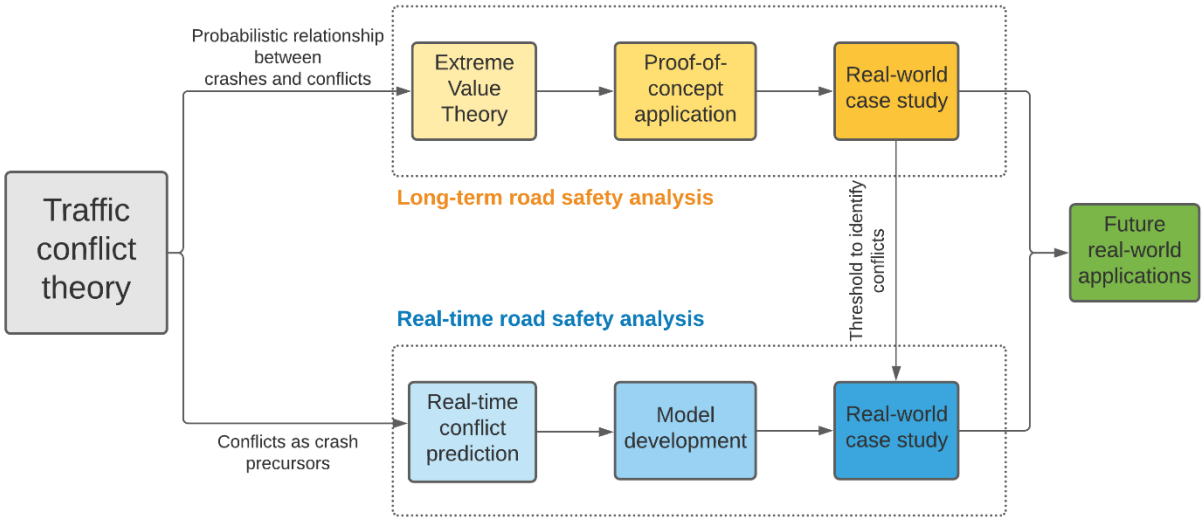

Figure 3. Outline of the dissertation

Chapter 2 addresses long-term road safety analysis (e.g., predicting annual crash rates) and is based on the premise that there is a probabilistic relationship between road crashes and traffic conflicts. As explained in Section 1.1.4, researchers have been looking for a relationship between traffic conflicts and crashes for some decades. Early works tried to identify a single coefficient based on crash and conflict counts, with mixed results, whereas, more recently, probabilistic approaches were able to provide better performance. Extreme value theory (EVT) is one of the most promising probabilistic approaches adopted in this field and has been gaining a lot of interest in the last few years. The base idea is that a dataset of surrogate measures of safety records can be used to estimate an EVT distribution, and, from that distribution is then possible to extrapolate a crash probability. There are several open questions regarding its application to road safety: which EVT approach is the most appropriate (block maxima or peak-over-threshold)? Is it possible to combine multiple surrogate measures of safety within the same model (e.g., bivariate EVT)? How does data collection time impact the goodness of the results, and should seasonality effect be taken into account? The first part of the dissertation aims to answer these questions.

Section 2.1 provides a literature review of EVT-based work in road safety carried out prior to this dissertation, highlighting the open questions listed above. Section 2.2 provides a theoretical overview of univariate and bivariate EVT. The application of EVT to road safety is illustrated, in practical terms, with two proof-of-concept studies, based on driving simulator experiments: the first (Section 2.3) applies univariate EVT to a roundabout scenario; the second (Section 2.4) compares univariate and bivariate EVT approaches in a unsignalized intersection scenario. The main case-study is presented in Section 2.5, and it consists in the application of EVT for the long-term road safety analysis of several real-world highway cross-sections. Further considerations on transferability and seasonality are reported in Appendix 1.

Chapter 3 deals with real-time road safety analysis, and it is based on the assumption that traffic conflicts are precursors to crashes (see Section 1.1.2). Although conflict analysis is increasingly established in road safety, it has not yet found its way into real-time road safety prediction models, except for very few works (see Section 3.1.2). To fill this gap, in this dissertation, a methodology is proposed to predict conflict-prone situations based on real-time traffic data.

In practical terms, a real-time conflict prediction model (RTConfPM) is structured the following way:

- The model is trained offline using historical traffic, weather, and road data as input and historical conflict data for calibration.
- The model is applied in real time using traffic, weather and road data collected at the present interval to predict whether an unsafe situation will exist in the near future.
- If an unsafe situation is predicted, then an intervention strategy is applied to prevent it. Several solutions can be applied, such as variable message signs or ramp metering.

The fundamental advantage of an RTConfPM over a traditional real-time crash-based model is that it can be developed even in the absence of highly spatial/temporal-accurate crash data, which are often difficult to retrieve, especially outside of the highway infrastructures of developed countries.

Again, there are several open questions in the literature.

First, how to define an "unsafe" situation in an RTConfPM? Here, we suggest adopting the peak-over-threshold EVT approach (developed in the first part of the dissertation) to identify a threshold that separates real traffic conflicts from normal traffic interactions. This represents a crucial and innovative contribution to the literature.

Another main novelty introduced by this work is that the results of the proposed conflict-based approach are compared to those of a traditional crash-based model calibrated on the

same dataset: this represents the first direct comparison between the two approaches in real-time road safety applications.

Finally, cross-sectional vehicle-by-vehicle data was used both to retrieve input traffic variables and to detect traffic conflicts. This was done by means of microwave radar sensors, which can produce reliable information with relatively low data collection and elaboration costs and has great potential in view of future large-scale practical implementations.

Section 3.1 presents a review of the literature on both traditional crash-based and alternative conflict-based models for real-time road safety analysis. Section 3.2 provides a description of the methodologies applied to develop the RTConfPM, and Section 3.3 applies it to a case study located on the same highway analyzed in Section 2.5. Here, there is a crucial connection between the two parts of the dissertation, as the identification of conflicts to calibrate the RTConfPM is made with a threshold defined with extreme value theory (see Section 3.3.2). Section 3.4 presents the development of a traditional crash-based model calibrated on the same datasets, and Section 3.5 directly compares the performance of the two approaches. A comparative study on the performance of different machine learning classifiers within a RTConfPM is presented in Appendix 2 and is used to justify the use of the support vector machine method in Section 3.3.

Chapter 4 concludes the dissertation with some general remarks and an overview of potential future practical applications of the models developed within the present dissertation.

2. Long-term road safety analysis

In this chapter of the dissertation, a conflict-based probabilistic approach to long-term road safety analysis is presented. It is based on the extreme value theory (EVT), which has been widely applied in many fields, such as hydrology and finance, but only relatively recently in road safety (see Section 2.1). EVT is a branch of statistics which deals with the extreme deviations from the median of probability distributions. It seeks to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than any previously observed. With respect to road safety, EVT is used to model the distribution of traffic conflicts, in order to estimate the probability of crashes (i.e., the "most extreme" traffic conflicts, see Figure 2).

There are two main approaches in EVT: the block maxima (BM) and the peak-over-threshold (POT). The fundamental differences between them are: (i) how they define extreme events, and (ii) which extreme value distribution they use.

Given a dataset of observed records of the variable of interest (such variable is generally a surrogate measure of safety in EVT applications to road safety), in the BM approach observations are aggregated into homogeneous time or space intervals, called blocks; the highest value in each block is considered an extreme event and it is sampled (Figure 4a). The dataset of extreme events is then used to fit a generalized extreme value (GEV) distribution.

In the POT approach, a threshold is defined, and only those higher than it are considered extremes and sampled (Figure 4b). Then, a dataset of exceedances over the threshold (i.e., differences between extreme events and the threshold) is used to fit a generalized Pareto (GP) distribution.

In EVT applied to road safety estimation, extreme events are usually characterized by low values of the variable of interest, since low values of surrogate measures of safety generally identify conflicts (see Section 1.1.3). Therefore, in order to apply the approaches described above, a negated transformation of the variable of interest is often performed, although other methods can be applied (e.g., considering the reciprocal).

In Section 2.2 a rigorous presentation of the two approaches is given.

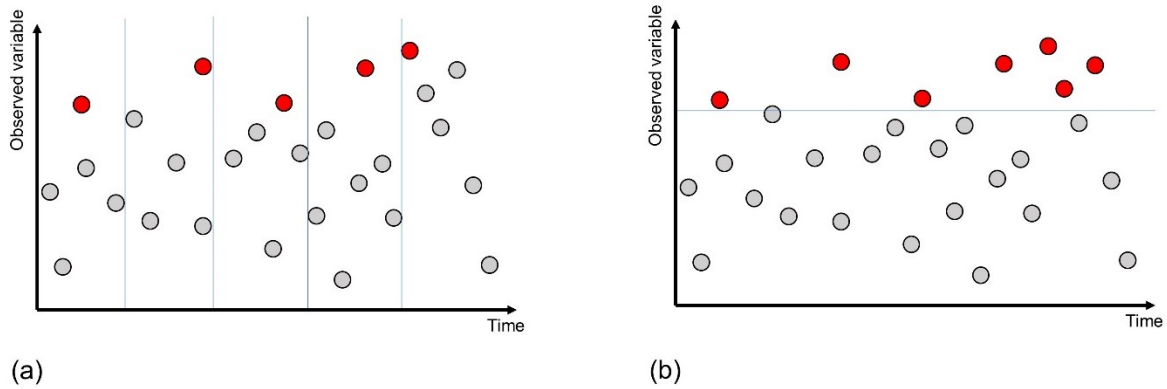

Figure 4. Graphical representation of EVT approaches: (a) block maxima and (b) peak-over-threshold.

2.1 Literature review

A report by Campbell et al. (1996) first and a research article by Songchitruksa & Tarko (2006) later were the first works dealing with the application of EVT to road safety. In the latter, the BM approach was used to analyze the safety of signalized intersections, using data collected with video cameras at 18 locations for 8 hours. In particular, the authors focused on the risk of right-angle collisions and used PET as a surrogate measure of safety. Their results showed a promising relationship between the crash estimates and the historical crash data; however, the authors highlighted the fact that their estimates were affected by large variance, due to the short observation period, and demonstrated with synthetic data that the method would require an observation period of at least 30 to 50 days, in order to achieve safety predictions with a precision comparable to that of 4 years of crash data.

A few years later, Tarko (2012) introduced the use of the POT approach to road safety, with a driving simulator experiment investigating the risk of road departures. Four participants were asked to drive 19 or 20 times on a 27-mile rural highway, during multiple experimental sessions. Several surrogate measures of safety were defined to study the risk of road departure: the lateral distance to departure, the time to road departure assuming constant speed and trajectory, and the time to road departure assuming constant lateral component of speed. Each of these measures was used to estimate a different model. The crash estimates were compared with the number of road departures observed during the experiment; for two out of three models, the observed number of road departures was included within the confidence interval of the crash estimate.

More recently, Tarko (2018b), proposed the use of another extreme value distribution, called Lomax, which is essentially a particular type of Pareto distribution, shifted so that its domain begins at zero. He compared different approaches for the estimation of parameters, with the use of synthetic data: Maximum Likelihood, Probability-Weighted Moments and Single Parameter Estimation. The latter was found to be more accurate and efficient.

The first comparative study between the BM and POT approaches in road safety was carried out by L. Zheng et al. (2014a). Their case study focused on lane-changing maneuvers in freeways using PET, and they collected data with video cameras in 29 locations for approximately 3 hours. As expected, due to the very short observation period, there was large variance in the estimated crashes; however, the authors were able to conclude that the POT approach performs better than the BM, in terms of data utilization, estimation accuracy and reliability. In a later study on the same dataset, the authors proposed using the Shifted Gamma-Generalized Pareto Distribution (L. Zheng et al., 2014b). In such work, they introduced a new method to transform the surrogate measure of safety, which consisted in considering the shifted-reciprocal, rather than the negated

value. They also compared the traditional Maximum Likelihood estimation with a Bayesian approach, which was found to provide more accurate and less uncertain estimated crashes.

Another comparative study between BM and POT was carried out by Farah & Azevedo (2017). They investigated the risk of head-on collisions during passing maneuvers with a driving simulator experiment involving 100 participants. The surrogate measure of safety they used was TTC, and the authors validated their model considering the number of collisions observed in the experiment. Contrary to L. Zheng et al. (2014a), their conclusion was that the BM approach provided more stable results, compared to the POT. The latter was, however, more consistently sensitive to the covariates of interest.

In the following years, great attention has been given to bivariate models, which are estimated using two sets of surrogate measures of safety, therefore combining more information. There is some confusion in road safety literature about the terminology of the methods adopted, but consistently with Beirlant et al. (2004), there are two main approaches, which can be considered as extensions to the univariate ones (more details in Section 2.2.2): the Component-wise Maxima (CM) and the Excesses Over a Threshold (EOT).

Using the same case study of Farah & Azevedo (2017), Cavadas et al. (2020) attempted to estimate the risk of head-on collisions with the CM approach: in addition to considering the dataset of TTC values, they included in the model an additional dataset containing the time headways between the overtaking vehicle and the vehicle being overtaken, calculated at the end of the passing maneuver. Compared to the univariate BM approach, the bivariate CM was able to improve the estimations.

L. Zheng et al. (2018; 2019a) applied the EOT approach to data on merging areas of highways (2018) and signalized intersections (2019a). In the first study, they compared univariate and bivariate approaches, showing that the latter was able to estimate crashes much more closely to the number of observed crashes. In the second study they compared bivariate models estimated with different combinations of surrogate of measures, concluding, after a correlation analysis, that using two sets of independent conflict indicators leads to better performance in crash estimation. It must be noted that both studies were carried out with short observation periods (a few hours, depending on the location) and therefore the variance in their estimates was rather significant. Another application of EOT was carried out by (Chen Wang et al., 2019), using videos recorded with drones at several intersections. Bivariate models were again found to be superior to univariate ones.

2.2 Methodology

2.2.1 Univariate extreme value theory

The following two sub-section will provide a theoretical overview on the GEV and GP distributions. For more details, the reader is referred to the works of Coles (2001) and Beirlant (2004).

2.2.1.1 Block Maxima approach and GEV distribution

In this approach, extreme observations are aggregated into time (or space) intervals, called blocks. In each block the event with the highest value, i.e., the block maximum, is considered as an extreme event and is sampled.

Mathematically, for independently and identically distributed random observations $\{X_1, X_2, \dots, X_n\}$, with unknown distribution function $F(x) = \Pr(X_i \leq x)$, the maximum $M_n = max\{X_1, \dots, X_n\}$ will converge to a GEV distribution when $n \to \infty$. The cumulative GEV distribution function is:

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

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ the scale parameter, and $-\infty < \xi < \infty$ the shape parameter. When the shape parameter is equal to zero, the distribution is called a type I GEV, also known as Gumbel; when the shape parameter is positive, it is a type II GEV/ Fréchet; when it is negative, a type III GEV/Weibull (Figure 5).

In the BM approach it is essential to select an appropriate block size: too small or too large intervals may lead to a poor fit (Gecchele et al., 2019). If blocks contain enough observations, r-largest order statistic is usually recommended, as it makes a more efficient use of data: in that approach, the estimation dataset is composed not just by the block maxima, but by the r-largest values in each block.

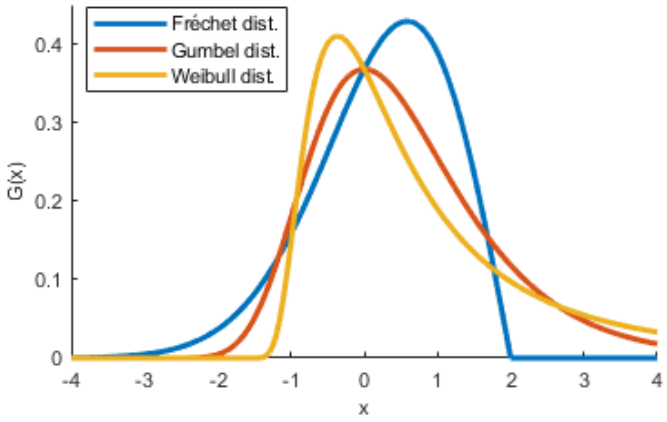

Figure 5. Examples of GEV distributions, characterized by different values of the shape parameter

2.2.1.2 Peak-Over-Threshold approach and GP distribution

In the POT approach, after a certain threshold has been defined, all observations with higher values, usually called exceedances, are considered extremes and sampled.

For independently and identically distributed random observations $\{X_1, X_2, \ldots, X_n\}$, the distribution function of exceedances X over a threshold u is $F_u(x) = \Pr(X - u \leq x | X > u)$. When threshold u is sufficiently high, the conditional distribution $F_u(x)$ can be approximated by a GP distribution. The cumulative GP distribution function is:

$$GP(x) = 1 - \left(1 + \xi \frac{x - u}{\sigma}\right)^{-\frac{1}{\xi}} \quad (2)$$

where $\sigma > 0$ the scale parameter, and $-\infty < \xi < \infty$ the shape parameter (Figure 6).

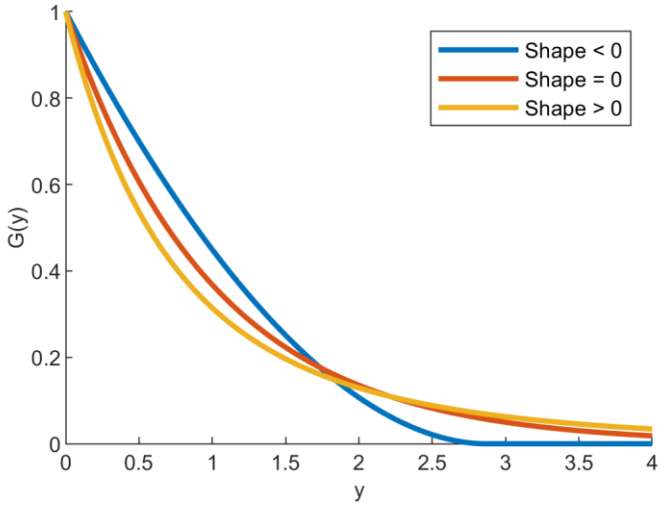

Figure 6. Examples of GP distributions, characterized by different values of the shape parameter

In this approach, threshold selection is crucial: if it is too low, this may result in sampling events which are not really extremes and lead to a poor fit; if it is too high, this may result in an overall sample size with not enough extremes which can lead, again, to a poor fit.

The threshold also has a physical meaning in road safety analysis, as it represents, intuitively, the value which discriminates between conflicts and normal interactions between vehicles. This separation between normal interactions and conflicts, however, is not straightforward: if the threshold is too low it may lead to select, along with genuine traffic conflicts, also aggressive but controlled interactions; therefore, a simpler approach is to define a threshold high enough to include only conflict events (Tarko, 2018b). Considering only severe conflicts inevitably increases the observation period needed to collect sufficient data for statistical analysis, but it simplifies the theory and strengthens the crash causality.

Several techniques to select the threshold exist; a comprehensive overview can be found in Scarrott & MacDonald (2012). The simplest and most used technique is the graphical diagnostics method, in which the GP distribution is fitted several times on the data, progressively increasing the threshold; the best threshold is then chosen by visually inspecting mean excess and stability plots (Coles, 2001). The main drawback of this technique is that it requires substantial expertise and can be rather subjective; moreover, it is quite time-consuming, so it is not easily applicable when several models have to be estimated. The graphical diagnostics method has been adopted in most of EVT applications in the field of road safety (Farah & Azevedo, 2017; L. Zheng et al., 2014a). L. Zheng et al. (2014b) used Bayesian inference to estimate the threshold along with the other model parameters; this procedure is however more complex and computationally demanding to implement.

A computationally inexpensive method which can be used to select in an automated way the optimal threshold, without visually analyzing the models one-by-one, is the Automatic Threshold Selection Method (ATSM), proposed by Thompson et al. (2009), and applied in several recent EVT works (Collet et al., 2017; Kumar & Kalyani, 2017; Postacchini et al., 2019). The basic idea of the ATSM is that, given an equally spaced candidate thresholds $u_1, \dots, u_n$, given one of these candidate thresholds u_k, fitted the GP parameters σ and ξ for each candidate threshold u_j higher than u_k, and defined:

$$\tau_{u_j} = \sigma_{u_j} - \xi_{u_j} u_j, \;\; j = k+1, \dots, n \tag{3}$$

it follows from (Coles, 2001) that, if u_k is a suitable threshold, then $E[\tau_{u_j} - \tau_{u_{j-1}}] \approx 0$ and $\tau_{u_j} - \tau_{u_{j-1}}$ is normally distributed, for any *j* such that $u_k \leq u_{j-1} < u_j$.

The procedure therefore is composed by the following steps:

1. The set of equally spaced candidate thresholds $u_1, \dots, u_n$ is defined.
2. For each candidate threshold, $\tau_{u_j} - \tau_{u_{j-1}}$ are computed (for any *j* such that $u \leq u_{j-1} < u_j$).
3. For each candidate threshold the Pearson's Chi-square Test is used as a goodness-of-fit to establish whether the set of differences $\tau_{u_j} - \tau_{u_{j-1}}$ is normally distributed with mean = 0.
4. If the null hypothesis is rejected it means that the candidate threshold is not a suitable threshold; otherwise, it is.

In the end, the method provides a set of suitable thresholds within the original set of candidate thresholds. This method aims to reduce the subjectivity in the threshold choice and the time required to select the threshold. Indeed, the procedure is computationally inexpensive, therefore it is suited for large-scale applications of EVT; analyst subjectivity is greatly reduced, but not completely eliminated: the choice of the initial set of candidate thresholds can influence the results, and, at the end of the procedure the analyst has still to choose among a (reduced) set of thresholds.

2.2.2 Bivariate extreme value theory

The focus of this Section is on bivariate EVT, which is used to model the joint distribution of two extreme variables. As in the univariate case, there are two classic methods, which, using the terminology of Beirlant et al. (2004) are: the Component-wise Maxima approach

(CM) and the Excesses Over a Threshold (EOT), which can be regarded as bivariate extensions of the univariate BM and POT approaches, respectively.

In short, in the CM approach the observation period is divided into blocks and, for each block, the maxima of both variables are included in a dataset used to estimate the parameters of a bivariate GEV distribution. In the EOT approach a threshold for each variable is selected and all the values above each threshold are included in a dataset used to estimate the parameters of a bivariate GP distribution. The following Sub-Sections provide a more rigorous, yet concise, presentation of the two approaches.

2.2.2.1 Component-wise Maxima approach

Let us consider a temporal sequence of vectors $(X_1, Y_1), \dots, (X_i, Y_i), \dots, (X_n, Y_n)$, which are independent versions of a random vector having distribution function $F(x, y)$. Similarly, to the univariate case, extremal behavior of bivariate (and, more in general, multivariate) extremes is based on the limiting behavior of block maxima. In the bivariate case the vector of Component-wise Maxima is defined as $M_n = (M_{X,n}, M_{Y,n})$, where $M_{X,n} = \max_{i=1,\dots,n} \{X_i\}$ and $M_{Y,n} = \max_{i=1,\dots,n} \{Y_i\}$. The index *i* for which the maxima of the *X* and *Y* occur do not need to be the same for the two variables; in other words, the CM vector does not necessarily correspond to an observed vector in the original series. Considered separately, $\{X_i\}$ and $\{Y_i\}$ are sequences of independent univariate random variables, and univariate EVT results can be applied to both components. As a result, assuming that they have standard Fréchet marginal distribution (Theorem 8.1 in (Coles, 2001)), if:

$$Pr\{M_{X,n} \leq x, M_{Y,n} \leq y\} \xrightarrow{d} G(x, y) \tag{4}$$

where *G* is a non-degenerate distribution function, which has the form:

$$G(x, y) = exp\{-V(x, y)\}, \;\; x > 0, y > 0 \tag{5}$$

where:

$$V(x, y) = 2 \int_0^1 max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w) \tag{6}$$

and H is a distribution function on [0,1], satisfying the mean constraint:

$$\int_0^1 w\, dH(w) = \frac{1}{2} \tag{7}$$

One issue is that an infinite number of parametric forms of G, satisfying the conditions in (6) and (7), exist. A possible solution is to use parametric sub-families for H, leading to sub-families of distributions for G. Several parametric families have been used in the literature: logistic, asymmetric logistic, negative logistic, asymmetric negative logistic, bilogistic, negative bilogistic and Husler-Reiss; a convenient list of the functional forms of these families can be found in L. Zheng, Ismail, et al. (2018); one of the most used is the logistic family, whose expression is:

$$G(x,y) = exp\left[-\left(x^{1/r} + y^{1/r}\right)^r\right] \tag{8}$$

where $0 < r \leq 1$ is a parameter to be estimated; when $r = 1$ the two variables are independent, whereas as $r \to 0$, there is perfect dependency.

Results from the previous theorem can be used in order to estimate the model parameters. Let us consider now a sequence of CM $\left(z_{1,j}, z_{2,j}\right)$ (j=1,..m, where m is the total number of blocks), whose components Z_i (i=1,2) are in fact sequences of independent block maxima, which can be modelled using the GEV distribution, defined by the set of parameters (μ_i, σ_i, ξ_i). Applying maximum likelihood to the separate series generates estimates of these parameters $(\hat{\mu}_i, \hat{\sigma}_i, \hat{\xi}_i)$. The transformed variable

$$\tilde{Z}_i = \left[1 + \hat{\xi}_i \left(\frac{Z_i - \hat{\mu}_i}{\hat{\sigma}_i}\right)\right]^{\frac{1}{\hat{\xi}_i}} \tag{9}$$

is therefore approximately distributed according with the standard Fréchet distribution. The pairs $\left(\tilde{z}_{1,j}, \tilde{z}_{2,j}\right)$ are a sequence of independent realizations of a vector with bivariate EV distribution within the family G, with probability density function $g(x,y)$ in which V has a parameter θ provided by the family G (e.g., if the family is logistic, θ is equal to the parameter r). This leads to the likelihood:

$$L(\theta) = \prod_{i=1}^{m} g\left(\tilde{z}_{1,j}, \tilde{z}_{2,j}\right) \tag{10}$$

and standard techniques yield maximum likelihood estimates and standard errors of parameters.

2.2.2.2 Excesses Over a Threshold

As in the previous paragraph, we consider a sequence of vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$, independent versions of a random vector having distribution function $F(x, y)$. Given suitable thresholds u_x and u_y, the marginal distributions of *F* are an approximation of the univariate Generalized Pareto distribution, with respective parameter sets (σ_x, ξ_x) and (σ_y, ξ_y). The distribution function $\tilde{F}$ of transformed variable $(\tilde{X}, \tilde{Y})$ has margins that are approximately standard Fréchet distribution for $X > u_x$ and $Y > u_y$, and, given equation (5), if the thresholds are large enough, it is possible to assume (see (Coles, 2001) page 155 for demonstration) that:

$$F(x, y) \approx G(x, y) = exp\{-V(\tilde{x}, \tilde{y})\}, \;\; x > u_x, y > u_y \tag{11}$$

where *V* is still defined by equation (6) and constrained by (7). Again, it is possible to choose among several parametric families to define *G*, for example the logistic (8).

In the EOT method a bivariate pair of variables may exceed the threshold in just one of its components. The plane (*x*,*y*) can be divided into four regions: $R_{0,0}$, $R_{1,0}$, $R_{0,1}$ and $R_{1,1}$. In region $R_{0,0}$ both x and y are below their respective thresholds; in regions $R_{1,0}$ and $R_{0,1}$ only one component is above the threshold; in region $R_{1,1}$ both are above the threshold. While in the $R_{1,1}$ region model (11) is applicable, in the other regions it is necessary to censor the likelihood component; for example, if (*x*,*y*) belongs to $R_{1,0}$ there is information on the marginal *x*-component (since $x > u_x$), but not for the *y*-variable; therefore the likelihood for a point in that region is: $Pr\{X = x, Y \leq u_y\} = \frac{\partial F}{\partial x}\Big|_{(x,u_y)}$.

The resulting likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} \psi\left(\theta; (x_i, y_i)\right) \tag{12}$$

where ψ depends on the region of the plane:

$$\psi\left(\theta; (x_i, y_i)\right) = \begin{cases} F\left(u_x, u_y\right) & if\ (x,y) \in R_{0,0} \\ \left.\dfrac{\partial F}{\partial x}\right|_{(x,u_y)} & if\ (x,y) \in R_{1,0} \\ \left.\dfrac{\partial F}{\partial y}\right|_{(u_x,y)} & if\ (x,y) \in R_{0,1} \\ \left.\dfrac{\partial^2 F}{\partial x\, \partial y}\right|_{(x,y)} & if\ (x,y) \in R_{1,1} \end{cases} \tag{13}$$

and, once again, maximum likelihood estimates and standard errors of parameters can be obtained with standard techniques.

2.3 Proof of concept 1: application of univariate EVT[1]

In this Section, the application of univariate EVT on data obtained from a driving simulator study is presented as a proof-of-concept to illustrate the methodology. The work presented this Section has also significant scientific relevance, because:

- It is one of the few that directly compares BM and POT approaches, and the literature review highlighted the need of more comparative studies, since the existing ones have reached somewhat inconsistent conclusions (see Section 2.1)
- It is the first one applying EVT to a case study involving roundabouts

Roundabout is a type of intersection control which has been shown to provide safety and operational benefits compared with other types of traffic control. Studies demonstrate that roundabouts can reduce delays, queue lengths, congestion, and degree of saturation, with lower power and maintenance costs (Burdett et al., 2017). The geometric features of roundabouts, which reduce the number of conflict points, have been shown to reduce fatal crashes and injuries when converted from other intersection control types (Elvik, 2003, 2017). Although the reduction in crashes is a positive step toward safer intersections, there is still room for improvement and an increased need for detailed analysis of their safety performance, as more roundabouts will be constructed in the future.

An NCHRP study of 39 roundabouts throughout the United States (Rodegerdts et al., 2007) found entering–circulating, exiting–circulating, and rear-end crashes to be the most frequent crash types for single-lane roundabouts. D. Zheng et al. (2010) carried out a comprehensive analysis of collision patterns at Wisconsin roundabouts, and found that 40% were entering-circulating collisions at single-lane roundabouts. Montella (2011) identified contributory crash factors for 15 urban roundabouts in Italy; in single-lane roundabouts, there was a greater proportion of crashes at the entry (68%), mainly due to entering–circulating conflicts. More recently, Polders et al. (2015) identified entering-circulating conflicts as one of the four dominant crash types in roundabouts, which account for 75% of all roundabouts collisions.

[1] The work presented in this Section is part of the following publication:

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2019). Collision prediction in roundabouts: a comparative study of extreme value theory approaches. *Transportmetrica A: transport science*, *15*(2), 556-572. (Orsini et al., 2019)

2.3.1 Driving Simulator experiment

Data used for analysis were obtained with a driving simulator. The simulation system is a dynamic-base driving simulator produced by STSoftware®, comprising a realistic cockpit, three networked computers, and five high-definition screens. It also has a Dolby Surround® sound system, the whole producing realistic virtual views of the road and the surrounding environment. The driving simulator was previously validated in multiple conditions, including gap-acceptance behavior at intersections (Rossi et al., 2011), and speed behavior in approaching bends (Rossi et al., 2014) and roundabouts (Rossi et al., 2018, 2020; Rossi, Gastaldi, et al., 2013).

The study site analyzed here is a single-lane roundabout on the urban road network of Noventa Padovana, in the province of Padova, Italy. The roundabout was built in virtual reality with 3D rendering software, and a circuit was created to allow drivers to repeat the entry maneuver at the roundabout several times in the same experiment. The circulating flow in the roundabout was composed by vehicles generated with headways measured in the real-world, following a pre-determined trajectory, without adapting their speed to the presence of the entering vehicle. Day time and good weather conditions, which allow good visibility, were adopted in this scenario.

The parameters of driver behavior provided and recorded by the simulator included global position, velocity, acceleration of all vehicles involved, and cabin parameters (frequency 10 Hz). For the purposes of this study, the minimum TTC surrogate measure of safety while entering the roundabout was also examined.

2.3.1.1 Time-To-Collison definition

In the research on traffic conflicts, Time-To-Collision (TTC) has proven to be an effective measure for evaluating the severity of conflicts (Van Der Horst & Hogema, 1993) and as a surrogate measure of safety (Gettman & Head, 2003). TTC is defined as "the time required for two vehicles to collide if they continue at their present speed and on the same path" (Hayward J.C., 1972).

Traditionally, at a given time *t*, TTC is computed as:

$$TTC = \frac{R}{RR} \tag{14}$$

where R (range) is the space separation between the vehicles, and RR (range rate) is the speed difference between the vehicles. Speeds are considered at time *t* and assumed constant.

In the present case, it is reasonable to assume constant the speed of the circulating vehicle, which proceeds on its trajectory regardless of the presence of the tester's entering vehicle. However, the same cannot be assumed for the entering vehicle. Therefore, a TTC equation which includes acceleration must be formulated. It is not unusual to take into account acceleration of road users in TTC; this was first introduced by Van Der Horst (1990) and it is often applied in rear-end conflicts, when vehicles are braking (Kusano & Gabler, 2011). In this roundabout entering-circulating conflict, instead, one of the vehicles is accelerating, a situation similar to one of those defined by Ito et al. (2017)

Considering the situation illustrated in Figure 7, the position of the circulating vehicle's front bumper at time *t* is:

$$x_1 = x_{1,0} + v_1 t \tag{15}$$

While the position of the entering vehicle rear bumper is:

$$x_2 = x_{2,0} + v_2 t + \frac{1}{2} a_2 t^2 \tag{16}$$

The condition for the collision to occur is $x_1 = x_2$. At time *t*, TTC is therefore obtained substituting (15) in (16):

$$TTC = \frac{RR - \sqrt{RR^2 - 2aR}}{a} \tag{17}$$

Where $R = x_{2,0} - x_{1,0}$ and $RR = v_1 - v_2$. The acceleration *a* of the entering vehicle is assumed constant. The two vehicles are on a collision course only if $RR^2 \geq 2aR$ and $TTC > 0$.

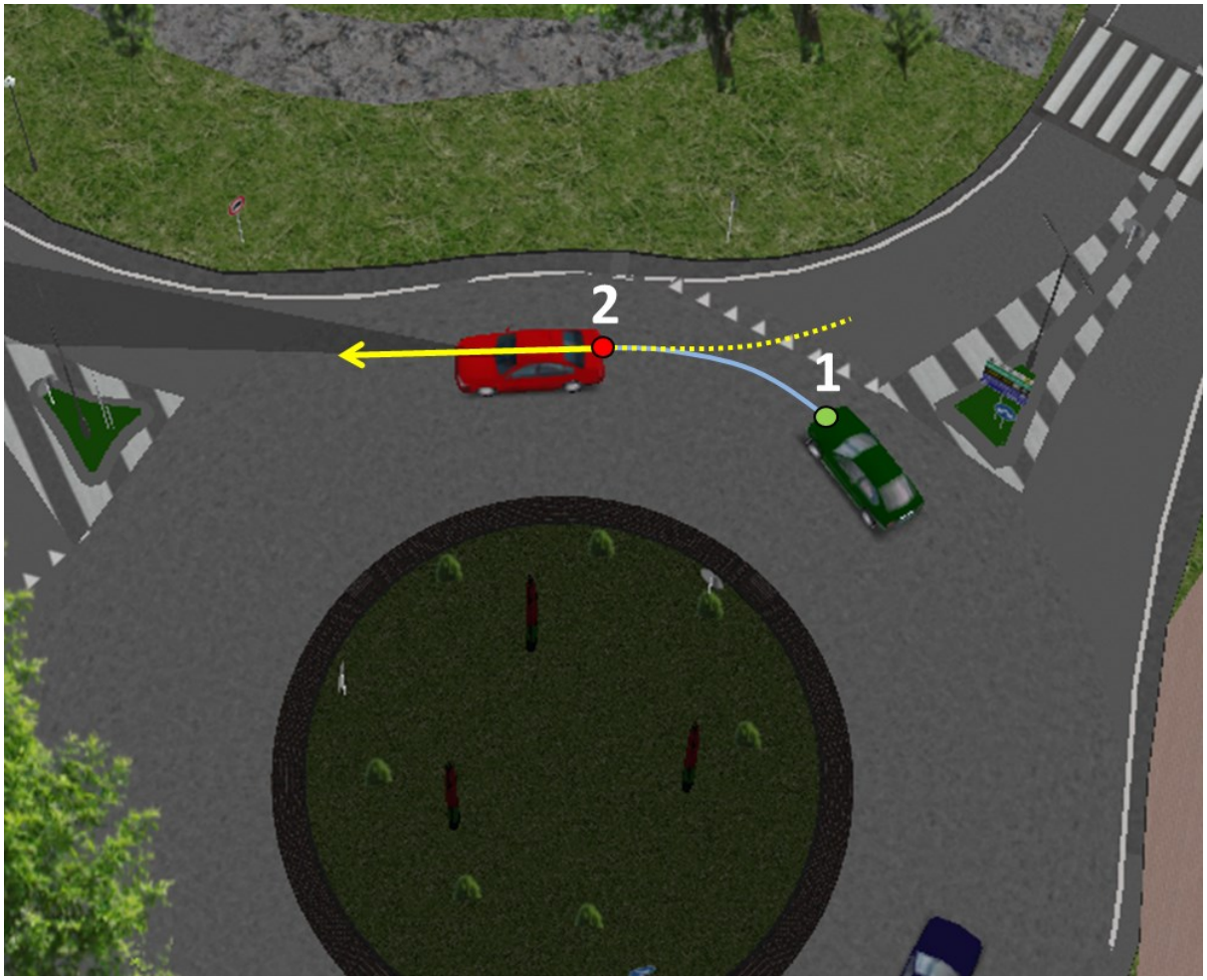

Figure 7. Entering-circulating conflict in the driving simulator experiment. Car number 1 is the computer-controlled circulating vehicle; car number 2 the tester-driven entering vehicle.

2.3.1.2 Participants and data

Sixty drivers (15 women, 45 men) were asked to drive along the road circuit. Participants' age ranged from 19 to 27, with an average of 23 years; this choice was suggested by past studies (ETSC, 2016), indicating that young drivers (< 25 years), particularly men, are at higher risk of accidents than more experienced, older drivers. All participants had regular driving licenses for at least two years, drove a minimum of 3000 km/year, and did not have any previous experience with driving simulators. Vision was normal or corrected-to-normal.

Participants were tested individually, after being properly instructed about the task. They were asked to drive along the circuit as they would normally do in the real world, respecting existing speed limits, and were given 10 minutes to become familiar with the simulator.

The familiarization scenario contained several intersections (4 roundabouts, 3 signalized and 1 non-signalized intersections) in urban and sub-urban context.

During the experiment, each driver completed 3 laps, approaching the roundabout 4 times in each of them. The basic roundabout, with the same geometric features of the real-world roundabout, was therefore repeated 12 times. For each of them, the circulating traffic in front of the approach reflected real-world observed sequences of vehicles (in terms of type and time headways). The experiment was designed so that the pattern of the circulating flow was selected randomly (among 500 possible real-world sequences) in each roundabout for each tester.

The resulting dataset gave 537 maneuvers in which the entering (driven by the tester) and circulating (controlled by the software) vehicles were on a collision course; 5 of these maneuvers ended with a rear-end collision, following the scheme presented in Figure 7, due to the tester's incorrect evaluation of the gap between two consecutive circulating vehicles, in relation with the minimum gap necessary to perform the maneuver safely. These observations were removed from the estimation datasets. The collision rate is relatively high compared to a real-world roundabout. This was caused by the circulating vehicles' behavior, which ignored the presence of the tester's entering vehicle, therefore not slowing down or performing evasive maneuvers to avoid contact. Setting this kind of behavior in the computer-controlled circulating vehicles was intentional, in order to obtain a number of incidents higher than zero during the experiments, to be compared with the model's estimation.

2.3.2 Results

The data obtained from the case-study were analyzed with various approaches: the Block Maxima using the GEV model and the Peak-Over-Threshold using the GP model. In the latter approach both negated-TTC (nTTC) and shifted-reciprocal-TTC (srTTC) were analyzed. GEV and GP distributions were fitted with MATLAB (Gilat, 2014), using maximum likelihood (ML) estimation method and considering only stationary conditions.

2.3.2.1 Block Maxima (BM)

In this analysis the block was defined as the maneuver of entering the roundabout, so there were as many blocks as the number of completed maneuvers observed in the experiment. In each block the maximum negated-TTC was considered.

Similarly to Farah & Azevedo's approach (2017), maxima were pre-processed, in order to eliminate measurement that were not useful as crash surrogates. A sensitivity analysis was carried out to assess limit value effect on parameters and crash estimations. The GEV was

fitted multiple times to the measurements above the limit, the latter varying between -2.00 s and -0.15 s. After the stability of GEV parameters had been plotted (Figure 8a), all limits above -0.80 s were excluded, due to the erratic trend observed. Among the remaining limits, the one which maximized the number of estimated collisions was chosen, for safety reasons (Figure 8b); this limit was -1.37 s.

After the original 532-maxima dataset had been filtered, 227 records had a negated-TTC above -1.37 s. The GEV model was fitted to these measurements and resulted in the parameters and respective standard errors shown in Table 2. Figure 11a presents the kernel probability density function of the empirical and modeled negated-TTC, and Figure 11b presents the simulated QQ plot. From these figures it can be concluded that the modeled distribution has satisfactory fitting results, since the points fall close to the 45° line in the simulated QQ plot.

Collision frequency was defined as the cumulative GEV probability of negated-TTC being higher than or equal to 0. The predicted collision number was then calculated as the product of this probability and 232 (the number of records above the limit, including actual collisions). This resulted in 5.32 predicted crashes, with a 95% confidence interval of [1.20; 10.40].

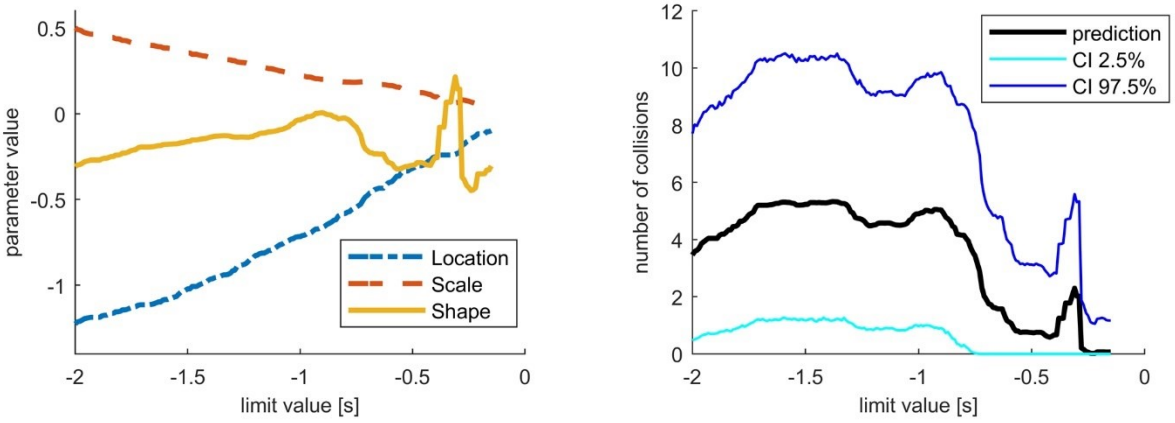

Figure 8. Block Maxima graphs: (a) parameter stability plot; (b) estimated collisions changing the "limit" value.

The confidence interval was computed with the same simulation-based inference method presented by Songchitruska & Tarko (2006). Collision frequency is a scalar function of the model parameters, which, under regularity conditions, can reasonably be assumed to follow the multivariate normal distribution. The parameter values were generated 10^6 times and the collision frequency was computed each time, giving its empirical distribution; the confidence interval was then calculated from the quantiles of the empirical distribution,

e.g., for the 95% confidence interval, the 2.5%-quantile was taken as lower bound and the 97.5%-quantile as upper bound.

2.3.2.2 Peak-Over-Threshold with negated-TTC (POT-nTTC)

In the Peak-Over-Threshold (POT) approach all the observed 532 collision-free maneuvers in the driving experiment were analyzed (Figure 9c). For each of these maneuvers the negated-TTC was examined, as usually found in literature (Farah & Azevedo, 2017; Tarko, 2012; L. Zheng et al., 2014a).

In the POT approach the first step is to determine the threshold. To do so, visual analysis of mean excess (Figure 9a) and stability (Figure 9b) plots was carried out. As a compromise between bias and variance, the threshold should be the higher possible value (L. Zheng et al., 2014a), provided that mean excess plot is linear and parameters are stable (Coles, 2001). In the mean excess plot, linearity is clearly maintained; in the stability plot, scale parameter and, more evidently, shape parameter can be observed to be relatively stable until -0.73 s. A -0.73 s threshold was therefore set and the number of exceedances was 94.

A GP distribution was fitted to the negated-TTC dataset, with a threshold of -0.73 s. The estimated parameters and their standard errors are shown in Table 2. A negative value of the shape parameter is consistent with the negative slope in the mean excess plot in Figure 9a (B. Das & Ghosh, 2016). Figure 11c and Figure 11d present the probability density function of the empirical and modeled negated-TTC and the simulated QQ plot for the best fitted model.

As in section 2.3.2.1, the collision frequency was defined as the GP cumulative probability of negated-TTC being higher than or equal to 0. This probability was then multiplied by the total number of maneuvers, including crashes, observed during the experiment, which was 537. This resulted in a predicted number of crashes of 0.00, with a 95% confidence interval of [0.00;2.46].

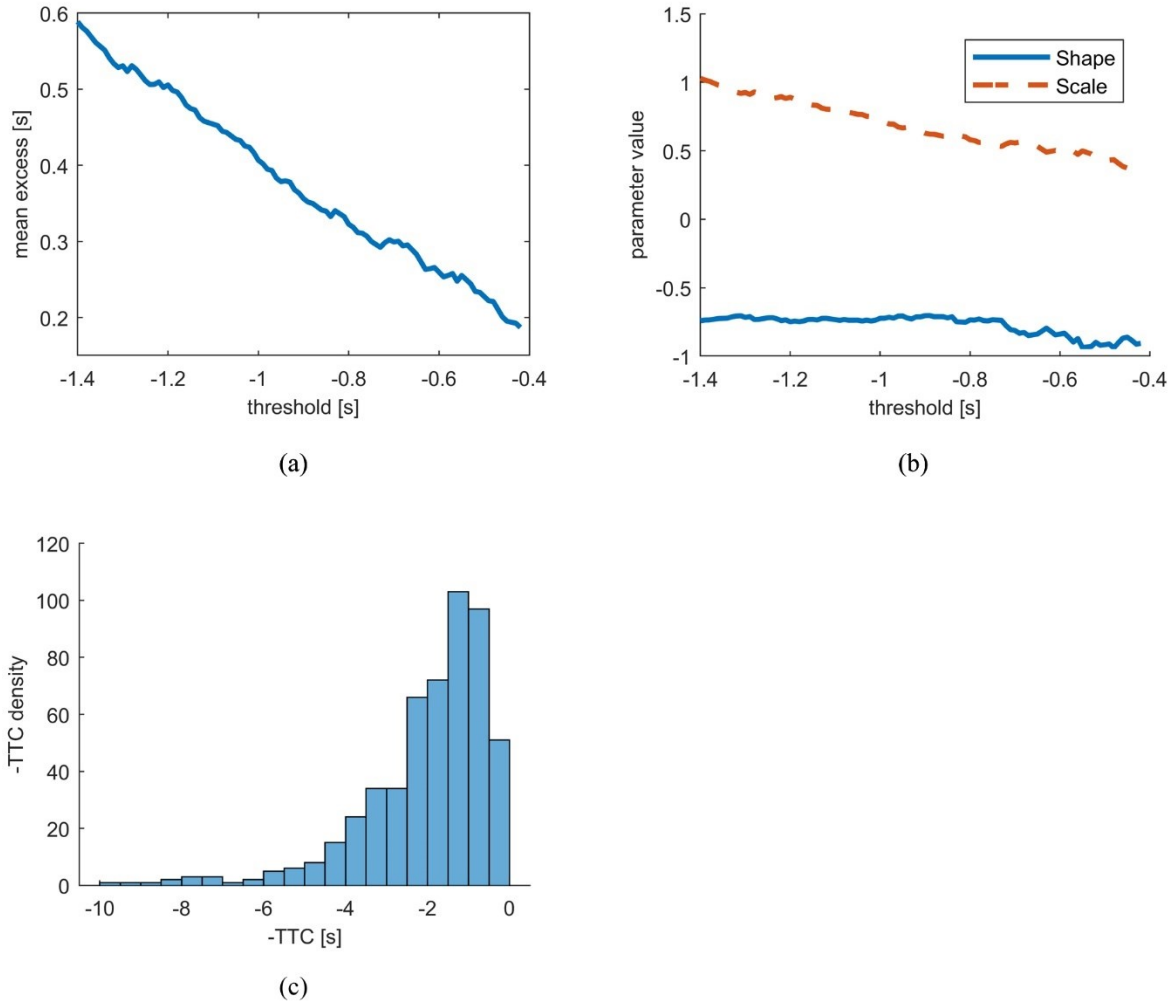

Figure 9. POT-nTTC graphs: (a) Mean excess plot; (b) Stability plot; (c) negated-TTC histogram.

2.3.2.3 Peak-Over-Threshold with shifted-reciprocal TTC (POT-srTTC)

Another POT approach was carried out, following a procedure derived from the work of L. Zheng et al. (2014b). The original positive TTC values from the 532 collision-free maneuvers were moved by a value of $\delta>0$ and the reciprocal were then computed. Therefore, the observed value of interest became $1/(TTC+\delta)$ for each maneuver. The GP distribution was to be fitted on these values (Figure 10d).

In this case, unlike the traditional POT approach, not only the threshold, but also the delta value, had to be set. An iterative procedure was followed: 1) a starting delta value was chosen; 2) the GP model was fitted multiple times changing progressively the threshold value; 3) mean excess and stability plot were evaluated to find the optimal threshold value; 4) the GP model was fitted multiple times, with the threshold value fixed, changing the delta value progressively; 5) a new delta value was selected in a similar way with respect to

the "limit" value in the BM approach, choosing the value that maximized the number of collision estimation, for safety reasons. The process continues until the new delta value is equal to the delta value obtained at the previous iteration.

This procedure represents an original contribution of this work. Selecting the delta value which maximizes crash prediction is an objective rule and allows the analyst to work without crash data. L. Zheng et al. (2014b) adopted a crash-based approach to choose this value, and selected the one that reduced the root-mean-square deviation of crash estimations in their sections of interest.

At the end of the procedure a 1.26 s^{-1} threshold (which corresponds to a TTC of 0.75 s) and a 0.07 s delta value were chosen; this resulted in 93 exceedances. Figure 10a shows the mean excess plot: a linear positive trend can be identified until 1.26 s^{-1}. In Figure 10b it can be observed that shape and scale parameters have stable values until 1.26 s^{-1}. Figure 10c shows how collision estimation changes with delta. The parameters and respective standard errors are shown in Table 2; as expected the shape parameter has a positive value, since the mean excess plot had a positive slope (B. Das & Ghosh, 2016). Figure 11e and Figure 11f present the probability density function and the simulated QQ plot for the estimated model. Both figures indicate a good fit between the modeled distribution and the empirical data.

The collision probability was then computed. In this case it was defined, as in L. Zheng et al. (2014b), as the GP cumulative probability of $1/(TTC+\delta)$ being higher than or equal to $1/\delta$, which is equivalent to TTC being lower than or equal to 0. As in section 2.3.2.2, collision probability was then multiplied by 537, resulting in 7.45 predicted crashes, with a 95% confidence interval of [0.41;19.79].

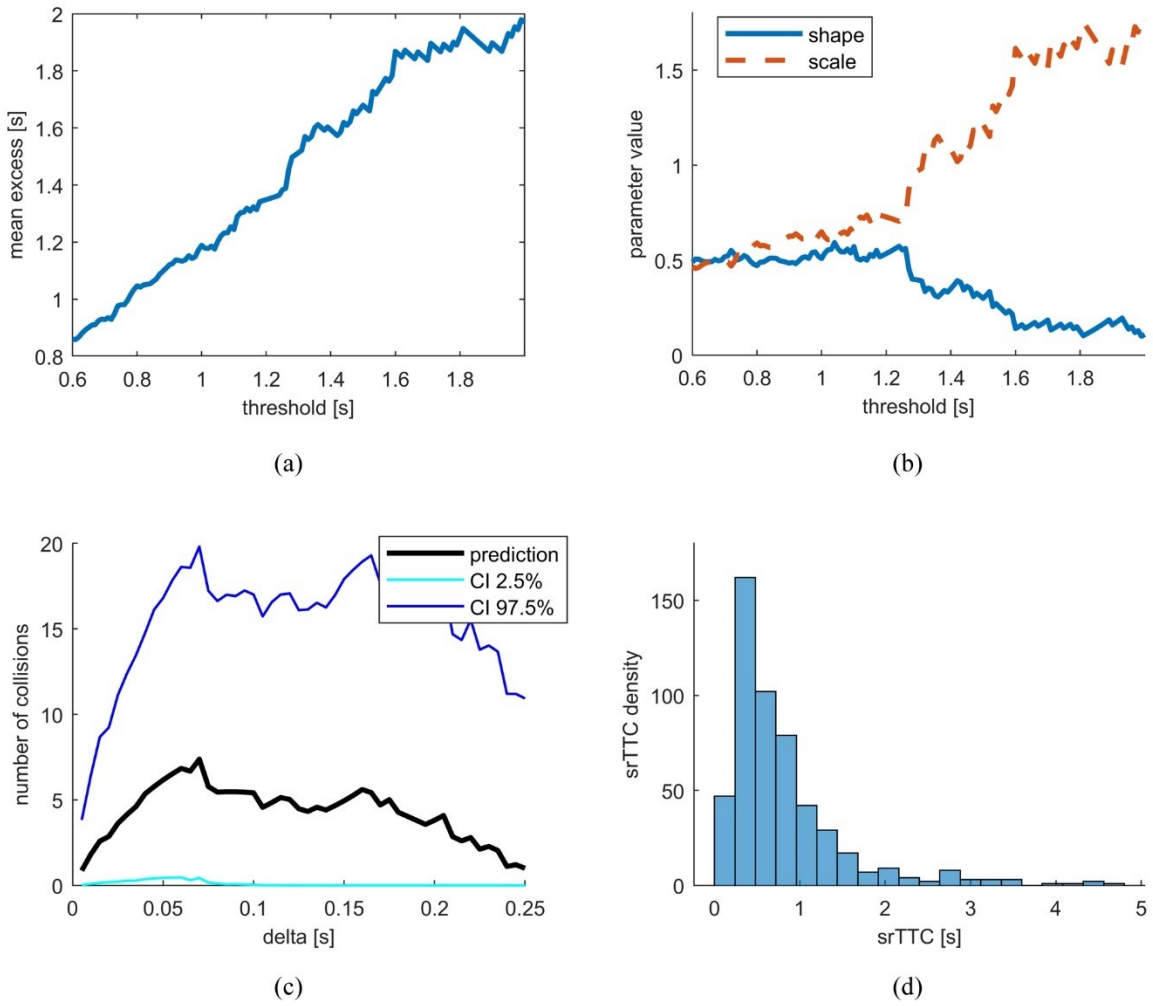

Figure 10. POT-srTTC graphs: (a) Mean excess plot when delta = 0.07 s; (b) Stability plot when delta = 0.07 s; (c) estimated collisions changing the δ value, when threshold = 1.26 s^{-1}; (d) shifted-reciprocal-TTC histogram

Table 2. Estimation results for the best fitted models.

Approach	**Samples [#]**	**Location (SE)**	**Scale (SE)**	**Shape (SE)**	**Threshold**	**Neg. Log-Likelihood**
BM	227	-0.9536 (0.0259)	0.3209 (0.0198)	-0.1308 (0.0758)	-	85.7943
POT-nTTC	94	-	0.5315 (0.0964)	-0.7381 (0.0632)	-0.73	-34.7944
POT-srTTC	93	-	0.7288 (0.1575)	0.5611 (0.1984)	1.26	115.7524
(-) = not applicable						

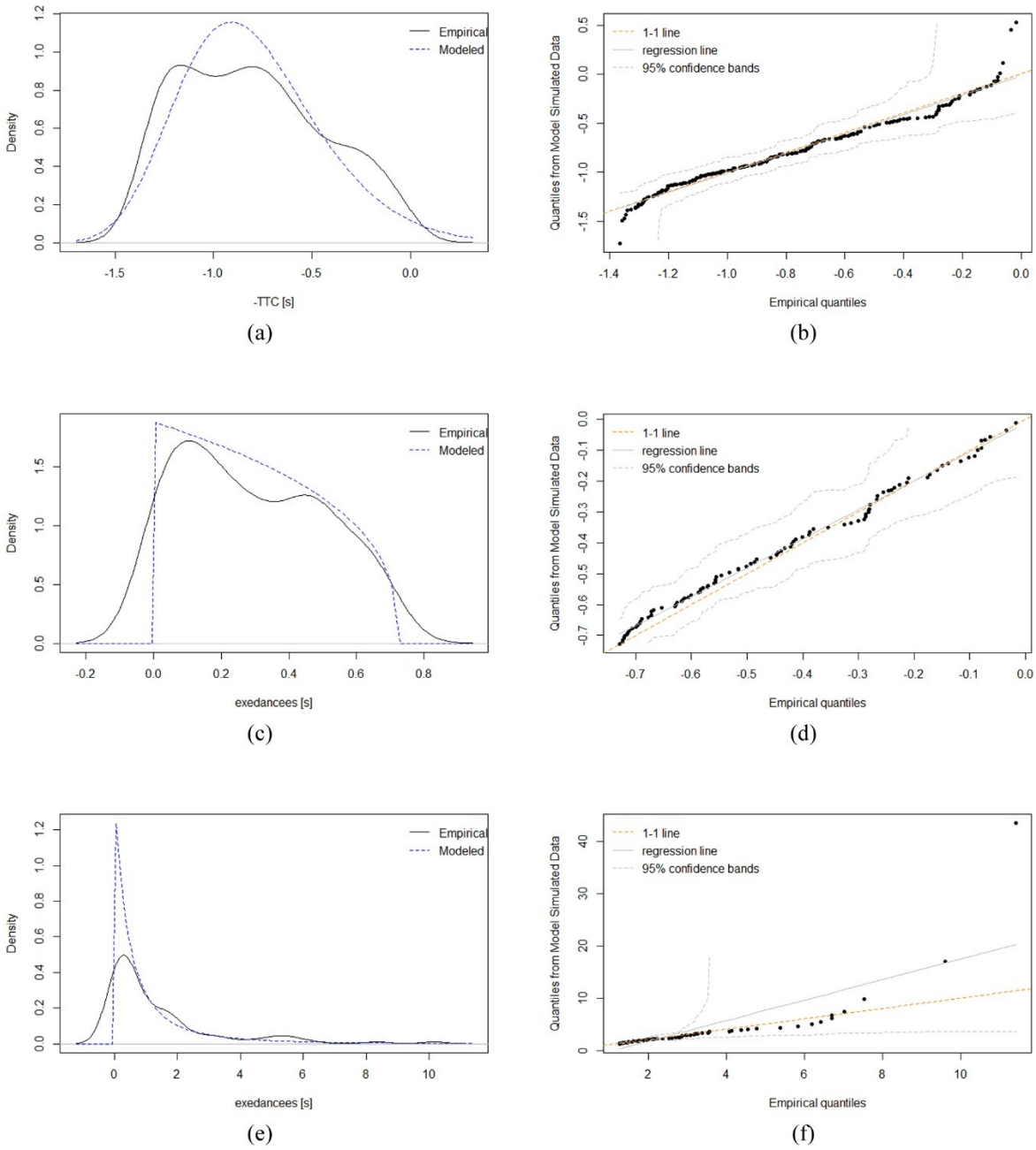

Figure 11. Model fitting for models: (a) BM approach - Kernel probability density plot; (b) BM approach – Simulated QQ plot; (c) POT-nTTC approach - Kernel probability density plot; (d) POT-nTTC approach – Simulated QQ plot; (e) POT-srTTC approach - Kernel probability density plot; (f) POT-srTTC approach – Simulated QQ plot.

2.3.3 Summary and discussion

The results of the three approaches described above were compared with respect to the actual number of collisions observed during the driving simulator experiment, which was five.

Table 3 lists the 95% confidence intervals for the three approaches. The approach that gives the worst results is the POT-nTTC, since the observed number of collisions is always outside the confidence intervals.

In both BM and POT-srTTC approaches the observed number of crashes is well inside the predicted crashes confidence interval. However, it should be noted that the latter approach has a larger confidence interval, and BM is the approach which actually gives a predicted number of crashes closer to the observed value. The quality of the collision estimations is comparable to, and in some cases better than, similar works in transportation research (L. Zheng et al., 2014a; Farah & Azevedo, 2017).

Another comparison can be made taking into account not only the observed value of collisions, but also its binomial confidence interval (Table 3). The idea is that the same experiment with the same participants could have produced a different number of crashes. A similar comparison was made by Farah & Azevedo (2017) in their work. It is particularly interesting to see that the observed crashes confidence interval is completely contained by the POT-srTTC and almost completely by the BM confidence intervals. For the POT-nTTC approach the two intervals marginally overlap.

Table 3. Comparison of results for the best fitted models.

Approach	**Collision frequency [CI]**	**Samples [#]**	**Predicted crashes [CI]**	**Observed crashes [CI]**
BM	0.0229 [0.0052;0.0448]	232	5.32 [1.20;10.40]	5 [1.63;11.50]
POT-nTTC	0.0000 [0.0000;0.0046]	537	0.00 [0.00;2.46]	5 [1.63;11.60]
POT-srTTC	0.0139 [0.0008;0.0368]	537	7.45 [0.41;19.79]	5 [1.63;11.60]

Although the traditional POT approach does not provide satisfactory results, BM and POT-srTTC performed very well. Among these two, the BM approach can produce the closest collision prediction to the actual observed value, and it is also the most efficient, having a confidence interval much narrower than the POT-srTTC interval.

The main difficulty with the BM approach is determining the limit. Ideally this should be the physical limit under which there is virtually no risk of collision; while, e.g., for head-on

collision there are well-recognized indications from the literature (Hydén, 1987; Jonasson & Rootzén, 2014), the same cannot be said about roundabout collisions. The approach used in this Section aimed at reducing the subjectivity of the choice of this limit, by selecting the value which provides the highest collision estimation; this represents an innovative contribution to the existing methodology.

POT-srTTC is a much less established approach in the literature, having been introduced only recently (L. Zheng et al., 2014b), but appears to be promising and, in the present case-study, actually performed significantly better than the traditional POT-nTTC method. From a statistical viewpoint this is because the distribution of the shifted-reciprocal-TTC has a much thinner right-hand tail, than that of the negated-TTC (see Figure 9c and Figure 10d); under a physical viewpoint, because this reciprocal transformation amplifies the sensitivity of the model to small TTC values, which are more reliable as crash surrogates. This case-study suggest that POT-srTTC approach should be preferred to the POT-nTTC approach when the negated-TTC distribution has a negative skew.

As in the case of L. Zheng et al. (2014b), confidence intervals for POT-srTTC are large when ML estimation is used. The above authors were able to reduce uncertainty in estimations by applying the Bayesian approach. Applying the Bayesian approach to this case study will be investigated in the future, as it also has the advantage of eliminating subjectivity and uncertainty in threshold selection.

It is worth noting that in neither of the two POT approaches applied here was a declustering operation adopted: observations were reasonably assumed to be completely independent, in view of the characteristics of the experiment.

2.4 Proof of concept 2: application of bivariate extreme value theory[2]

In this Section, the application of bivariate EVT on data obtained from a driving simulator study is presented as a proof-of-concept to illustrate the methodology. The work presented in this Section has also significant scientific relevance, because:

- Unlike previous works, both CM and EOT methods were applied and compared (see Section 2.1)
- To the best of authors' knowledge, the safety of left-turn from minor road at unsignalized intersection was never analyzed with EVT before

The present study deals with the safety analysis of a three-leg unsignalized intersection (see Figure 15a), specifically concerning the left-turn maneuver from the minor road. The bivariate EVT approach was applied under two different points-of-view: (i) considering one surrogate measure of safety and the two conflict points (one for each lane of the major road); (ii) considering one conflict point and two different surrogate measures of safety. In both cases the bivariate analysis was performed applying two methods: Component-wise Maxima (CM) and Excesses Over a Threshold (EOT).

The aim of the study is to analyze the relationship between univariate and bivariate EVT approaches applied to unsignalized intersections, taking into account the correlation of two datasets of extreme variables and giving more insights on the methodological procedure to follow in order to apply these models.

2.4.1 Surrogate measures of safety

The proximity to a collision can be measured by Post-Encroachment Time (PET) or Time-To-Collision (TTC) (Tarko, 2018a); many variations of these measures exist and several other indicators have been proposed in the past (Johnsson et al., 2018). The PET and TTC indicators used in this work are defined in the following Sub-Sections.

[2] The work presented in this Section is part of the following publication:

Gastaldi, M., Orsini, F., Gecchele, G., & Rossi, R. (2021). Safety analysis of unsignalized intersections: a bivariate extreme value approach. *Transportation letters*, *13*(3), 209-218. (Gastaldi et al., 2021)

2.4.1.1 Post-Encroachment Time

Post-Encroachment time (PET) was introduced by Allen et al. (1978) and it involves two elements: a conflict area, and the order in which two vehicles enter that area (Tarko, 2018a). It is conventionally defined as the time between the first road user leaves the conflict area and the second enters it (see Figure 12a). In other words it "indicates the extent to which they miss each other" (Johnsson et al., 2018).

This conventional definition works well when applied to 90° right-angle conflicts, less so when the angles are different. In the latter case entry and exit from the conflict area are no longer time instants, but time periods; in addition, the two vehicles may both be simultaneously inside the conflict area without being involved in a collision (Figure 12b). Laureshyn et al. (2010) proposed to define PET, more generally, as "the minimal delay of the first road user which, if applied, will result in a collision course and a collision".

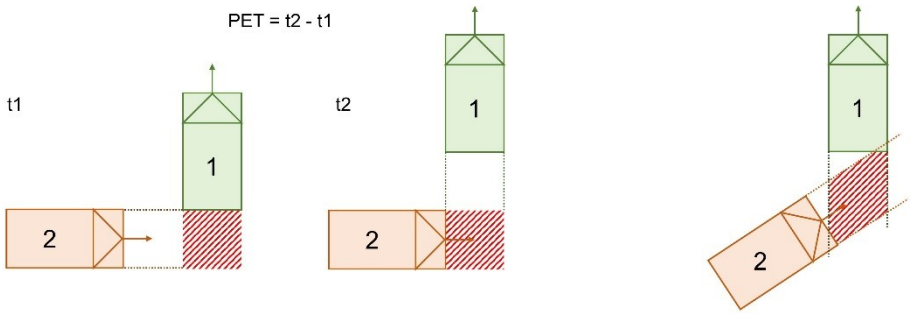

Figure 12. (a) Conventional definition of PET; (b) problem with conventional definition of PET: both road users may appear in conflict area without colliding (adapted from (Laureshyn et al., 2010)).

In the present work, PET was computed according to that definition. During the left-turning maneuver from minor road, the corner coordinates of both the left-turning vehicle (vehicle number 1 in Figure 13) and the conflicting vehicle in the major road (vehicle number 2 in Figure 13) were recorded at each time instant.

At time *t* the position of the left-turning vehicle was identified by the coordinates of its corners. For every successive time step, the position of the conflict vehicle on the major road (the coordinates of its corners) was recorded, and a check was made to verify whether the line segment representing its front bumper intersected any of the line segments representing the sides of the left-turning vehicle at time *t* (i.e. its "footprint"). Time *t'* was the first time step in which the conflict vehicle occupied the left-turning vehicle's "footprint" (see Figure 13). Therefore, PET at time *t* was:

$$PET(t) = t' - t \tag{18}$$

Time *t'* only exists if the left-turning vehicle is located, at time *t*, on the conflict vehicle's future path. If the two vehicles' trajectories intersect, there is at least one time step *t* in which PET exists.

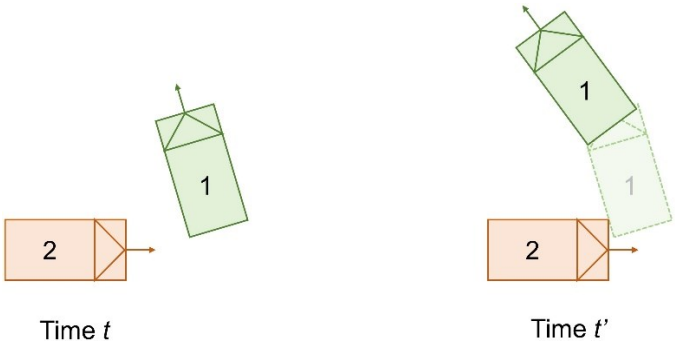

Figure 13. In this work *PET(t)* was defined as the difference between *t'* and *t*.

2.4.1.2 Time-To-Collision

Time-To-Collision is defined as "the time required for two vehicles to collide if they continue at their present speed and on the same path" (Hayward J.C., 1972). Traditionally, at a given time t, TTC is computed as the ratio between the space separating the two vehicles, and the speed difference between them. Speeds are recorded at time t and assumed to be constant. TTC is a continuous variable and is computed when the two vehicles are on a collision course. In practical applications, TTC is measured at given time steps.

Several modifications to the Hayward definition have been proposed, in order to increase TTC efficiency; a comprehensive review can be found in (Nadimi et al., 2020).

TTC is often used as a surrogate measure for rear-end collisions (Jonasson & Rootzén, 2014); in this case the trajectories of the vehicles overlap and the only condition to be satisfied, for the vehicles to be on a collision course, is the speed of the following vehicle being greater than that of the leading vehicle. In the case of a left-turn maneuver from minor road in an unsignalized intersection, the condition regarding speeds is not sufficient, and trajectories must also be taken into account. In addition, various types of collisions are possible: if the two vehicles are represented as rectangles, it is always the corner of one of the vehicles which hits the side of the other one, resulting in several possible configurations (Laureshyn et al., 2010) (Figure 14).

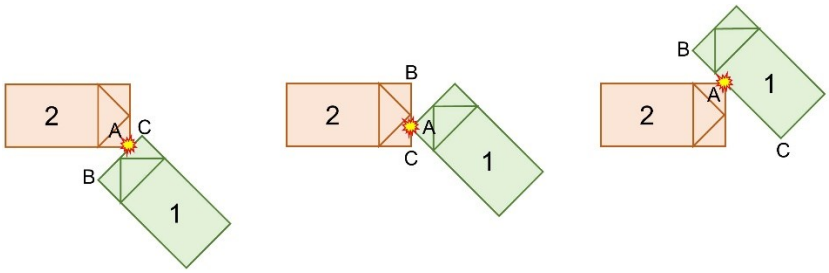

Figure 14. Some of the possible collision configurations between left-turning vehicle from minor road (1) and major road vehicle (2). Adapted from (Laureshyn et al., 2010).

In this study TTC was computed under a microscopic point-of-view, following the most general definition, first proposed by Van der Horst (1990), which says that the TTC is the time required for one point (the corner of one vehicle) to meet a line segment (the side of the other vehicle), provided that they maintain their present speed on the same trajectory. According to Nadimi et al. (2020) classification, this is "TTC for a moving line and an object".

Let (x_A, y_A) be the point coordinates at time t, and (vx_A, vy_A) its speed vector; let (x_B, y_B) and (x_C, y_C) be the coordinates at time t of the line segment ends, and (vx_{BC}, vy_{BC}) the line segment speed vector. Line segment BC lies on line l, with slope k. If speeds and the trajectories do not change, the time of collision between point A and line l containing line segment BC is:

$$TTC(t) = \frac{(y_A - y_B) - k(x_A - x_B)}{(vy_{BC} - vy_A) - k(vx_{BC} - vx_A)} \tag{19}$$

Point A and line l are on a collision course if TTC>0, although this does not necessarily mean that point A and line segment BC are. Another condition must be met: at t=TTC, collision point (x_{coll}, y_{coll}), which lies on line l must be inside line segment BC. More details on this method to compute TTC can be found in Laureshyn et al. (2010).

At each time step, the TTC of all the corner-side combinations is computed. The lowest of these measurements is used, as it identifies the side and corner which will come into contact first in a collision.

2.4.2 Driving simulator experiment

Data used for analysis were obtained with the same dynamic-base driving simulator used in the experiment presented in Section 2.3. The study involved twenty-three drivers (5 women, 18 men) whose age ranged from 20 to 35; this choice was supported, as in Section

2.3.1.2, by the fact that young drivers, particularly men, are at higher risk of accidents than older drivers (ETSC, 2016). All participants had regular driving license for at least two years, drove a minimum of 3000 km/year, and did not have any previous experience with driving simulators. Vision was normal or corrected-to-normal. One of the participants dropped out of the experiment due to simulator sickness and was replaced with another one.

For the experiment, a real-world three-leg intersection located in the urban road network of Dolo (province of Venice, Italy) was built in virtual reality with 3D rendering software (Figure 15b and Figure 15c). Participants were tested individually, after being properly instructed about the task. They were asked to drive as they would normally do, and they were given 10 minutes to become familiar with the simulator. During the experiment, each driver was asked to perform ten left-turn maneuvers from the minor road. Simulator variables (e.g. vehicles positions, speed) were recorded with a frequency of 50 Hz.

The major road traffic was derived from real-world observed sequences, in terms of vehicle type and time headways, using data collected with microwave radar sensors (see Table 4). Two shifted log-normal (SLN) distributions (one for each direction) were fitted on time headways recorded during afternoon peak hour; SLN distribution was first applied to fit headways by Tolle (1971), it has been applied in several works (see literature review in Saha et al., 2019) and it has been shown to provide results comparable to more sophisticated models (Ha et al., 2012). For each maneuver performed by participants, vehicles on the major road were randomly generated from these distributions, and they maintained a constant speed equal to the observed real-world average speed. To further improve the coherence with observed conditions, some of the vehicles in the major road randomly performed a right turn into the minor road.

Table 4. Major road traffic characteristics, for each direction.

	Direction: from Left	Direction: from Right
Cars	97%	96%
Heavy-duty trucks	3%	4%
Vehicles turning into the minor road	30%	0%
Speed [km/h]	64.5	64.5
SLN location parameter	0.49	0.35
SLN scale parameter	1.42	1.28
SLN shape parameter	0.80	1.10
Volumes [veh/h]	600	524

The experiment involved 230 maneuvers. Seven of them ended with a collision: five collisions involved vehicles in the major road coming from the left, two involved vehicles from the right. These observations were removed from the estimation datasets. The collision rate was relatively high compared to a real-world intersection; this was due to the behavior of the major road drivers, who ignored the presence of the tester's left-turning vehicle, and therefore did not slow down or perform any evasive maneuvers in order to avoid contact.

As in the proof-of-concept application presented in Section 2.3, setting this kind of behavior in the computer-controlled vehicles was intentional, in order to obtain a number of incidents higher than zero during the experiments, to be compared with the model estimation. As mentioned at the beginning of this Section, its aim is to illustrate and validate a statistical analysis to predict collisions at unsignalized intersections, rather than analyzing the safety of the specific intersection built for the driving simulator experiment. Therefore, the fact that drivers on the major road may not necessarily have a realistic behavior, does not affect the conclusions that can be drawn from the experiment.

Figure 15. (a) Satellite image of the intersection; (b) road-side picture of the intersection; (c) driving simulator view of the intersection

2.4.3 Univariate benchmarks

Two conflict points for left-turning vehicles at the unsignalized intersection analyzed in this study can be identified: one involving vehicles on the major road coming from the left-hand side (conflict point "L"), the other involving vehicles coming from the right-hand side ("R"), as can be observed in Figure 15a.

In the analyses performed in this Section, two different extreme variables were used to estimate EVT distribution: the PET and TTC (see Section 2.4.1).

To evaluate and compare the bivariate EVT methods, univariate EVT models were fitted for each conflict point (L and R) and each surrogate measure of safety (PET and TTC), using both BM and POT methods. In the remainder of this sub-paragraph we will refer to

PET data when describing the application of BM and POT, but the same process was applied also for TTC data.

The same procedure as in Section 2.3 was followed for the BM approach: (i) the block was defined as the single left-turn maneuver by each tester; (ii) the negated values of PET were considered; (iii) maxima were pre-processed to exclude observations that were not real extremes, by setting a limit value chosen with a sensitivity analysis; (iv) the remaining dataset was used to estimate GEV parameters reported in Table 5; (v) collision probability was defined as the cumulative GEV probability of negated-PET being higher than or equal to zero; (vi) the predicted number of collisions was computed by multiplying the collision probability and the number of blocks; (vii) confidence intervals were computed with the simulation-based inference method originally introduced by Songchitruksa & Tarko (2006).

For the POT approach, some clarification is needed on how the fitting dataset is composed. All the PET values collected during the left-turning maneuver of each tester were, in general, eligible to be part of the fitting dataset. However, one of the model requirements is that the dataset has to contain independent observations. Clearly, as PETs recorded throughout the same maneuver could not be considered as independent events, only one value of PET for each maneuver was included in the fitting dataset. The chosen value was the most extreme in the maneuver, i.e., the minimum PET.

Shifted-reciprocal value of PET (srPET) was considered, as proposed by L. Zheng et al. (2014b); therefore the extreme variable was: $srPET = 1/(PET + \delta)$. It has been observed in Section 2.3 that this transformation can produce better results for POT applications. The following procedure was then applied: (i) threshold and delta values were selected with the iterative procedure introduced in Section 2.3; (ii) exceedances were used to estimate GP parameters reported in Table 5; (iii) collision probability was the cumulative GP probability of srPET being higher than or equal to 1/delta; (iv) the predicted number of collisions was computed by multiplying the collision probability and the number of exceedances, as indicated in Coles (2001), Tarko (2018b); (v) confidence intervals were computed as in the BM approach. Predicted number of collisions and confidence intervals for univariate models are reported in Table 8 and Table 11.

Table 5. Parameter estimates and standard errors (in parenthesis) of all univariate EVT models. μ, σ and ξ are respectively location, scale and shape parameters; u is the threshold, δ is the delta value, N is the number of blocks (BM) or the number of exceedances (POT); n/a: not applicable.

Model	μ	σ	ξ	u	δ	N
BM (PET,L)	-1.5662 (0.0835)	0.7114 (0.0654)	-0.3343 (0.1043)	n/a	n/a	100
BM (PET, R)	-2.2450 (0.0789)	0.6478 (0.0622)	-0.1322 (0.1257)	n/a	n/a	107
BM (TTC,L)	-0.5276 (0.0252)	0.1817 (0.0211)	0.2069 (0.1602)	n/a	n/a	93
BM (TTC,R)	-1.2589 (0.0377)	0.2841 (0.0300)	-0.1964 (0.1388)	n/a	n/a	90
POT (PET,L)	n/a	0.6007 (0.2302)	0.4928 (0.1251)	0.64	0.12	56
POT (PET,R)	n/a	0.4832 (0.2996)	0.5179 (0.1735)	0.84	0.19	31
POT (TTC,L)	n/a	1.4088 (0.3746)	0.7238 (0.2570)	1.50	0.08	69
POT (TTC,R)	n/a	0.2817 (0.0796)	0.4388 (0.2235)	0.94	0.15	37

2.4.4 Bivariate approach #1: two conflict points

In the first proposed bivariate approach two conflict points were considered (L and R); the extreme variable observed and used to estimate the models was PET for both conflict points (see Figure 16). A preliminary analysis showed no correlation between the two PET datasets (Pearson-correlation value is -0.008, p-value is 0.96, with the null hypothesis being correlation equal to zero), suggesting that univariate and bivariate approaches should produce similar results. The aim of the analysis of this paragraph is to validate the methodologies used in the bivariate approach, by comparing its results with those obtained with the univariate models.

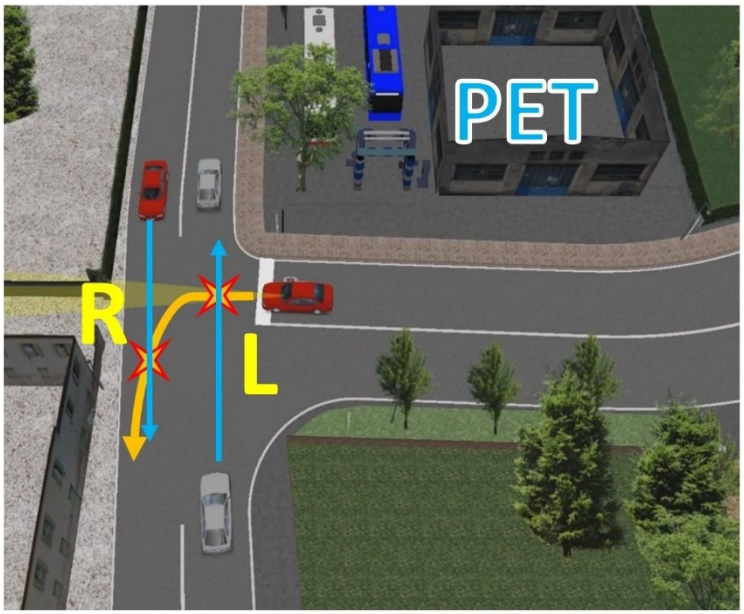

Figure 16. Bivariate approach #1 considers two conflict points (L and R) and one surrogate measure of safety (PET).

2.4.4.1 CM method

In the CM approach the block was defined as the single left-turn maneuver performed by each tester, as in the univariate BM case. Each CM vector was composed by the two maximum negated-PETs registered during the left-turning maneuver, considering respectively conflicts with vehicles coming from the left (L) and from the right (R). As mentioned in Section 2.2.2.1, this vector is not necessarily an observed set of variables, because the two values of negated-PET are, in general, observed at different time instants within the block. Only negated-PETs above the limit values defined in the univariate case were considered: as a result, some CM vectors contained only one value, because the other was below the limit value and therefore not a real extreme. This is different from the choice made by Cavadas et al. (2020), who considered only blocks in which both components were above the limit value; however we believe that this would cause a loss in sample information, which may compromise estimation results, especially if correlation between extreme variables is low.

During the model fitting stage, the distribution family for *G* must be chosen: the model was fitted several times, testing different families. As in L. Zheng, Ismail, et al. (2018), the chosen family should be the one with the lowest Akaike Information Criterion (AIC), which is an indicator of relative quality for statistical models, dealing with the trade-off between model goodness-of-fit and simplicity. The lowest AIC value was obtained for the logistic family (8) (see Table 6). Model fitting then produced estimates and standard errors for all GEV parameters in Table 7 and a *r* logistic parameter of 0.987, which indicates independence (Coles, 2001).

Collision probability is the probability of negated-PET to be higher than or equal to zero in either one of the two conflict points. In mathematical terms:

$$Pr(-PET_L > 0 \vee -PET_R > 0) = \\ Pr(-PET_L > 0) + Pr(-PET_R > 0) - Pr(-PET_L > 0 \wedge -PET_R > 0) \qquad (20)$$

Then the predicted number of collisions is computed by multiplying these probabilities by the respective number of blocks (e.g., the probability of $-PET_L>0$ is multiplied by the number of blocks in which the negated-PET of vehicles on the major road coming from the left is above the limit value). Confidence intervals were then defined adapting to the bivariate analysis the same procedure used in the univariate case (see Table 8).

2.4.4.2 EOT method

The fitting datasets used in the univariate POT approach were used also for the bivariate EOT analysis. The delta values defined in the univariate case were used to compute srPETs. However, as pointed out in Section 2.2.2, equation (11) is valid only if an optimal pair of thresholds are chosen, and these thresholds are not necessarily the same as in the univariate case. Beirlant et al. (2004) proposed to use the spectral measure plot to determine the thresholds, and this approach, in combination with the traditional stability plot, was used in this work, as it was in previous bivariate EVT applications in road safety (Chen Wang et al., 2019; L. Zheng, Ismail, et al., 2018). According to Beirlant's theory, at suitable thresholds, the empirical spectral value should be close to the theoretical spectral measure, which, for the specific bivariate EVT case, is 2. More details are available in Beirlant et al., (2004). In this case the *k*-value was 87, therefore, threshold values equal to the 87th-highest value in each srPET dataset were chosen. Stability plots then confirmed that these two thresholds could be considered acceptable.

As in the CM approach, the model was then fitted several times, evaluating different families for *G*. The logistic family was again the best in terms of AIC value (see Table 6). Parameter estimates are listed in Table 7, whereas the *r* logistic parameter was 0.999, confirming that the two variables were independent. The collision probability was then defined as:

$$Pr\left(srPET_L > \frac{1}{\delta_{PET,L}} \vee srPET_R > \frac{1}{\delta_{PET,R}}\right) = Pr\left(srPET_L > \frac{1}{\delta_{PET,L}}\right) + Pr\left(srPET_R > \frac{1}{\delta_{PET,R}}\right) - Pr\left(srPET_L > \frac{1}{\delta_{PET,L}} \wedge srPET_R > \frac{1}{\delta_{PET,R} l}\right) \quad (21)$$

and the predicted number of collisions by multiplying probabilities and number of exceedances over the thresholds (see Table 8).

2.4.4.3 Discussion

As can be observed in Table 8, the results, in terms of predicted number of collisions, are similar between univariate and bivariate approaches. As mentioned before, this was expected, because the two sets of PETs were not correlated: in such situation it is

reasonable to assume that the collision probability is independent for the two conflict points, so the probabilities can be calculated separately with univariate models and then simply added up. BM/CM and POT/EOT results are comparable, and they both tend to underestimate the number of accidents. However, the BM/CM approach is always able to include the observed number of collisions within the confidence interval.

Table 6. Bivariate approach #1. AICs of estimated models with different parametric families G, for both CM and EOT methods. Lowest AIC for each method is in italics.

Family *G*	CM method	EOT method
Logistic	*457.3936*	*503.3757*
Asymmetric logistic	460.0174	n/a
Bilogistic	464.2841	n/a
Negative logistic	457.4163	503.4068
Asymmetric negative logistic	n/a	n/a
Negative bilogistic	457.7321	505.4087
Husler-Reiss	457.4164	504.7249

Table 7. Bivariate approach #1. Parameter estimates and standard errors for both CM and EOT bivariate methods.

	$\mu_{PET,L}$	$\sigma_{PET,L}$	$\xi_{PET,L}$	$u_{PET,L}$	$\mu_{PET,R}$	$\sigma_{PET,R}$	$\xi_{PET,R}$	$u_{PET,R}$
CM	-1.5667 (0.0836)	0.7113 (0.0655)	-0.3341 (0.1042)	n/a	-2.2453 (0.0789)	0.6475 (0.0621)	-0.1311 (0.1259)	n/a
EOT	n/a	0.4220 (0.0825)	0.5331 (0.1701)	0.43	n/a	0.3419 (0.0733)	0.5494 (0.2037)	0.40

Table 8. Bivariate approach #1. Predicted number of collisions. 95% confidence intervals in parenthesis.

	Univariate			**Bivariate**
	PET, L	**PET,R**	**PET, L + PET, R**	**Combined**
BM/CM	1.94 [0.00;5.15]	1.05 [0.00;4.15]	2.99 [0.34;7.39]	2.87 [0.34;7.09]
POT/EOT	1.24 [0.02;3.65]	1.12 [0.00;2.99]	2.36 [0.47;5.37]	2.57 [0.65;5.55]
Observed	5	2	7	7

2.4.5 Bivariate approach #2: two surrogate measures of safety

In the second approach the bivariate procedure was applied in a situation in which the two sets of extreme variables were correlated. In this case, conflict point L is analyzed considering two surrogate measures of safety (PET and TTC) (see Figure 17). All models were estimated following the same procedures described in the previous Sub-Section. Even though the two measures have different definitions (e.g., TTC makes assumptions on vehicle speeds, whereas PET does not), the two sets of observed measures are correlated (Pearson-correlation value is 0.262, p-value 0.02). An even slightly higher correlation was observed considering conflict point R and the same two surrogate measures of safety (Pearson-correlation value 0.368, p-value <0.01); however, the full analysis of that conflict point is not reported in this Section for the sake of brevity.

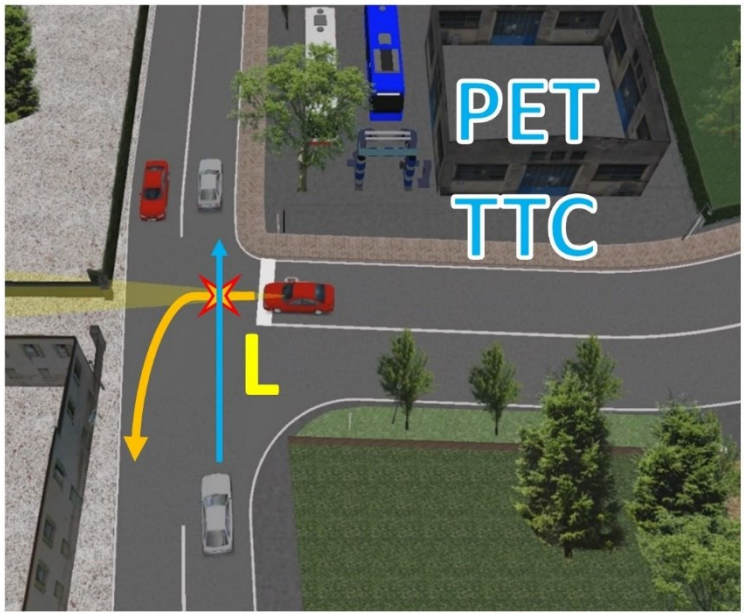

Figure 17. Bivariate approach #2 considers one conflict point (L) and two surrogate measures of safety (PET and TTC).

2.4.5.1 CM method

In the CM approach, negated-PET and negated-TTC block maxima above univariate limit values were excluded from the fitting dataset. In this case the family with the lowest AIC was the negative bi-logistic (see Table 9), whose formulation is:

$$G(x,y) = exp\{-x - y + xq^{1+\alpha} + y(1-q)^{1+\beta}\} \quad (22)$$

where $\alpha > 0, \beta > 0$, and $q = q(x,y,\alpha,\beta)$ is the root of the following equation:

$$(1+\alpha)xq^{\alpha} - (1+\beta)y(1-q)^{\beta} = 0 \tag{23}$$

EVT parameter estimates are listed in Table 10; as for the negative bilogistic parameters *a* and β, their estimates were respectively 19.66 and 0.100, indicating an intermediate situation between independence and complete dependence (Coles & Tawn, 1994). Collision probability was defined as:

$$\begin{gathered} Pr(-PET_L > 0 \vee -TTC_L > 0) = \\ Pr(-PET_L > 0) + Pr(-TTC_L > 0) - Pr(-PET_L > 0 \wedge -TTC_L > 0) \end{gathered} \tag{24}$$

producing the predicted number of collisions and confidence intervals shown in Table 11.

2.4.5.2 EOT method

In the EOT approach delta values from univariate models were used to compute srPET and srTTC. Spectral measure plot returned a *k*-value of 55, and threshold values were consequently defined. Logistic family was chosen (see Table 9) and model fitting produced the parameter estimates in Table 10 and a *r* parameter of 0.652, which indicates some dependence between the extreme variables (Coles, 2001). Collision probability was:

$$\begin{gathered} Pr\left(srPET_L > \frac{1}{\delta_{PET,L}} \vee srTTC_L > \frac{1}{\delta_{TTC,L}}\right) = \\ Pr\left(srPET_L > \frac{1}{\delta_{PET,L}}\right) + Pr\left(srTTC_L > \frac{1}{\delta_{TTC,L}}\right) - \\ Pr\left(srPET_L > \frac{1}{\delta_{PET,L}} \wedge srTTC_L > \frac{1}{\delta_{TTC,L}}\right) \end{gathered} \tag{25}$$

and predicted number of collisions and confidence intervals are shown in Table 11.

2.4.5.3 Discussion

Table 11 allows to compare univariate and bivariate results. Although the PET and TTC datasets are correlated, they produce significantly different collision predictions in the univariate approach: PET-based models tend to underestimate collisions, whereas TTC-based models tend to overestimate them. On the other hand, the combination of PET and

TTC inputs in the bivariate approach returns much better results, especially in the case of EOT approach, although both CM and EOT confidence intervals include the observed value of collisions. In this case, the bivariate approach is able to acquire and combine information from both surrogate measures of safety, providing results that are more reliable than in the univariate approach.

Table 9. Bivariate approach #2. AICs of estimated models with different parametric families G, for both CM and EOT methods. Lowest AIC for each method is in italics.

Family *G*	CM method	EOT method
Logistic	216.8516	*679.3367*
Asymmetric logistic	218.0411	n/a
Bilogistic	212.5689	n/a
Negative logistic	216.5191	680.7934
Asymmetric negative logistic	217.4818	n/a
Negative bilogistic	*212.5506*	682.7934
Husler-Reiss	216.3539	688.2287

Table 10. Bivariate approach #2. Parameter estimates and standard errors for both CM and EOT bivariate methods

	$\mu_{PET,L}$	$\sigma_{PET,L}$	$\xi_{PET,L}$	$u_{PET,L}$	$\mu_{TTC,L}$	$\sigma_{TTC,L}$	$\xi_{TTC,L}$	$u_{TTC,L}$
CM	-1.5267 (0.0805)	0.7456 (0.0682)	-0.4306 (0.0767)	n/a	-0.5396 (0.0235)	0.1761 (0.0203)	0.2234 (0.1407)	n/a
EOT	n/a	0.6902 (0.2001)	0.7411 (0.2952)	0.64	n/a	0.5543 (0.1856)	1.7287 (0.3787)	1.63

Table 11. Bivariate approach #2. Predicted number of collisions. 95% confidence intervals in parenthesis.

	Univariate		**Bivariate**
	PET,L	**TTC,L**	**Combined**
BM/CM	1.94 [0.00;5.15]	9.58 [4.09;14.76]	8.79 [4.20;13.52]
POT/EOT	1.45 [0.02;3.65]	8.14 [2.99-12.46]	5.25 [0.75;11.32]
Observed	5	5	5

2.4.6 Summary and discussion

In this Section a bivariate extreme value theory application to road safety analysis of a left-turn maneuver at a three-leg unsignalized intersection has been presented. The bivariate analysis has been applied under two alternative perspectives: two conflict points and the same surrogate measure of safety (PET); only one conflict point and two different surrogate measures of safety (PET and TTC). In the former case, the two sets of extreme variables were independent, and therefore bivariate results were similar to the univariate EVT benchmarks. In the latter case, a significant improvement in model accuracy was obtained: the univariate EVT models produced discordant results, with the PET-based models underestimating the number of collisions and the TTC-based models overestimating it. In such situation, it was difficult to claim which of the two surrogate measures produced the best results; on the other hand, the bivariate approach allowed to gain information from both measures and returned a combined predicted number of collisions closer to the observed number of collisions.

There are some limitations to this work that will be addressed in future research. First of all, in order to generalize results, the application of the same analysis to different sites and maneuvers (e.g., left-turn from major road) and a replication of the study extending size and composition of the drivers' sample would be of great value. More bivariate approaches, such as the copula (Cavadas et al., 2020; S. Das & Maurya, 2020), will be tested. The effect of covariates, such as traffic flow or waiting time at the intersection, will also be investigated in the future, as well as a multivariate model to include more than two surrogate measures of safety.

Even though this line of research is recent and relatively unexplored, these results are already interesting under a practical point-of-view. Indeed, it is not always clear which surrogate measure performs better in a given situation, and specifically to analyze certain maneuvers at unsignalized intersections (Shekhar Babu & Vedagiri, 2018). With the proposed approach, multiple measures can be collected and combined, in order to obtain more robust collision predictions.

2.5 Case study: large-scale highway application[3]

The ultimate objective of this part of the dissertation is to apply the EVT models to a real-world scenario.

Both BM and POT approaches are applied in a large-scale motorway scenario, in order to evaluate the risk of rear-end collisions. Several cross-sections are evaluated using models calibrated on data collected in each cross-section for a year. Model performance is evaluated by comparing predicted and observed rear-end collisions.

It is worth noting that, due to the characteristics of the type of conflict analyzed and the data available, only a univariate EVT approach was viable.

This Section contributes significantly to the existing road safety EVT literature in several ways:

- For the first time EVT is applied to a large-scale (in terms of observation period) real-world application.
- It proposes the use of radar technology to collect data, which is a less expensive (than video recording) and convenient way to collect data for long period of times.
- It investigates for the first time in EVT literature motorway read-end collisions, which are a serious safety concern.
- It introduces in road safety EVT applications an automatic method to select the threshold of the POT approach, which can be very practical in real-world large-scale safety analysis.

Finally, this approach may allow in future works to deal with other still unexplored and open issues in EVT, such as transferability and seasonality, as well as the impact of observation period on the quality of results.

[3] The work presented in this Section is part of the following publication:

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2020). Large-scale road safety evaluation using extreme value theory. *IET Intelligent Transport Systems*, *14*(9), 1004-1012. (Orsini et al., 2020a)

2.5.1 Data collection

2.5.1.1 Data for calibration: TTC values

Traffic data were collected for one year (2013) in 19 cross-sections located in two motorways in North-Eastern Italy, using micro-wave Doppler radars. These were all the available sections which were recording traffic data on the two motorways continuously in 2013.

In the rest of the Section, these cross-sections are identified by a code ("Section ID"): IDs 1 to 14 refer to cross-sections located on a three-lane motorway, IDs 15 to 19 on a two-lane motorway. In each cross-section a radar is permanently installed over each lane of the carriageway and collects the following vehicle-by-vehicle information: speed, time gap, time stamp and vehicle class.

With these data it is possible to calculate the TTC between each couple of vehicles consecutively detected on the same lane, with the classic definition:

$$TTC = \frac{R}{RR} = \frac{v_L * t_{GAP}}{v_F - v_L} \tag{26}$$

where v_L and v_F are the speeds of the leading and the following vehicles, respectively; t_{GAP} is the time gap between them, recorded at the cross-section.

The reader is reminded that two vehicles are on a collision course if the speed of the following vehicle is higher than the leading vehicle (RR must be higher than zero); TTC equal to zero represents a collision. For more details on how to compute it and on alternative definitions, the reader is referred to Nadimi et al. (2020).

In this work we considered the following normalized definition of TTC:

$$nTTC_j = TTC_j - min\{TTC_j\}_{j=1}^{n} \tag{27}$$

where j is the generic vehicle which crosses the section during the observation period, n the total number of vehicles recorded during the observation period, TTC_j the instantaneous TTC computed for the j-vehicle, and $nTTC_j$ the normalized value of TTC for the j-vehicle.

This choice was suggested by a previous work, which showed that better model performance can be obtained with such normalization when there is high variability in the observed surrogate measure of safety (Cavadas et al., 2020). A preliminary analysis confirmed that this was the case also in the present study; for the sake of brevity, and since

demonstrating the effectiveness of TTC normalization was not the aim of this Section (that was already done in Cavadas et al., 2020), this analysis is not reported in the present work.

In the remainder of the Section, we will use the term "TTC" instead of "nTTC" for notation lightness, even though all analyses were performed with the normalized values of TTC.

2.5.1.2 Data for validation: historical crash counts

The models calibrated with TTC values are used to estimate annual number of rear-end crashes (see Section 0); in order to validate the results, these estimates are compared to the observed number of rear-end crashes.

A 6-year crash database was available, covering both motorways. Data were collected by the administrative department of the motorways from 1st January 2011 to 31st December 2016. During this period, 5632 crashes occurred.

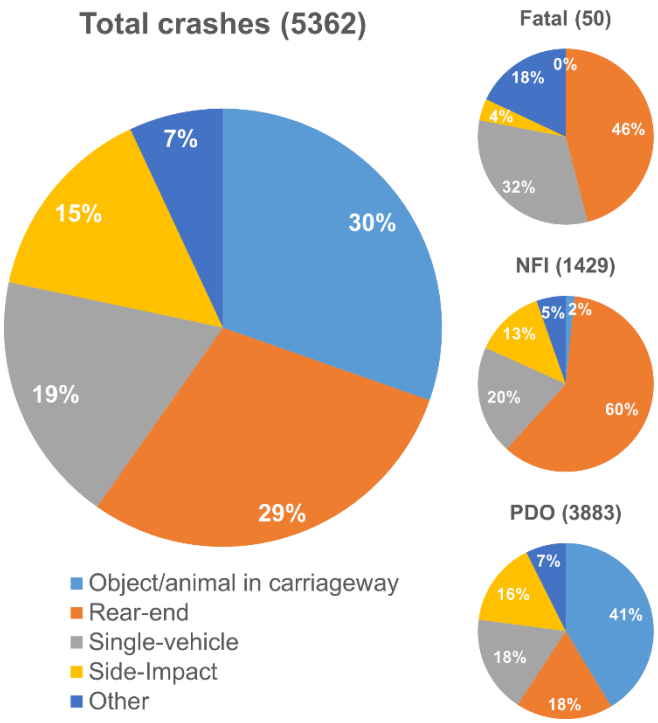

Figure 18. On the left: total crashes per type, observed in the motorways of this work between 2011 and 2016. On the right: fatal, non-fatal-injury (NFI) and property-damage-only (PDO) crashes per type, observed in the motorways of this work between 2011 and 2016.

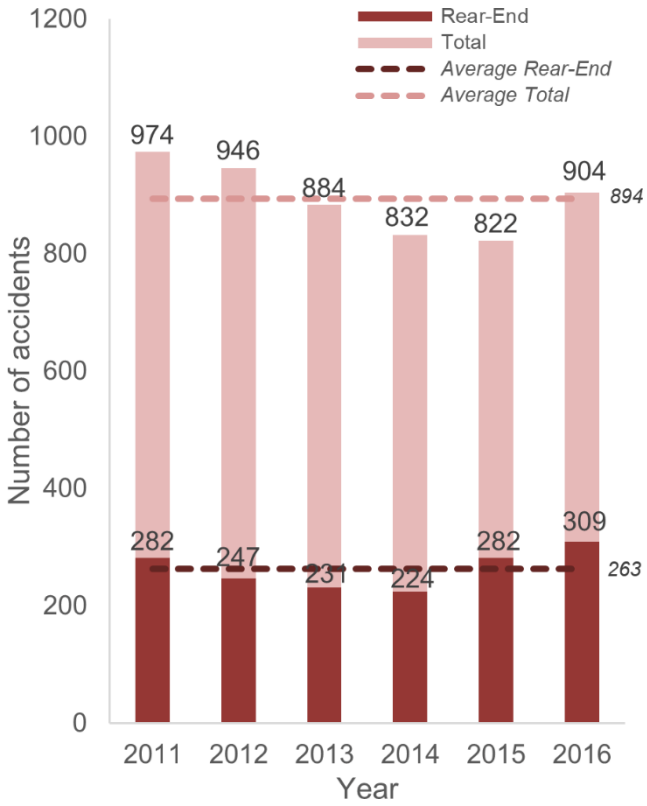

Figure 19. Total crashes and rear-end crashes per year, observed in the motorways of this work between 2011 and 2016.

As can be observed in Figure 19, rear-end collisions represent an important slice of the total crashes (1575 accidents, 29% of the total), second only to crashes caused by object or animals on the carriageway. However, most of the latter crashes do not have serious consequences, whereas the rear-end collisions account for the 46% of fatal crashes (23 out of 50) and the 60% of crashes with injured people (859 out of 1429).

These statistics underline the need to develop tools to model motorway rear-end crashes, as they are by far the most frequent among the crashes with serious consequences. It is also worth noting that also the single-vehicle crashes are particularly relevant, but as conflict-based approach is generally not able to model this kind of collisions, other approaches should be pursued (Astarita & Giofré, 2019).

As mentioned in Section 2.5.1.1, parameters estimation was carried out with data collected during 2013. During validation (Section 2.5.2.3), the predicted number of collisions was compared with the average annual number of observed collisions, assuming that crashes are constant in the 2011-2016. This is indeed confirmed by analyzing crash data in period 2011-2016: the total annual number of crashes does not show a clear trend (see Figure 19), ranging between 822 in 2015 and 974 in 2011, with an average value of 894; as for the rear-end crashes, they range between 224 in 2014 and 309 in 2016, with an average value of 262.

In 2013 both the total number of crashes (884) and the number of rear-end crashes (231) are close to the average value of the period 2011-2016.

2.5.2 Results: annual collision predictions

2.5.2.1 Block Maxima approach

In the Block Maxima approach the following procedure was followed in order to obtain the annual predicted number of rear-end collisions. Each motorway section was analyzed independently from the others.

Block definition: The observation period has to be divided into several blocks, homogeneous between each other, whose size is chosen as a trade-off between bias and variance: blocks that are too small can produce bias in estimation and extrapolation of GEV parameters; large blocks generate few block maxima, leading to large estimation variance (Coles, 2001). In this analysis a block length of one day was defined, according to the findings of a previous analysis in the same motorway, which showed that a one-day block size produced better performance than one-hour block size (Gecchele et al., 2019). It is important to stress that the block size choice strongly depends on the context, in particular on the type of collision analyzed and the infrastructure. In this case, the observed extreme events (i.e. near-crashes) are relatively rare, therefore a rather big block size is chosen; in other contexts, for example intersections, conflicts are more frequent and therefore smaller block size can be chosen: L. Zheng & Sayed 2019a) used a block size of 20 minutes (selected after an appropriate sensitivity analysis) to estimate crashes at signalized intersections.

GEV estimation dataset: EVT is designed to estimate the probability of extremely high values of the variable of interest. In this analysis extreme events are represented by extremely low values of TTC, therefore a transformed variable should be used. The common practice in road safety is to consider negated values of TTC (Farah & Azevedo, 2017; Songchitruksa & Tarko, 2006; L. Zheng et al., 2014a).

In the BM approach the estimation dataset is composed by the block maxima. If the block size is relatively big compared to the observation period, the dataset may be too small to obtain a reliable estimation; in that case r-largest values are selected for each block (Gecchele et al., 2019; Songchitruksa & Tarko, 2006).

In this work, the estimation dataset is composed by the maximum value of –TTC in each day-block. Days in which only partial data were recorded, due to instrument maintenance or failure, were excluded from analysis.

GEV parameters estimation: The estimation dataset is then used to estimate the parameters of a GEV distribution. In this work estimation is performed with maximum likelihood (ML) method, which returns parameter estimates, confidence intervals of parameters estimate and variance-covariance matrix. In order to have reliable ML estimates, shape parameter must be higher than -0.5 (Smith, 1985). Table 12 presents GEV parameters for each motorway section.

Table 12. Estimated GEV parameters and respective standard errors and 95% confidence intervals, for each motorway section

ID	Blocks	Shape	Scale	Location
1	307	-0.383 (0.034) [-0.449;-0.317]	0.169 (0.007) [0.155;0.184]	-0.392 (0.010) [-0.412;-0.371]
2	302	-0.392 (0.031) [-0.450;-0.329]	0.149 (0.007) [0.137;0.163]	-0.348 (0.009) [-0.369;-0.333]
3	306	-0.276 (0.020) [-0.314;-0.237]	0.183 (0.007) [0.169;0.198]	-0.612 (0.011) [-0.634;-0.590]
4	309	-0.396 (0.038) [-0.469;-0.322]	0.216 (0.010) [0.198;0.236]	-0.478 (0.013) [-0.504;-0.452]
5	309	-0.336 (0.019) [-0.373;-0.299]	0.169 (0.007) [0.156;0.183]	-0.483 (0.010) [-0.503;-0.462]
6	308	-0.487 (0.028) [-0.541;-0.430]	0.259 (0.012) [0.237;0.283]	-0.518 (0.016) [-0.550;-0.488]
7	309	-0.427 (0.031) [-0.489;-0.365]	0.214 (0.010) [0.196;0.234]	-0.476 (0.013) [-0.502;-0.450]
8	309	-0.279 (0.019) [-0.317;-0.241]	0.172 (0.007) [0.159;0.186]	-0.564 (0.010) [-0.585;-0.544]
9	308	-0.331 (0.032) [-0.395;-0.269]	0.198 (0.009) [0.182;0.215]	-0.517 (0.012) [-0.541;-0.493]
10	307	-0.321 (0.016) [-0.351;-0.289]	0.171 (0.007) [0.158;0.184]	-0.514 (0.010) [-0.534;-0.494]
11	310	-0.499 (0.030) [-0.559;-0.440]	0.262 (0.012) [0.240;0.287]	-0.505 (0.016) [-0.537;-0.474]
12	309	-0.408 (0.026) [-0.458;-0.355]	0.204 (0.009) [0.188;0.222]	-0.476 (0.012) [-0.501;-0.452]
13	307	-0.299 (0.031) [-0.357;-0.237]	0.175 (0.008) [0.160;0.189]	-0.504 (0.011) [-0.526;-0.484]
14	305	-0.346 (0.025) [-0.392;-0.296]	0.183 (0.008) [0.169;0.199]	-0.486 (0.011) [-0.510;-0.467]
15	249	-0.460 (0.026) [-0.497;-0.408]	0.281 (0.014) [0.249;0.295]	-0.597 (0.019) [-0.620;-0.555]
16	286	-0.372 (0.024) [-0.427;-0.335]	0.325 (0.014) [0.299;0.354]	-0.832 (0.021) [-0.856;-0.778]
17	306	-0.447 (0.021) [-0.486;-0.403]	0.270 (0.011) [0.247;0.292]	-0.593 (0.016) [-0.625;-0.561]
18	280	-0.442 (0.033) [-0.498;-0.380]	0.286 (0.014) [0.260;0.311]	-0.615 (0.018) [-0.647;-0.579]
19	285	-0.325 (0.034) [-0.392;-0.265]	0.337 (0.015) [0.308;0.366]	-0.880 (0.022) [-0.917;-0.833]

GEV collision probability and predicted collisions: Once the parameters have been estimated, it is possible to use the cumulative GEV distribution (1) to obtain collision predictions. According to the definition of TTC and the assumptions made in Section 2.5.1.1, a collision occurs when TTC=0. Therefore, the collision probability can be defined, for each block as:

$$P_{inc,block} = Pr\{-TTC \geq 0\} = 1 - GEV(0) \quad (28)$$

Considering observation time *t*, the predicted number of collisions in *t* is:

$$Inc_t = P_{inc,block} * N_{block} \quad (29)$$

Where N_{block} is the number of blocks in which the observation period has been divided. To obtain the predicted number of collisions for a period of time *T* (with *T*>*t*, for example 1 year), Inc_t must be multiplied by the ratio T/t, under the assumption that the collision probability is constant throughout *T*. Therefore, the predicted number of collisions in *T* is:

$$Inc_T = Inc_t * \frac{T}{t} = P_{inc,block} * N_{block} * \frac{T}{t} \quad (30)$$

Predicted number of collisions for each section and *T*=1 year are reported in Table 13.

Confidence interval of collision predictions: As mentioned before, GEV parameter estimates are returned with a certain degree of uncertainty. To properly evaluate the performance of the models it is therefore appropriate to give some indication also on the uncertainty of the resulting collision predictions. This is usually done in road safety applications of EVT by computing the confidence interval of collisions number.

Under regularity conditions, it is possible to assume that GEV parameters follow the multivariate normal distribution (Coles, 2001, page 56). The values of parameters are generated from this distribution several times, then the collision probability and predicted collisions are computed each time, producing empirical distributions and standard errors; the confidence intervals are then calculated from the quantiles of the empirical distributions. These assumptions were further tested by using a non-parametric method, bootstrapping with 1000 draws (see Gilli & Këllezi, 2006; Tarko, 2012 for further details), which produced similar confidence intervals.

Table 13 contains 95% confidence intervals for annual predicted collisions.

2.5.2.2 Peak-Over-Threshold approach

In the Peak-Over-Threshold approach the following procedure was followed in order to obtain the annual predicted number of rear-end collisions. Each motorway section was analyzed independently from the others.

***GP estimation dataset**:* As in the BM approach (see Section 2.5.2.1), the variable of interest, TTC, must be transformed. Typically, the negated transformation is applied also in the POT approach (Farah & Azevedo, 2017; L. Zheng et al., 2014a; L. Zheng & Sayed, 2019a), but in some other cases shifted-reciprocal transformation is used (L. Zheng et al., 2014b). In the present case negated transformation is used, in order to reduce the number of parameters to estimate (in the shifted-reciprocal transformation also a delta-parameter has to be defined, see Section 2.3.2.3).

TTC values lower than zero are excluded from analysis because, according to the definition of TTC (see Section 2.5.1.1) there is no risk of rear-end collision, as the following vehicle is slower than the leading vehicle.

In principle, every other TTC value recorded during the observation period is eligible to be included in the estimation dataset. However, in order to have a smaller and more manageable dataset, also TTC higher than 5 seconds were excluded from analysis. These TTC values represented situations of very low collision risk and, as the reader can see in Table 14, all thresholds used in the models were significantly lower than 5 seconds, in absolute terms, therefore excluding these data from analysis had no impact on the results.

***Threshold selection**:* Before estimating the GP parameters with ML method, a threshold has to be selected. As mentioned in Section 2.2.1.2, traditionally the threshold is chosen via graphical diagnostics, which is however practically unfeasible when the number of models to estimate is high. For this reason the ATSM developed by Thompson et al. (2009) was adopted (see Section 2.2.1.2 for further details).

In the ATSM, the analyst must choose a set of candidate thresholds. In this case the set was the largest possible, ranging from the lowest value of negated-TTC (i.e., all datasets values are extremes) and the 30th highest value of negated-TTC. The latter condition was adopted because both literature (Farah & Azevedo, 2017) and analysts' experience suggest that model estimations with less than 30 values are usually unreliable.

The ATSM returns a range of suitable threshold. The largest value of this range was chosen, in analogy to the graphical diagnostics method, where the highest possible threshold is usually chosen.

All threshold values were considered acceptable also according to graphical diagnostics method. To exemplify this procedure, Figure 20 shows the result of this graphical test for

two of the motorways Sections (IDs 1 and 2): in both cases the ATSM-proposed threshold is within the stable region of the parameters, and the linear region of the mean excess plot. Threshold values and number of exceedances are presented in Table 14.

GP parameters estimation: Once a threshold has been selected for each motorway section, GP parameters can be estimated with ML method. As for the GEV distribution, shape parameter should be higher than -0.5 (Smith, 1985). Parameter values are reported in Table 14.

GP collision probability and predicted collisions: Once the thresholds have been selected and the parameters estimated, it is possible to use the cumulative GP distribution (2) to obtain collision predictions. Considering observation time *t*, the collision probability in *t*, conditional to the two vehicles being in a conflict situation (negated-TTC higher than the threshold *u*), can be defined as follows:

$$P_{inc,t} = Pr\{-TTC \geq 0 | -TTC > u\} = 1 - GP(0) \quad (31)$$

the number of predicted collisions in that time period is obtained by multiplying the collision probability $P_{inc,t}$ (produced estimating the GP model using only TTC values collected during *t*), by the number of exceedances $N_{exc,t}$, i.e. the number of observed situations in which the negated-TTC is higher than the threshold, during *t*. In other words, the predicted number of incidents in the observation period is:

$$Inc_t = P_{inc,t} * N_{exc,t} \quad (32)$$

As in Section 2.5.2.1, to obtain the predicted number of collisions in *T*, Inc_t must be multiplied by the ratio T/t, which is valid only under the assumption that the collision probability is constant throughout *T*:

$$Inc_T = Inc_t * \frac{T}{t} = P_{inc,t} * N_{exc,t} * \frac{T}{t} \quad (33)$$

Predicted number of collisions for each section and T=1year are reported in Table 13.

Confidence interval of collision predictions: The same procedure illustrated in Section 2.5.2.1 was followed. 95% confidence intervals for POT approach are in Table 13.

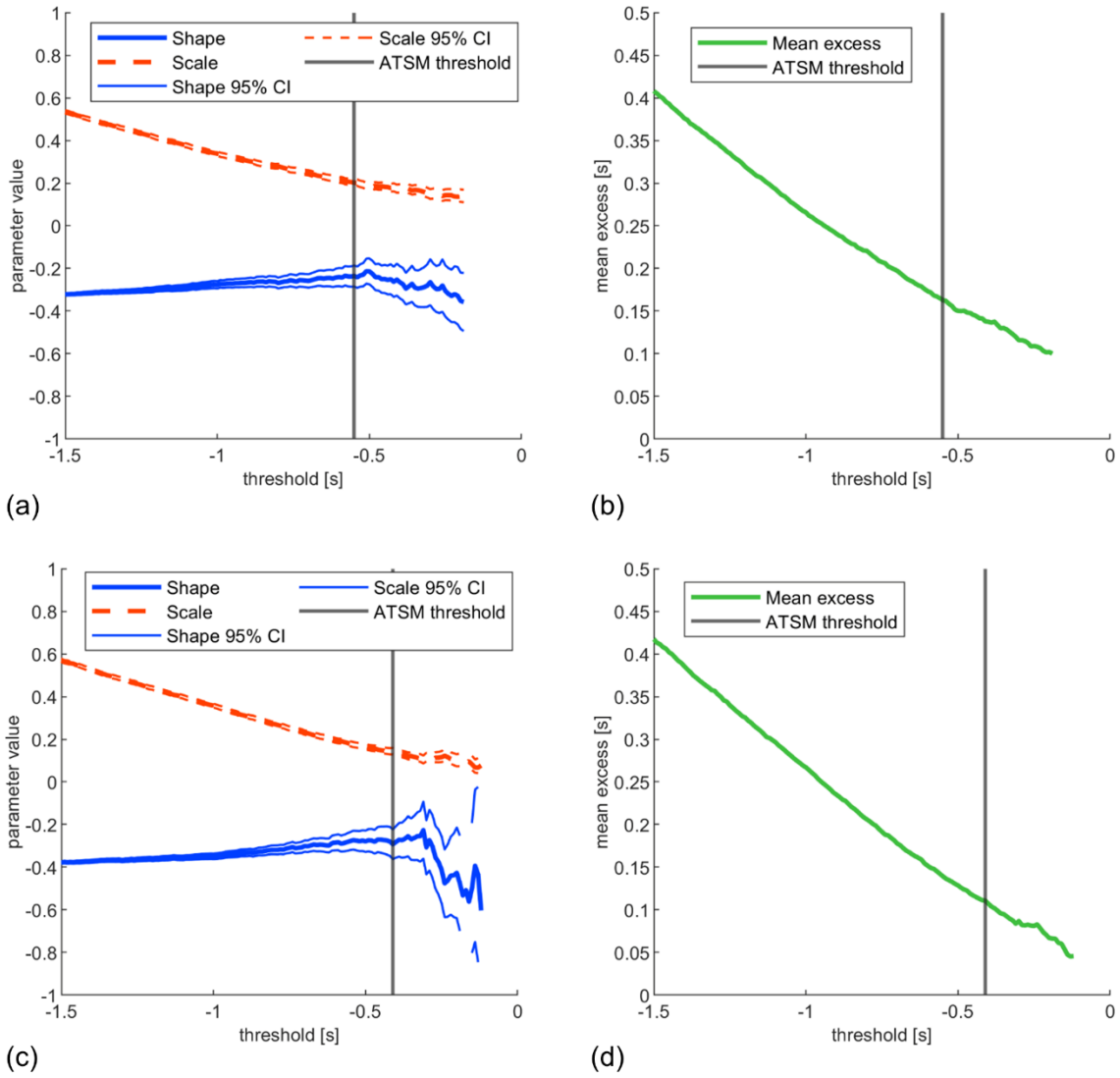

Figure 20. Graphical diagnostics test for Sections 1 and 2:(a) stability plot of Section 1; (b) mean excess plot of Section 1; (c) stability plot of Section 2; (d) mean excess plot of Section 2.

Table 13. Predicted number of collisions and respective 95% confidence intervals for each motorway section, obtained with BM and POT approaches

	BM		POT	
Section ID	Pred.	95% C.I	Pred.	95% C.I.
1	1.18	[0.02;3.59]	0.68	[0.00;2.67]
2	0.63	[0.00;2.51]	1.32	[0.06;3.82]
3	0.03	[0.00;0.32]	0.66	[0.02;2.29]
4	1.90	[0.08;5.09]	1.84	[0.01;4.88]
5	0.03	[0.00;0.24]	0.18	[0.00;1.11]
6	0.20	[0.00;1.32]	0.25	[0.00;1.44]
7	0.34	[0.00;1.84]	0.43	[0.07;1.23]
8	0.06	[0.00;0.37]	0.36	[0.01;1.36]
9	0.87	[0.01;2.99]	1.60	[0.15;4.21]
10	0.01	[0.00;0.10]	0.37	[0.00;1.78]
11	0.48	[0.00;1.98]	0.18	[0.01;0.65]
12	0.22	[0.00;1.21]	0.06	[0.00;0.40]
13	0.48	[0.00;2.06]	0.42	[0.00;2.06]
14	0.26	[0.00;1.18]	0.51	[0.01;1.88]
15	0.10	[0.00;0.83]	0.39	[0.00;1.65]
16	0.10	[0.00;0.73]	0.47	[0.00;1.90]
17	0.04	[0.00;0.37]	0.14	[0.00;0.94]
18	0.38	[0.00;2.05]	0.17	[0.00;1.14]
19	1.11	[0.03;3.52]	0.64	[0.00;2.34]

Table 14. Estimated GP parameters and respective standard errors and 95% confidence intervals, for each motorway section

Section ID	Threshold	Exceedances	Shape	Scale
1	-0.55	528	-0.265 (0.038) [-0.340;-0.189]	0.174 (0.010) [0.156;0.195]
2	-0.41	572	-0.293 (0.035) [-0.362;-0.223]	0.143 (0.008) [0.129;0.159]
3	-0.58	283	-0.122 (0.052) [-0.225;-0.020]	0.133 (0.010) [0.114;0.155]
4	-0.44	219	-0.334 (0.069) [-0.470;-0.199]	0.182 (0.017) [0.151;0.219]
5	-0.51	474	-0.222 (0.035) [-0.292;-0.153]	0.136 (0.008) [0.122;0.152]
6	-0.44	183	-0.451 (0.052) [-0.554;-0.349]	0.208 (0.018) [0.177;0.246]
7	-1.37	11989	-0.265 (0.006) [-0.277;-0.253]	0.388 (0.004) [0.380;0.396]
8	-0.55	354	-0.114 (0.040) [-0.192;-0.036]	0.113 (0.007) [0.100;0.129]
9	-0.61	571	-0.204 (0.038) [-0.278;-0.130]	0.176 (0.010) [0.158;0.196]
10	-0.46	571	-0.250 (0.035) [-0.318;-0.182]	0.136 (0.007) [0.122;0.151]
11	-0.83	1372	-0.277 (0.023) [-0.323;-0.231]	0.275 (0.010) [0.256;0.294]
12	-1.36	6484	-0.324 (0.007) [-0.339;-0.310]	0.453 (0.006) [0.441;0.466]
13	-0.53	163	-0.294 (0.069) [-0.429;-0.160]	0.187 (0.019) [0.153;0.228]
14	-0.54	402	-0.216 (0.037) [-0.289;-0.143]	0.151 (0.009) [0.134;0.170]
15	-0.75	846	-0.393 (0.021) [-0.435;-0.352]	0.232 (0.010) [0.213;0.253]
16	-0.70	700	-0.298 (0.026) [-0.348;-0.248]	0.212 (0.017) [0.181;0.249]
17	-0.69	225	-0.241 (0.051) [-0.340;-0.142]	0.187 (0.012) [0.164;0.213]
18	-0.54	290	-0.318 (0.036) [-0.389;-0.246]	0.471 (0.017) [0.439;0.505]
19	-1.15	975	-0.397 (0.021) [-0.437;-0.357]	0.261 .028) [0.211;0.322]

2.5.2.3 Analysis of results

In EVT applications to road safety it is sometimes possible to compare predicted collisions with the number of collisions which actually occurred during the observation time. This is usually the case of driving simulator experiments (Farah & Azevedo, 2017; Tarko, 2012).

In real-world it is not possible to directly compare predicted and observed collisions because, in general, crashes are rare events. Therefore, usually predicted collisions are compared with historical crash data, aggregated over several years.

The physical location where collision happens is usually the same location in which surrogate measure is recorded. For example, Songchitruska & Tarko (2006) used as surrogate measure PET to estimate the number of right-angle collisions at signalized intersections. They collected these data videotaping traffic at 12 intersections in Lafayette area, Indiana. For each of these intersections they compared the estimated number of

crashes with the historical number of crashes (collected over a 4-year period). In this case, collection and accident point were virtually the same exact place, i.e., the conflict point.

In a freeway case study, L. Zheng et al. (2014) were interested in predicting lane-change-maneuver collisions using as surrogate measure PET. They videotaped traffic on 29 road segments in Jinzhu, Yuegan and Kaiyang freeways in China. They compared the crash estimation in each segment with the historical crashes that happened within the observed segments. In their case collection and accident point were not the same place, but both points were located inside a certain road segment, identified by the camera angle.

In the present work, (as well as in Gecchele et al., 2019), the surrogate measure (TTC) is collected in a motorway cross-section, but the objective is to predict collisions that happen inside a road segment. The issue is identifying the road segment size, as the number of actual crashes depends on its length: the longer the segment, the more accident happened and vice-versa. In Gecchele et al. (2019) it was not possible to give a definitive answer to this issue, since only one section was analyzed; on the other hand in this work we can identify the segment length for which we obtain, on average across the 19 road sections, the best performance.

Six different segment lengths were considered: 100, 200, 500, 1000, 2000 and 5000 meters, defined so that the motorway section is located exactly half-way inside the segment. Taking into account the precision on the crash data location and the uncertainty in the predicted number of collisions, in this analysis the aim is not to find the exact segment length for which the model performance is optimal, but rather to identify the order of magnitude of segment length for which the model performs best.

The model performance was analyzed considering the following indicators: mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and the proportion of sections for which the observed number of collisions fell within the predicted confidence interval (Obs. Within CI). Table 15 compares results from BM approach and observed collisions, for each segment length; Table 16 compares results from POT approach and observed collisions, again for each segment length.

It is evident that the models do not perform well when the segment has a 100-meter length (overestimation of crashes) and when the segment has 2000 or 5000-meter length (underestimation of crashes). It is less straightforward to determine the best segment between 200, 500 and 1000 meters of length: the 1000-meter segment provides the lowest ME in both approaches, but higher MAE and RMSE; the 200-meter segment provides the lowest MAE with the BM approach; the 500-meter segment provides the lowest MAE with the POT approach and the lowest RMSE in both approaches, as well as including for 94.7% of the segments (18 out of 19) the observed collisions within the 95% POT confidence interval. Considering these results, the best segment length appears to be 500 meters.

Table 15. Performance indicators of BM approach predictions for six motorway segment lengths

Indicator	L=100	L=200	L=500	L=1000	L=2000	L=5000
Mean Error	0.330	0.251	0.102	-0.083	-0.679	-2.232
Mean Abs. Error	0.369	0.357	0.374	0.451	0.786	2.313
RMSE	0.579	0.544	0.531	0.587	0.938	2.570
Obs. Within CI	89.5%	89.5%	89.5%	89.5%	68.4%	26.3%

Table 16. Performance indicators of POT approach predictions for six motorway segment lengths

Indicator	L=100	L=200	L=500	L=1000	L=2000	L=5000
Mean Error	0.456	0.395	0.220	0.035	-0.561	-2.114
Mean Abs. Error	0.483	0.421	0.395	0.487	0.677	2.146
RMSE	0.667	0.607	0.580	0.611	0.858	2.465
Obs. Within CI	84.2%	89.5%	94.7%	89.5%	73.7%	42.1%

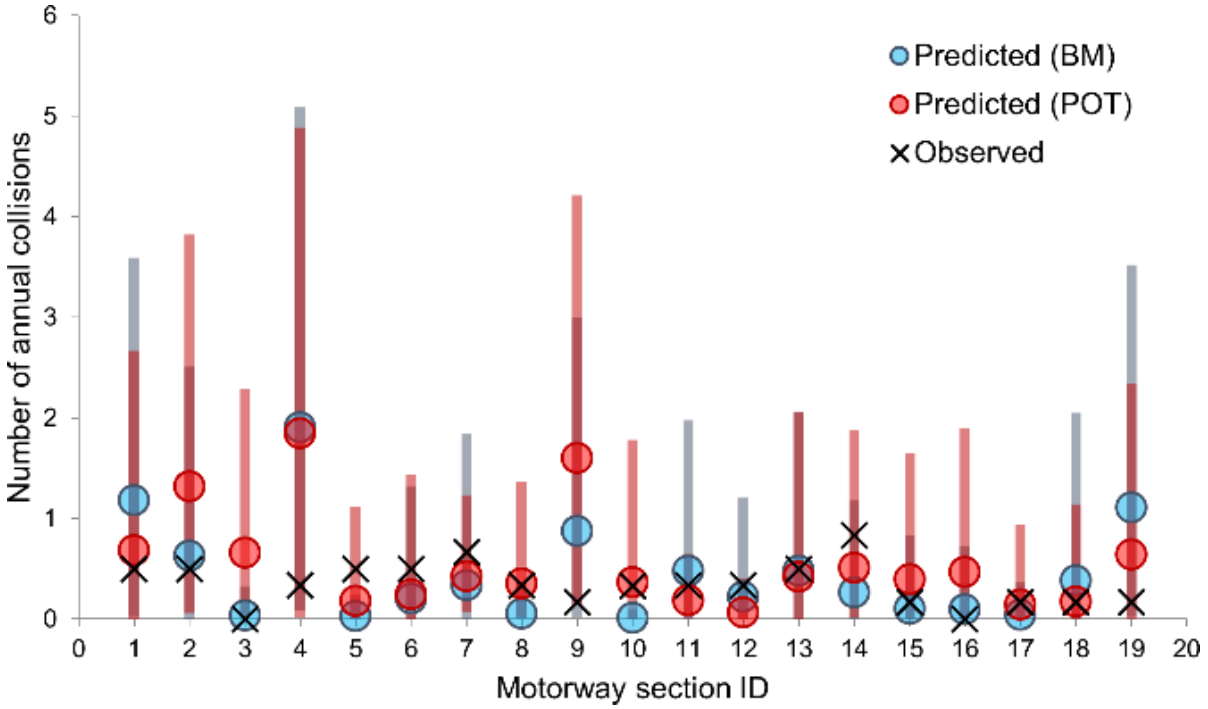

Figure 21. Comparison between predicted and observed (within a 500-meter segment) annual number of collisions, for each motorway section. Blue and red bars represent respectively predicted BM and POT 95% confidence intervals.

Even though for some sections (see Table 13 and Figure 21) BM and POT approaches produce quite different results, the overall performance of the models, considering a segment length of 500 meters, is similar. The POT approach has slightly higher values of ME, MAE and RMSE, but a higher percentage of sections for which the observed number of collisions falls within the confidence interval. This work shows once more that it is not easy to claim which of the two approaches performs better. However, previous studies suggested that the POT method is able to make a more efficient use of the data (L. Zheng et al., 2014a). It is also something that Coles affirms in more general terms, describing the Block Maxima as a "wasteful approach to extreme value analysis" (Coles, 2001). Indeed, as can be observed by looking at Table 15 and Table 16, the POT approach is in general able to extract information from more data, as the number of exceedances is often higher than the number of blocks of the BM approach; this may not affect model performance when the observation time is long and both the number of blocks and exceedances are high, but could result crucial for shorter observation times.

Compared to other works in EVT (Songchitruksa & Tarko, 2006; L. Zheng et al., 2014a; L. Zheng, Ismail, et al., 2018), the indicators analyzed for both BM and POT approaches are significantly better, although it must be noted that in this case-study the observation period was much longer than in the previously mentioned works.

2.5.3 Conclusion

This Section presented a large-scale application of extreme value theory in the field of road safety. With this method it is possible to obtain reliable crash predictions without the use of crash data, but rather analyzing surrogate measures of safety.

The EVT case study presented in this Section was the largest, in terms of observation period, retrieved in the road safety literature up to now: 19 cross-sections on two Italian motorways were monitored for one year. The two more established EVT approaches were applied: the block maxima and the peak-over threshold. The main objective of the work, i.e. applying EVT to reliably predict motorway rear-end collision, was achieved, since both approaches produced average annual crash predictions close to the observed number of crashes.

This work showed how an extended observation period can produce reliable and consistent rear-end collision risk predictions in motorway cross-sections using EVT. With this approach it is possible to exploit traffic data which are already collected for other purposes (e.g., traffic management) to provide quick (compared to the use of crash data), reliable and inexpensive road safety evaluations. With the introduction of ATSM in the POT approach, it is possible to fit EVT distributions avoiding the time-consuming and subjective operations of the traditional graphical diagnostics method, with hugely positive practical implications.

The results of this work can be used as benchmark for future and further analysis. In particular, the large observation period and study area could be exploited to investigate issues such as seasonality, transferability (see Appendix 1), and the effect of the observation period length. Moreover, other EVT (or non-EVT) modelling approaches could be compared using data collected during this study.

In the future, EVT could be applied on data coming from road segment observations, employing information obtained with new ITS technologies, thanks to vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communications, which are being more and more utilized in the field of transport (Learn et al., 2018; Shaikh & Thalkar, 2019; R. Wang et al., 2019).

3. Real-time road safety analysis[4]

Although conflict analysis is well established in road safety, it has not yet found its way into real-time road safety prediction models, except for very few works, which are discussed in Section 3.1.2. To fill this gap, in this Chapter, we propose a methodology to predict conflict-prone situations based on real-time traffic data.

In practical terms, a real-time conflict prediction model (RTConfPM) is structured similarly to an RTCPM. The flowchart in Figure 22 shows its main features:

- The model is developed offline using historical traffic, weather, and road data as input and historical crash/conflict data for calibration.
- The model is applied in real time using traffic, weather, and road data collected at the present interval to predict whether an unsafe situation will exist in the next future.
- If an unsafe situation is predicted, then an intervention strategy is applied to prevent it. Several solutions can be applied, such as variable message signs or ramp metering (Pande & Abdel-Aty, 2005).

[4] The work presented in this Chapter is part of the following publication:

Orsini, F., Gecchele, G., Rossi, R., & Gastaldi, M. (2021). A conflict-based approach for real-time road safety analysis: comparative evaluation with crash-based models. *Accident Analysis & Prevention*, *161*, 106382. (Orsini, Gecchele, Rossi, et al., 2021)

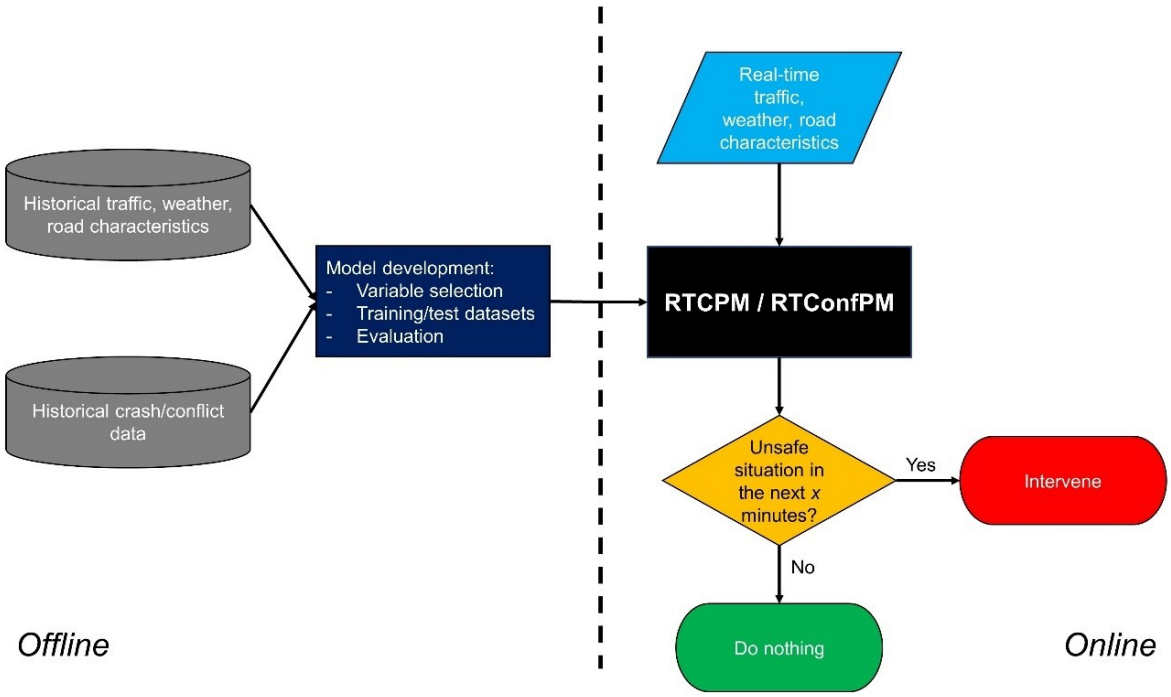

Figure 22. Flowchart representing the offline development of an RTCPM/RTConfPM (left) and the real-time application of an RTCPM/RTConfPM.

A major advantage of an RTConfPM is that it can be developed even in absence of highly spatial/temporal-accurate crash data,[5] which can be a huge benefit in practical applications, since inaccuracies in crash location and time are among the most serious quality issues of crash data (Imprialou & Quddus, 2019). In this work, we further suggest that this approach can identify real-time risky conditions more reliably than traditional RTCPMs. This is discussed in quantitative terms in the next sections of the dissertation, but is also intuitively illustrated in Figure 23. Let us consider a first situation, in which a set of traffic characteristics induced the occurrence of several traffic conflicts within a certain time window, but no actual collision ensued. Let us now consider a second situation, in which the same set of traffic characteristics produced the same amount of traffic conflicts, and, due to a road user's ineffective intervention, a crash ensued. With the RTConfPM that set of traffic characteristics induced an unsafe situation in both cases; with the RTCPM, the same set of traffic conditions was considered, in one case, a crash precursor, and in the other case not, although the only difference between the two situations was the effectiveness of a road user's intervention, which is independent from the input data of the model (i.e., traffic characteristics). This inconsistency can have negative consequences on the effectiveness of model training.

[5] However, crash data are still needed, at least in some aggregate form, to initially validate the definition of conflicts and unsafe situations used in the RTConfPM. This is further discussed in Section 3.3.2.

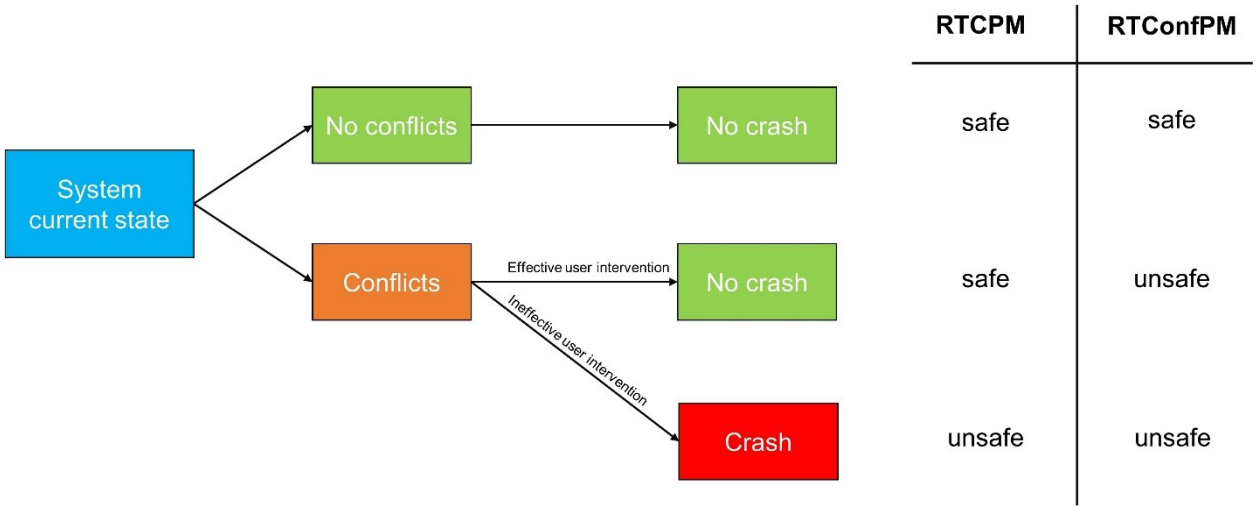

Figure 23. Safe and unsafe situations for RTCPMs and RTConfPMs.

Another advantage of dealing with near-crashes is that they are much more frequent than crash events (see the classic "safety pyramid" introduced by Hydén, 1987), resulting in shorter observation times to collect enough data to train an RTConfPM. Moreover, recent developments in sensing and ITS technologies significantly improved the availability and accuracy of traffic measures (Sadeq & Sayed, 2016; Tarko, 2012), especially on highways (Eltrass & Khalil, 2018; St-Aubin et al., 2015).

Starting from the considerations discussed above, here we aim to develop an RTConfPM to predict rear-end crashes on motorways.

The first main novelty introduced by this work is that the results of the proposed conflict-based approach are compared to those of a traditional crash-based model calibrated on the same dataset. To the best of the authors' knowledge, this represents the first direct comparison between the two approaches in real-time road safety applications.

The second main novelty is that cross-sectional vehicle-by-vehicle data were used both to retrieve input traffic variables and to detect traffic conflicts. This was done by means of microwave radar sensors, which can produce reliable information with relatively low data collection and elaboration costs, and have great potential in view of future large-scale practical implementations.

Finally, as discussed in Section 3.5, this work can be viewed as a starting point for several research developments.

3.1 Literature review

The rationale behind an RTConfPM is not dissimilar to that of an RTCPM: given the system current state (composed by a set of roads, traffic, weather characteristics, etc.), an RTConfPM aims to predict the occurrence of an unsafe situation in the immediate future. The only difference, as discussed before, consists of how unsafe situations are defined in the two approaches. Therefore, methodologically, an RTConfPMs adopts the same tools as RTCPMs do to identify relationships between the system current state and unsafe situations and, therefore, the choice of which tools to use cannot be made without an appropriate review of the methods employed in RTCPMs.

3.1.1 Real-time crash prediction models

The first works dealing with real-time assessment of road safety date back to the end of the last century. Madanat and Liu (1995) included traffic conditions measured by surveillance sensors in their incident detection model in order to produce estimates of prior probabilities of incidents; Hughes and Council (1999) analyzed the relationship between freeway safety and peak period operations, observing that macroscopic measures of flow (e.g., annual average daily traffic, hourly volume) were shown to correlate poorly to real-time system performance, whereas variability in vehicle speeds was identified as a measure that significantly affects crashes. A few years later, Lee et al. (2002) developed a log-linear model to predict crashes through the estimation of crash precursors, and Abdel-Aty et al. (2004) developed a logistic regression model to predict crashes based on real-time traffic flow variables collected from loop detectors.

Since then, several RTCPMs have been developed. Recently, a systematic review by Hossain et al. (2019) identified and analyzed 78 articles published between 2003 and 2018 dealing with real-time prediction of crashes.

The main objective of RTCPMs is, of course, to predict crashes; some studies also considered crash severity (Park et al., 2018; C. Xu, Tarko, et al., 2013) and/or crash type (Wu, Abdel-Aty, Cai, et al., 2018; Yu & Abdel-Aty, 2013a). Other papers dealt with crash mechanisms (Yu et al., 2013) or intervention strategies (Abdel-Aty & L. Wang, 2017; Pande & Abdel-Aty, 2005).

The source of traffic data is, more frequently, loop detectors (Dimitriou et al., 2018; Roy et al., 2018); some studies employed microwave vehicle detection systems (L. Wang et al., 2017; Yasmin et al., 2018), automated vehicle identification (Ahmed et al., 2012; Basso et al., 2018), Bluetooth detectors (Yuan et al., 2018; Yuan & Abdel-Aty, 2018), remote traffic microwave stations (Wu, Abdel-Aty, Cai, et al., 2018; Yu et al., 2014), probe vehicles (Park & Haghani, 2016), or combinations of these sources.

The main input in RTCPMs consists of traffic flow variables, typically a subset of average, standard deviation, and other statistics or transformations, of speed, flow and occupancy; in addition to these, it has been shown that considering traffic flow composition can improve model performance (Basso et al., 2020). Data are aggregated at different temporal scales, depending mainly on the available raw data, usually in the range of 1 to 5 minutes, with the vast majority of studies using 5-minute aggregation (Hossain et al., 2019). Moreover, several other non-traffic variables have been used as input in RTCPMs, including: geometrical features of the road, such as the presence of curves (Ahmed et al., 2012) and pavement condition (L. Wang et al., 2017); weather conditions, such as the presence of fog (Wu, Abdel-Aty, & Lee, 2018) and lighting conditions (Yasmin et al., 2018); and other variables, such as the presence of neighborhoods with younger populations (Paikari et al., 2014).

This large number of variables often produces a large variable space, compared to the usually relatively small sample size of crashes, leading to the need to apply variable selection methods, to discard the less meaningful ones. The recent trend is to apply artificial intelligence and data mining techniques, such as classification trees (Pande & Abdel-Aty, 2006), random forest (Abdel-Aty & Haleem, 2011), or random multinomial logit (Hossain & Muromachi, 2012).

Modeling methods used to develop RTCPMs are generally divided into statistical and machine learning approaches. Among the former, multinomial logistic regressions (Abdel-Aty et al., 2004) and Bayesian logistic regression (Ahmed et al., 2012) have been widely applied. Among the latter, several classification methods have been used: support vector machine (SVM - Qu et al., 2013; Yu and Abdel-Aty, 2013b), classification trees/random forest (Hossain & Muromachi, 2011), k-Nearest Neighbor (kNN - Lv et al., 2009a), Bayesian networks (Hossain & Muromachi, 2012), and various forms of neural networks (Pande & Abdel-Aty, 2006) and deep neural networks (P. Li et al., 2020; Yuan et al., 2019). Recently, Theofilatos et al. (2019) applied and compared several machine learning and deep learning techniques on the same dataset: kNN, Naïve Bayes, classification tree, random forest, SVM, shallow neural network, and deep neural network, observing that the deep learning approach yielded the best results, but other far less complex methods such as Naïve Bayes performed only marginally worse.

Regardless of the model employed, an RTCPM is, in general, a binary classification problem, in which the two classes are crash and no-crash. As crashes are rare events, the dataset used to train the model is unbalanced, with far more time intervals in which no crashes were observed, than those in which a crash happened. In an unbalanced classification problem, models tend to classify all of the sample into the major class, due to the inherent motivation of their objective functions, which minimize the sum of errors by assigning the same weight to both the major and minor classes, whereby samples in the minor class make little contribution (Sun et al., 2007). There are two possible approaches to deal with this issue: algorithm-level solutions and data-level solutions (Ke et al., 2019).

The former approach consists in applying cost-sensitive learning techniques, in which different weights are assigned to the classes in the objective function, to penalize the misclassification of minority class samples. In the latter approach, the training dataset is resampled to obtain a more balanced problem, with several sampling methods existing in the literature (He & Garcia, 2009). In RTCPM literature, data-level solutions are usually employed, typically with a matched case-control method (Abdel-Aty et al., 2004), although recently, other more sophisticated approaches, such as synthetic minority oversampling technique (SMOTE), have started to be used (Basso et al., 2018; P. Li et al., 2020).

3.1.2 Real-time conflict prediction models

There is a limited, but rapidly growing, literature on real-time conflict prediction models. To the best of the authors' knowledge, the only attempts to predict conflict-prone conditions on a road segment in real-time were made by Caleffi et al. (2017), Formosa et al. (2020), Katrakazas et al. (2017, 2018).

Similarly to this work, Katrakazas et al. (2017, 2018) motivated the need for an RTConfPM on the difficulty of retrieving the high spatial-temporal precision crash data that an RTCPM needs. Contrarily to the present work, Katrakazas et al. (2017, 2018) did not have access to real-world vehicle-by-vehicle traffic data, but only to 15-minute aggregated data; to overcome this, they used such data as input to a microsimulation model, to reproduce vehicle trajectories and retrieve simulated conflicts. Conflicts were identified based on some surrogate measures of safety thresholds/rules of thumb from the literature, but no ad hoc analysis of crash-conflict relationships was performed. They applied three machine learning classifiers: SVM, random forest, and kNN; however, differently from the authors of this work, they did not compare the results of their RTConfPM to those of an RTCPM trained on the same dataset.

Caleffi et al. (2017) applied linear discriminant analysis to predict conflicts on a Brazilian freeway section. They were able to detect conflicts in a disaggregate manner, retrieving them from video recordings. Their approach to detecting conflicts, however, was significantly different from that of the present work: conflicts were defined in qualitative terms and detected only by observing evasive actions, instead of being quantitatively determined by surrogate measures of safety (see Section 3.3.2).

Recently, Formosa et al. (2020), proposed an RTConfPM in which input traffic data were collected with both loop detectors and in-vehicle sensors, and conflicts were identified with an instrumented vehicle. As in Katrakazas et al. (2018), conflicts were defined according to general rules and thresholds from the literature. The predictive model adopted was a deep neural network, and the results obtained were promising. The data acquisition process in their study is quite different from that of the present study, as it involves instrumented vehicles; its complexity can be regarded as providing an opportunity to handle complex relationship between heterogeneous variables, as well as acting as a potential limitation for

a large-scale real-world implementation, at least in the short-medium term. No direct comparison with a crash-based approach was made in their work.

In a different environment (i.e., intersections), the recent works of Essa and Sayed (2018, 2019) and L. Zheng & Sayed (2020) developed a conflict-based model for real-time safety evaluation of signalized intersections, using generalized linear models, full-Bayesian and extreme value theory approaches. Traffic input variables included queue length, shock wave speed and area, and the platoon ratio; time-to-collision (TTC), modified TTC, and the deceleration rate to avoid the crash were used as traffic conflict indicators. Data were gathered at several intersections for some hours using video cameras. Their works lacked a direct comparison with crash-based approaches.

Another work worth mentioning is by Osman et al. (2019), who predicted near-crashes based on the vehicle kinematics variables recorded immediately beforehand, using several machine learning classifiers. A similar approach was previously proposed by Pin Wang & Chan (2017). The point of view adopted by these studies is quite different from that of the present work, as it was exclusively vehicle-based, rather than infrastructure-based: their aim was not to predict unsafe situations on a certain road segment, but rather to develop a tool to support crash avoidance systems and/or ADAS.

3.2 Methodology

In this work, we developed and compared an RTConfPM and an RTCPM. Methodologically, they both follow the same typical scheme, as shown in Figure 22.

3.2.1 Variable selection model: Random forest

Random Forest (RF) is an ensemble machine learning classification model composed by multiple tree classifiers. The main advantage of RT over a simpler decision tree is that it can provide more stable and unbiased estimates of the classification error. Moreover, as the number of trees within the RF grows, the results are more robust against over-fitting.

Each tree of the RF is grown following this procedure (Breiman L., 2001): (i) given N records in the training set, a bootstrap sample n of N is drawn for growing the tree; (ii) given M variables, a constant number of $m<M$ variables are selected at each node of the tree, and the best split among m is used; (iii) the tree is grown fully, without pruning.

The $N - n$ left-out records are called out-of-bag (OOB) data and they are in fact an internal test dataset. OOB data are used to obtain an unbiased error estimate, and to assess the variables importance. Given p, one of the M variables, the following procedure is followed:

- OOB data are fed to all the trees in RF and the number of correct classifications counted.
- All the values of variable p are permuted randomly, the permuted data are put through the RF and again correct classifications are counted.
- The raw importance of variable p is the average difference between the two counts obtained at the previous steps.

An alternative importance measure is the total decrease in Gini node impurities from splitting on the variable p, averaged over all trees. The two importance metrics are generally consistent with each other; however, the former has the drawback that it tends to overestimates the variable importance of highly correlated variables (Strobl et al., 2008).

Random Forest are widely used in RTCPMs for variable selection (Abdel-Aty et al., 2008; Abdel-Aty & Haleem, 2011; Ahmed et al., 2012; Basso et al., 2018; Shi & Abdel-Aty, 2015; C. Xu, W. Wang, et al., 2013), and in some cases also as a classifier itself (Theofilatos et al., 2019).

3.2.2 Training dataset resampling: SMOTE

As mentioned in Section 3.1.1, the unbalanced classification issue can be resolved by resampling the training dataset in order to make it more balanced. The easiest way to achieve this is to randomly under-sampling the majority class or to randomly over-sampling

the minority class. In the first case randomly selected samples of the majority class are removed from the dataset, in the second case, randomly selected samples of the minority class are replicated in the dataset. However, in the case of random under-sampling, removing samples from the majority class can cause the classifier to miss important aspects of the problem, leading to poor model performance, especially if the number of minority class samples is very small; in the case of random over-sampling, the replication of certain samples may induce the model to give too much importance to certain aspects of the problem and to cause over-fitting problems (He & Garcia, 2009).

A more advanced solution to the problem is to apply the Synthetic Minority Oversampling TEcnhique (SMOTE), proposed by Chawla et al. (2002). Instead of using duplicate samples, SMOTE generates synthetic samples of the minority class. The algorithm follows a relatively simple and computationally inexpensive procedure:

- For each minority class sample, k nearest neighbor minority class samples are selected.
- For each neighbor, the difference between the original sample feature vector and the neighbor feature vector is computed.
- The difference is multiplied by a random number between 0 and 1 and added to the original sample feature vector, generating a new synthetic sample of the minority class.

For example, in a bidimensional feature space, the synthetic sample would be a random point on the line segment between the original sample and the neighbor sample.

In this way, given k, the minority class is oversampled by a factor k+1. Moreover, SMOTE can be used in combination with the under-sampling of the majority class, in order to obtain the desired ratio between the classes (Chawla et al., 2002).

SMOTE has recently been applied with good results in several recent RTCPMs (Basso et al., 2018; Katrakazas et al., 2019; Ke et al., 2019; P. Li et al., 2020; Parsa et al., 2019; Yuan et al., 2019).

3.2.3 Classifier: support vector machine

Support Vector Machine (SVM) is a supervised machine learning method for classification and regression analysis introduced by Vapnik (1995). In a classification problem, the basic idea behind SVM is to find the hyperplane that separates all data points, maximizing the distance between the classes (i.e., the "margin").

Given a set of vectors x_j (j=1,…,n) with dimension d (i.e. the dimension of the feature space) and their respective classes y_j, where y_j can either be equal to 1 (e.g., in our case an unsafe situation) or -1 (e.g. a safe situation). In the case of linearly separable data, the hyperplane is:

$$f(x) = wx + b \tag{34}$$

If $y_j = 1,\ x_j w + b \geq 1$, if $y_j = -1,\ x_j w + b \leq -1$. This can be combined into:

$$y_j\left(x_j w + b\right) \geq 1 \quad \forall j \tag{35}$$

It can be proved (Cortes & Vapnik, 1995) that, in order to maximize the margin, the *w* vector must be the solution of the following optimization problem:

$$\min_w \frac{1}{2}\|w\|^2 \quad s.t.\ \ y_j\left(x_j w + b\right) \geq 1,\ \ j = 1, \dots, n \tag{36}$$

However, in many applications, RTConfPM/RTCPM included, data are hardly linearly separable. In that case, SVM can use a soft margin, i.e. a hyperplane that separates many (yet not all) data points. This is done by adding slack variables ξ_j to penalize misclassification, and optimization problem becomes (Cortes & Vapnik, 1995):

$$\min_w \frac{1}{2}\|w\|^2 + C\sum\nolimits_j \xi_j \tag{37}$$

$$s.t.\ \ y_j\left(x_j w + b\right) \geq 1 - \xi_j,\ \ \xi_{,} \geq 0,\ \ j = 1, \dots, n$$

It is computationally simpler to solve the dual quadratic programming problem. By introducing Lagrange multipliers a_j, the optimization problem has the form:

$$\max_\alpha \sum\nolimits_j \alpha_j - \frac{1}{2}\sum\nolimits_j \sum\nolimits_k \alpha_j \alpha_k y_j y_k K\left(x_j, x_k\right) \tag{38}$$

$$s.t.\ \ \sum\nolimits_j \alpha_j y_j = 0,\ \ 0 \leq \alpha_j \leq C,\ \ j = 1, \dots, n$$

In (5), *K* is a kernel function, which in the linear case is simply $K\left(x_j, x_k\right) = x_j^T x_k$. However, some binary classification problems do not have a simple hyperplane as a useful separating criterion, and in that case non-linear kernel functions, such as polynomial or radial basis (RBF), can be employed.

SVM is one of the most commonly used classifiers in RTCPMs; it was introduced in the works of (Lv, Tang, Zhao, et al., 2009; Qu et al., 2013; Yu & Abdel-Aty, 2013b), and recently applied in (Theofilatos et al., 2019; J. Wang et al., 2019; L. Wang et al., 2019), sometimes in combination with SMOTE (Basso et al., 2018; Parsa et al., 2019).

3.2.4 Performance indicators

Several performance metrics can be used to evaluate an RTCPM or an RTConfPM. Many of the indicators can be derived from the confusion matrix, which, in the case of binary classification, is a 2x2 array composed of the following components (see Figure 24): true positive, true negative, false positive, and false negative.

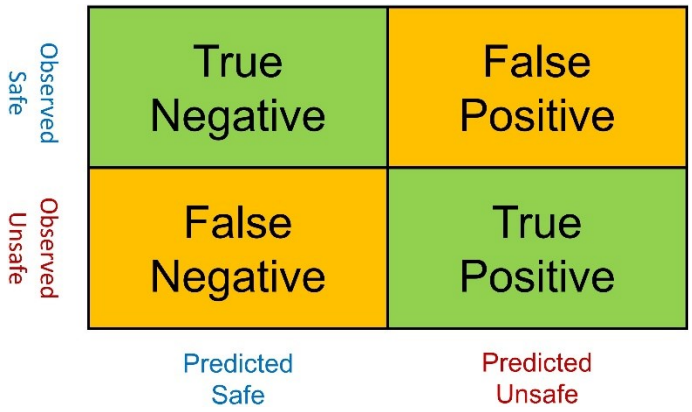

Figure 24. Confusion matrix for the binary unsafe/safe classification problem.

The indicators considered in this work, and computed as in Table 17, are as follows (Stehman, 1997; Tharwat, 2018):

- Accuracy: the number of safe and unsafe situations correctly predicted by the model, divided by the total number of instances. This measure should be treated very cautiously when classes are unbalanced.
- Recall (also known as sensitivity or true positive rate): the number of correctly predicted unsafe situations divided by the total number of observed unsafe situations. A low-recall model indicates the model missed some unsafe situations.
- Specificity (also known as true negative rate): the number of correctly predicted safe situations divided by the total number of observed safe situations. A low-specificity model indicates the model considered many of the safe situations unsafe. The complementary measure (1 − specificity) is the false alarm rate (or false positive rate).

Another useful tool to evaluate a classification model is the receiver operating characteristic (ROC) curve. The ROC curve is created by plotting the true positive rate (i.e., recall) against

the false positive rate (i.e., false alarm rate) at various threshold settings. From this curve it is possible to derive a widely used performance indicator: the area under the ROC curve (AUC). AUC is the probability that a classifier will rank a randomly chosen positive instance higher than it would a randomly chosen negative one (Fawcett, 2006). AUC is considered a measure of aggregated classification performance (Flach et al., 2011).

To summarize, accuracy and AUC can be used for an overall evaluation of the model performance, whereas recall and specificity can assess specific aspects of the model. In road safety applications, recall is particularly important because it quantifies the influence of the number of false negatives: a model that marks situations "safe" when they are actually unsafe must be avoided because it precludes the possibility of intervening and avoid a crash. Specificity quantifies the impact of false positives; although high-specificity models are preferable, a model that triggers too many false alarms is a problem mainly from an operational point of view instead of a safety point of view.

This set of performance indicators is commonly used to evaluate RTCPMs and RTConfPMs (Caleffi et al., 2017; P. Li et al., 2020; Pengfei Wang et al., 2020; Yuan et al., 2019).

Table 17. Equations of performance indicators used to evaluate the models developed in this work. TP=true positive; TN=true negative; FP=false positive; FN=false negative.

Indicator	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Recall	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
False alarm rate	$\frac{FP}{TN+FP} = 1 - \frac{TN}{TN+FP}$

3.3 Case study: real-time conflict prediction

An RTConfPM was developed using data collected on a three-lane Italian motorway for 1 year, from January 1st to December 31st, 2013 (the same as in the case study presented in Section 2.5). The motorway is about 150km long, and there are 20 cross-sections in each direction (40 in total) that are equipped with microwave Doppler radars. The average spacing of the detectors is 7.58 km, with minimum spacing of 0.76 km and maximum of 16.07 km.

In an RTConfPM, input data (see Section 3.3.1) collected during a given interval are used to predict whether the risk of an unsafe situation occurring (defined according to Section 3.3.2) will increase in the following interval. Conflict data are derived from surrogate measures of safety for model calibration and validation, whereas they are not used during the live real-world application of the model (see Figure 22).

3.3.1 Traffic data

Over each lane of each motorway cross-section, a dedicated radar collects the following vehicle-by-vehicle information: speed, time gap, time stamp and vehicle class. These data, originally available in disaggregated form, were then aggregated in 5-minute intervals, a length chosen to align this dissertation with the vast majority of works in the RTCPM/RTConfPM literature (see Section 3.1).

Data aggregation provided the following variables for each of the three lanes:

- Traffic volume (*Vol*): the total number of vehicles.
- Heavy-duty traffic volume (*HVol*): the percentage of vehicles classified as heavy-duty trucks.
- Speed (*Sp*): harmonic mean of vehicle speeds.
- Speed variance (*SdSp*): standard deviation of vehicle speeds.

Therefore, for each cross-section, 12 variables are available in 5-minute aggregation. In the remainder of the Section, suffixes R, C, and L will indicate the rightmost lane, the center lane, and leftmost lane, respectively. Descriptive statistics are presented in Table 18.

Traffic volume, average speed, and standard deviation of speed are variables commonly used in real-time road safety analysis, and they are particularly relevant in the case of rear-end crashes (Dimitriou et al., 2018). In this case study, speeds are recorded on a cross-section, but the aim of the model is eventually to detect conflict-prone and crash-prone situations near the cross-section (i.e., on a road segment). For this reason, to avoid overestimating the influence of faster vehicles on the mean speed of such segment, we considered space-mean-speed instead of time-mean-speed (Knoop et al., 2009). The

former, under the assumption of stationary conditions, is given by the harmonic mean of speed (Daganzo, 1997).

Information on vehicle type is not often used in real-time road safety analysis. However, Basso et al. (2020) underlined the importance of considering flow composition in an RTCPM. Another peculiarity of the variables adopted in this work is that they were considered separately for each of the three motorway lanes: As can be observed in Table 18, each of the three lanes has specific characteristics and aggregating variables over the whole cross-section would cause the loss of such information.

Table 18. Descriptive statistics of candidate input traffic variables

Traffic variable	Description	Unit	Mean (SD)	95% quantile range
Vol_R	Traffic volume on right lane	veh/5min	30.0 (27.9)	0; 72
Vol_C	Traffic volume on central lane	veh/5min	42.3 (38.8)	0; 111
Vol_L	Traffic volume on left lane	veh/5min	27.4 (62.6)	0; 111
HVol_R	% of heavy-duty vehicles on right lane	%	51.2 (27.4)	0.00; 92.3
HVol_C	% of heavy-duty vehicles on central lane	%	7.19 (10.2)	0.00; 36.7
HVol_L	% of heavy-duty vehicles on left lane	%	0.47 (2.83)	0.00; 3.41
Sp_R	Harmonic mean of speed on the right lane	km/h	90.2 (9.58)	72.2; 110.0
Sp_C	Harmonic mean of speed on the central lane	km/h	112.3 (11.1)	90.6; 131.4
Sp_L	Harmonic mean of speed on the left lane	km/h	128.4 (14.4)	104.0; 154.5
SdSp_R	SD of speed on the right lane	km/h	10.5 (4.27)	3.84; 19.6
SdSp_C	SD of speed on the central lane	km/h	13.7 (4.01)	8.04; 22.6
SdSp_L	SD of speed on the right lane	km/h	11.5 (5.78)	0.00; 23.4

3.3.2 Conflict data

Considering the above-mentioned traffic data (see Section 3.3.1) in disaggregate, vehicle-by-vehicle form, it is possible to calculate the TTC between each couple of vehicles consecutively detected on the same lane, with the well-known definition reported in equation (26).

In each 5-minute interval, a TTC value for each couple of consecutive vehicles on each lane of the cross-section is available. With this kind of information, determining whether one or more conflicts have happened in that 5-minute interval is possible. However, it is neither a straightforward nor a trivial operation.

A TTC threshold *u* should be set to separate non-risky interaction events and proper conflict events; however, such a threshold cannot be defined univocally, because certain low TTC values may represent intentionally aggressive behaviors, which may still produce close but controlled interactions between vehicles (Tarko, 2019). Setting a high threshold allows consideration of all the "true conflicts" (in accordance with the definition provided in Section 1) as conflicts, but also many controlled interactions; setting a low threshold has the consequence of discarding some real conflicts, but also excludes all controlled interactions. Tarko (2018b) suggests choosing "a threshold separation sufficiently short to claim only traffic conflicts," because including controlled interactions may bias the results.

A further complication arose in this work: we needed to define if, considering the *n* conflicts observed in the 5-minute interval, that interval should be considered safe or unsafe. In this case, operational needs come into play: the lower value of *n* is selected, the lower crash risk is accepted, but the higher numbers of alarms are triggered. In addition, the choice of *n* may significantly influence model performance.

To sum up, two parameters must be chosen:

1. The TTC threshold *u* under which an interaction between vehicles is considered a conflict
2. The minimum number *n* of conflicts within the 5-minute interval to consider it "unsafe."

In order to choose the threshold *u* in a robust and objective way, this dissertation propose to apply the extreme value theory (see Section 2).

The extreme value distribution adopted in the present study is the generalized Pareto (GP), see Section 2.2.1.2. The aim of the analysis in this Section is to select an appropriate threshold *u* for a GP distribution fitted on the (negated, see Section 2.5.2) TTC values recorded in this work's case study. Such threshold will be used in the remainder of the Section to identify traffic conflicts.

The negated TTC values were used to fit several different GP distributions, progressively increasing the threshold value by 0.01 seconds. Note that we were allowed to fit a single distribution for all cross-sections because a transferability study on data from the same motorway showed that it is possible to aggregate data from several road sections, provided that they share similar geometric, traffic and weather characteristics (see Appendix 1). Otherwise, separate GP distributions should be fitted for each cross-section.

The stability plot is presented in Figure 25a, which shows that the shape parameter remains constant until the threshold is around −0.80 s. ATSM returns a value of −0.78 seconds,

which is consistent with the stability plot. Visual inspection of the goodness-of-fit of the GP distribution with threshold equal to −0.78 seconds is satisfactory (see Figure 25b). The estimated value of the scale parameter is 0.352 (standard error 0.0011); the estimated value of the shape parameter is −0.445 (standard error 0.0013).

From this distribution, extracting the crash probability is possible under the assumption that –TTC ≥ 0 represents a crash, resulting in $P_{inc} = 1 - GP(0) = 7.22 * 10^{-5}$. This probability is associated with crashes in the close vicinity of the cross-sections (about 250 m before and after the section, according to Section 2.5.2.3). Assuming a constant crash probability within 2500 m before and after the cross-section (i.e., considering a 10× multiplier), and multiplying this probability by the number of −TTC values above −0.78 (i.e., 111,859), we obtain an estimated number of 80.7 crashes within 2,500 m from the cross-sections during the observation period, with a 95% confidence interval [73.4, 90.9]. The observed number of crashes in the same area and period is 88, which falls within the confidence interval, therefore validating the goodness of the model estimated with a threshold value of 0.78.

Note that, in order to validate this procedure and choose the threshold, crash data were needed. So, even in this conflict-based RTConfPM, crash data are still necessary. It is worth pointing out, however, that these data need not to have a high spatial/temporal precision, as they are only used in aggregated form. From a practical point of view this is crucial, because it means that, contrarily to a traditional RTCPM, the proposed conflict-based approach can be adopted even in absence of highly spatial/temporal-accurate data, which may not be available for many real-world highways. Furthermore, after appropriate transferability studies, EVT models could be transferred even to highway segments where crash data are completely unavailable.

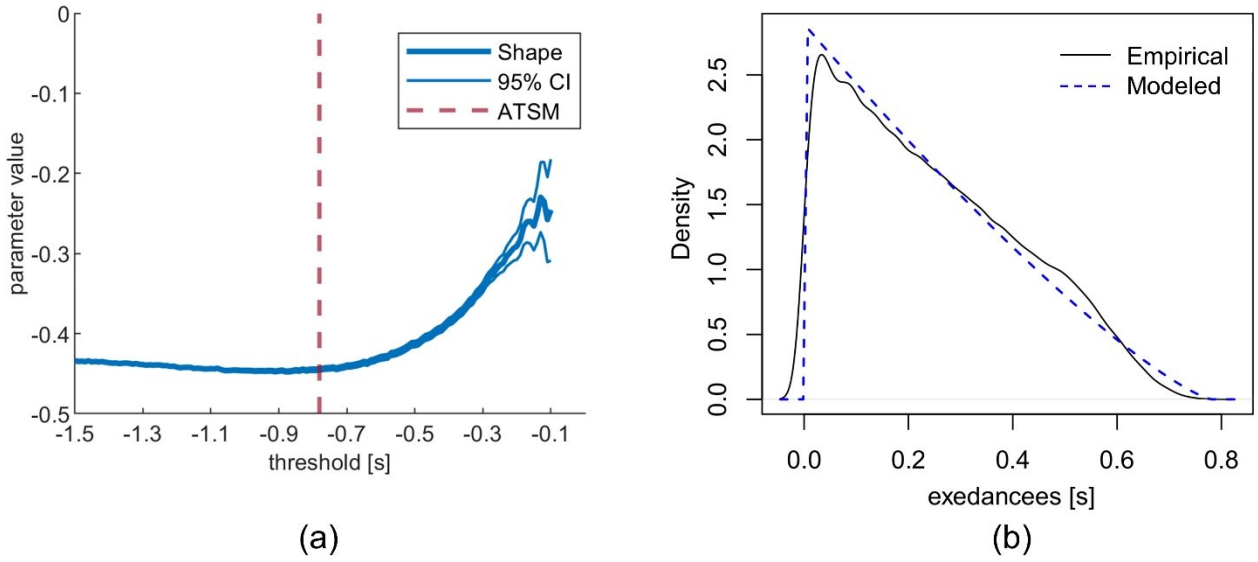

Figure 25. (a) Stability plot for shape parameter; (b) Probability density plot for a model fitted with u = −0.78s.

P_{inc} can be interpreted as the probability of a conflict resulting into a crash. Therefore, assuming independence between conflicts, if we predict that n conflicts will happen within a given 5-minute interval, then the probability of observing a crash during such interval is $n \times P_{inc}$.

Contrary to the selection of parameter u, which should be supported by a study on the conflict–crash relationship, the choice of n primarily depends on operational aspects. The choice is a compromise between the level of risk one wishes to accept, the amount of alarms that one can receive compatibly with intervention strategies, and model predictive performance.

Increasing n (i.e., accepting a higher crash risk) quickly decreases the number of "unsafe" intervals, as can be observed in Table 21. Having a high value of n may reduce the number of unsafe intervals to the point that it becomes comparable or even lower than the actual number of observed crashes, cancelling one of the advantages of the conflict-based approach over the crash-based approach (i.e., retrieving more data for training the model). However, the more intervals are considered as "unsafe," the more often alarms are triggered (regardless of the predictive power of the model), which may be undesirable under an operational point of view.

Table 19. Effect of parameter n on the number of unsafe intervals.

n	Number of unsafe intervals
1	98,397
2	10,738
3	2,241
4	408
5	64
6	10
7	1
8+	0

In terms of model performance, choosing $n = 1$ may loosen the relationship between the input traffic variable and unsafe situations. Conflicts, as crashes, may be caused not just by traffic characteristics, which are input variables to the RTConfPM, but also, for example, by unprovoked human error (see Figure 1), which is not and cannot be an input variable in the RTConfPM presented in this work. If a 5-minute interval is considered unsafe because a single conflict occurred, then we might include in the unsafe set some situations which are not strictly unsafe from a traffic variables point of view, with negative

consequences on model performance. However, choosing a high value of n reduces the numerosity of the unsafe class of intervals, which in turn reduces the size of the training dataset, potentially leading to a reduction of model performance. Figure 26 presents a sensitivity analysis of the performance of the model in the full-test dataset, while increasing n from 1 to 5. Parameters used in this sensitivity analysis are 0.78-second threshold, k = 5, 1:1 ratio of safe versus unsafe situations, linear kernel function, and Set #1 input variables. For each n, 100 SVMs were trained with Monte Carlo cross-validation technique, and average AUC, specificity and recall were computed. As n increases, the relationship between unsafe situations and traffic conditions becomes stronger and the predictive power of the model increases. However, this increase becomes marginal when n is higher than 2, and it starts to decrease when n is 5, because the size of the training dataset becomes small enough to affect the reliability of the SVM.

Therefore, according to this sensitivity analysis, the best recall performance is obtained with n = 3, and because about 25 times more unsafe intervals than crashes exist for the value of n (therefore significantly increasing the amount of data available for training the model), we chose to adopt this parameter.

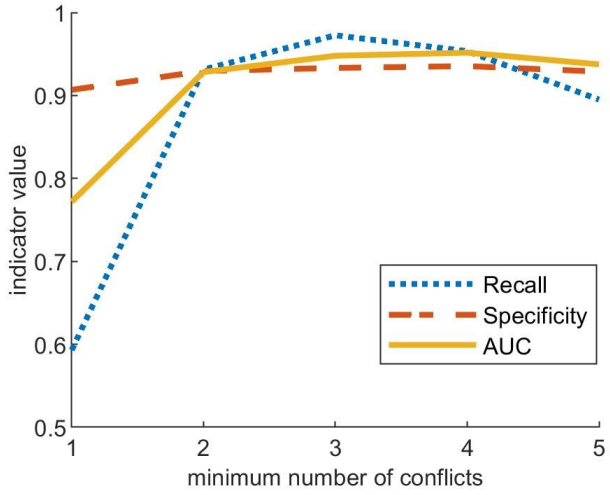

Figure 26. Sensitivity analysis of the effect of parameter n on performance indicators (full-test dataset).

To sum up, this analysis allowed us to determine a TTC threshold (u = 0.78 s) to identify rear-end conflicts. Subsequently, a sensitivity analysis on model performance, and some considerations on the effect of n on the size of the "unsafe" class of observations, lead to the selection of n = 3. In other words, in the RTConfPM, a 5-minute interval was considered unsafe if at least three TTC values under 0.78 seconds were recorded during it.

As a result, 2,241 unsafe situations were detected and matched with the traffic variables recorded on the same cross-section in the previous 5-minute interval.

3.3.3 Training and test datasets

From the procedure of Sections 3.3.1 and 3.3.2, a huge dataset with almost 3.5 million records was built, one for each of the 5-minute intervals and each cross-section of the year. The total number could potentially have been 4,204,800 records (288 5-minute intervals in 1 day × 365 days × 40 cross-sections), but some of the radars were not able to record all of the data during the year due to technical or operational problems. As mentioned before, only 2,241 of these records were matched to an unsafe situation, producing an extremely unbalanced classification problem (one unsafe record every 1,554 safe records). With reference to Figure 27, this is the full dataset.

This dataset was then split into a full training dataset (80% of the full dataset) and in a full test dataset (20%), keeping the unsafe-to-safe ratio of 1:1,554 in both datasets. The latter was then used exclusively for model evaluation.

On the other hand, the full training dataset was used for variable selection (Section 3.3.4) and then divided into unsafe situation and safe situation datasets, respectively containing only unsafe records and only safe records.

SMOTE was then applied on the unsafe situations dataset to increase the number of unsafe records, whereas a limited number of records was randomly sampled from the safe situations dataset to have a 1:1 ratio between unsafe and safe records. The two sets of records were merged in the SMOTEd dataset. The combination of SMOTE and under-sampling is commonly adopted (Chawla et al., 2002). The choice of random under-sampling was suggested by recent literature praising its performance over that of the matched case-control method (Peng et al., 2020).

This dataset was again split into a SMOTEd Training dataset (80% of SMOTEd dataset) and a SMOTEd test dataset (20%), with the same 1:1 unsafe-to-safe ratio in both datasets. The former was used to train the SVM (see Section 3.3.5.2), the latter for model evaluation.

It is worth underlining that although the original full dataset contained more than 3.5 million records, the SMOTEd Training dataset is much smaller (about 14,000 records), therefore the SVM is able to handle it without the need to adopt advanced algorithmic strategies required for large-scale learning (Menon, 2009).

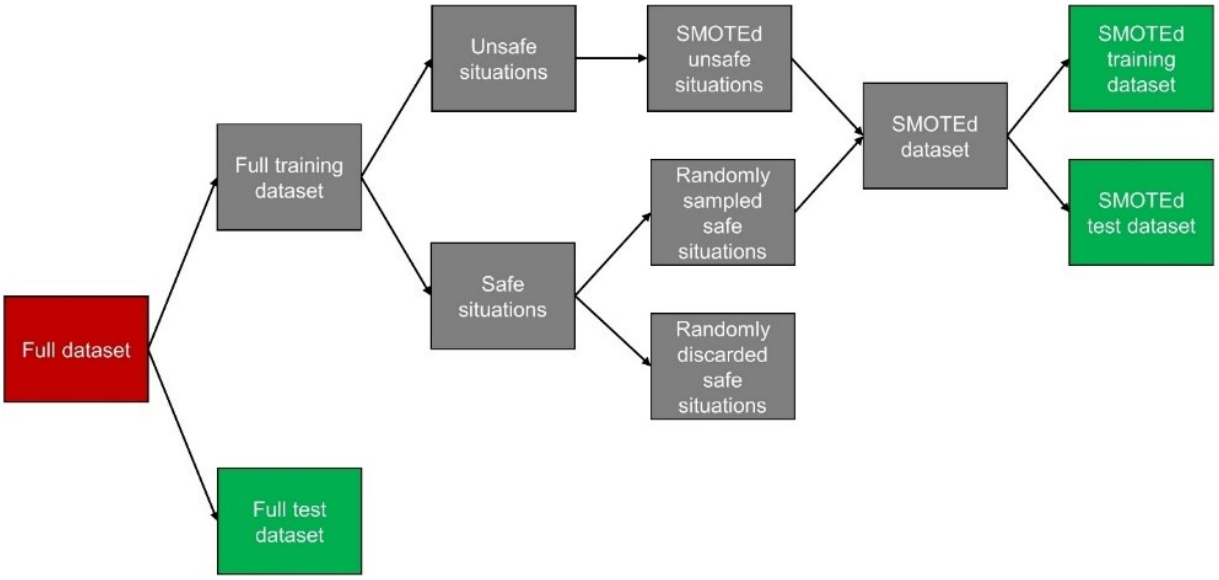

Figure 27. Overview on training, test and full-test dataset formation.

3.3.4 Variable selection

Before applying a classifier, it is good practice to select a subset of variables, excluding those less important and those correlated with other variables. In the case of unbalanced datasets, the variable selection procedure should be performed before applying the resampling procedure, therefore on the full training dataset (see Figure 27). This is especially true when applying SMOTE, which introduces correlation between samples, whereas most variable selection methods assume sample independence (Blagus & Lusa, 2013).

Pearson's correlation was computed for each pair of the 12 recorded variables to test for linear correlation. Results are presented in Figure 28. In RTCPMs/RTConfPMs there is not a clear rule on when to consider two variables as "too much correlated": for example, L. Wang et al. (2019) excluded variables with correlation coefficients higher than 0.3, whereas P. Li et al. (2020) used a 0.5 threshold, Yuan et al. (2019) a 0.65 threshold, and Basso et al. (2018) a 0.95 threshold. Therefore, we decided not to remove any of the variables a priori, and to perform further investigations with random forest.

	Vol_R	Vol_C	Vol_L	HVol_R	HVol_C	HVol_L	Sp_R	Sp_C	Sp_L	SdSp_R	SdSp_C	SdSp_L
Vol_R	1.00	0.59	0.28	0.19	0.26	0.06	-0.17	-0.24	-0.16	-0.28	-0.01	0.12
Vol_C	0.59	1.00	0.40	-0.08	-0.06	0.05	-0.01	-0.13	-0.14	-0.13	-0.26	0.14
Vol_L	0.28	0.40	1.00	0.05	0.04	0.04	-0.09	-0.14	-0.11	-0.12	-0.08	0.03
HVol_R	0.19	-0.08	0.05	1.00	0.57	0.06	-0.39	-0.23	-0.06	-0.56	0.36	-0.01
HVol_C	0.26	-0.06	0.04	0.57	1.00	0.17	-0.36	-0.48	-0.22	-0.44	0.37	0.04
HVol_L	0.06	0.05	0.04	0.06	0.17	1.00	-0.14	-0.19	-0.21	0.05	0.15	0.09
Sp_R	-0.17	-0.01	-0.09	-0.39	-0.36	-0.14	1.00	0.68	0.53	0.34	-0.16	0.02
Sp_C	-0.24	-0.13	-0.14	-0.23	-0.48	-0.19	0.68	1.00	0.74	0.34	-0.08	0.01
Sp_L	-0.16	-0.14	-0.11	-0.06	-0.22	-0.21	0.53	0.74	1.00	0.20	0.03	0.00
SdSp_R	-0.28	-0.13	-0.12	-0.56	-0.44	0.05	0.34	0.34	0.20	1.00	0.02	0.11
SdSp_C	-0.01	-0.26	-0.08	0.36	0.37	0.15	-0.16	-0.08	0.03	0.02	1.00	0.16
SdSp_L	0.12	0.14	0.03	-0.01	0.04	0.09	0.02	0.01	0.00	0.11	0.16	1.00

Figure 28. Correlation matrix for the full training dataset.

The random forest model for variable selection was built using the RandomForest package in R (Liaw & Wiener, 2002). Three variables were randomly sampled at each split and 500 trees were grown. Figure 29 presents variable importance as the total decrease in Gini node impurities.

According to the model, the most important variables are those related to speed. In addition to the information on total traffic volumes, the presence of heavy-duty trucks on the rightmost and central lanes appears to be significant. The least important variable is the percentage of heavy-duty trucks in the leftmost lane; this is not surprising because this percentage is always close to zero, even during congestion periods, as trucks are forbidden to use that lane, according to the Italian Highway Code (Art. 176-9).

This "importance ranking" does not determine, on its own, the optimal set of variables to include in the model; instead, it can be used to define several candidate sets to test and evaluate using performance indicators (see Sections 3.3.5 and 3.3.6).

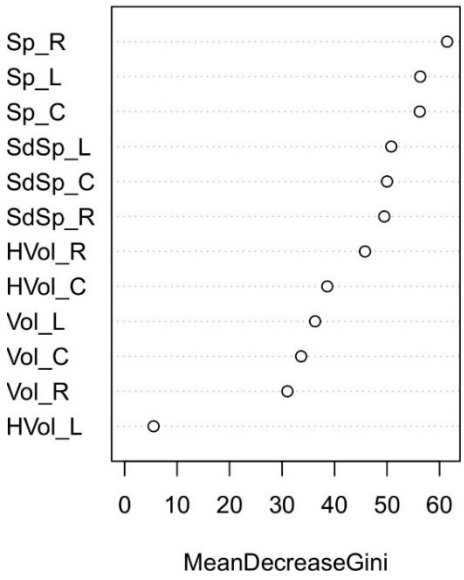

Figure 29. Change in Gini impurity index to determine variable importance in the RTConfPM.

3.3.5 Model development

3.3.5.1 Resampling

A combination of SMOTE over-sampling of unsafe records and random under-sampling of safe records was applied to achieve a balanced dataset. Two parameters should be chosen in this phase:

- The factor k of the SMOTE (i.e., how many new samples are generated for each original sample).
- The desired ratio between unsafe and safe records.

In the literature, a typical value of k is 5, which was originally proposed by Chawla et al. (2002). For this work, we carried out a sensitivity analysis, which showed that a further increase in the value of k did not produce a significant improvement in performance. Another sensitivity analysis was carried out to choose the unsafe/safe ratio. A 1:1 ratio was chosen because higher ratios of safe situations would have produced a significant decrease in the recall indicator. For the sake of brevity, details on this sensitivity analysis are not reported in this dissertation. SMOTE procedure was implemented in MATLAB by the authors.

3.3.5.2 SVM

Having obtained a SMOTEd dataset (see Figure 27), 80% of this dataset was used to train several SVM classifiers, using combinations of different sets of input variables and kernel

functions. SVM models were implemented in MATLAB using 'fitcsvm' function (cost = 1; gamma = 1).

The choice of SVM was made after an extensive comparative study between several candidate classifiers, whose details are reported in Appendix 2.

Different sets of input variables were selected:

- Set #1, with all 12 recorded variables (i.e., without variable selection).
- Set #2, with 11 variables, which excluded the least important according to random forest (i.e., *HVol_L*; see Section 3.3.4).
- Set #3, with the six most important variables according to random forest (i.e., average and standard deviation of speeds; see Section 3.3.4).

Moreover, SVMs with either a linear or radial basis (RBF) kernel function were trained, resulting in six possible combinations.

For each combination of input variable set and kernel function, Monte Carlo cross-validation (Simon, 2007; Q. S. Xu & Liang, 2001) with 100 repetitions was employed. Therefore, for each combination, 100 different 80%-20% partitions of the SMOTEd dataset (see Figure 27) were performed, resulting in the training of 100 different SVMs.

3.3.6 Results

3.3.6.1 Predictive performance

For each of the 100 SVMs of each input variable – kernel function combination, accuracy, recall, specificity, and AUC were computed on SMOTEd training and test datasets and on the full test dataset; in the remainder of the paragraph, we will refer to these three datasets simply as training, test, and full-test datasets, respectively. Averages across the 100 repetitions are reported in Table 20.

Training and test indicators indicate how well the SVM models adapted to the SMOTEd dataset; then, full-test is used to evaluate the models' performance on a proper real-world dataset, with a realistic, strongly unbalanced ratio between crashes and non-crashes.

In particular, it is possible to observe that, for all the models, training and test indicators have similar values, meaning that none of the SVM models suffers from overfitting.

Table 20. Performance indicator values, averaged on the 100 Monte Carlo repetitions, for training, test, and full-test dataset of the RTConfPM

Training dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
	Set #1	0.958	0.982	0.934	0.066	0.952
Linear	Set #2	0.957	0.981	0.934	0.066	0.951
	Set #3	0.788	0.975	0.601	0.399	0.945
	Set #1	0.981	0.992	0.970	0.030	0.995
Radial basis	Set #2	0.981	0.992	0.971	0.029	0.995
	Set #3	0.825	0.998	0.651	0.349	0.995

Test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
	Set #1	0.958	0.982	0.934	0.066	0.953
Linear	Set #2	0.957	0.981	0.933	0.067	0.952
	Set #3	0.788	0.976	0.600	0.400	0.945
	Set #1	0.972	0.986	0.959	0.041	0.985
Radial basis	Set #2	0.972	0.986	0.959	0.041	0.985
	Set #3	0.809	0.979	0.639	0.361	0.981

Full-test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
	Set #1	0.934	0.983	0.933	0.067	0.948
Linear	Set #2	0.933	0.981	0.933	0.067	0.947
	Set #3	0.600	0.978	0.600	0.400	0.942
	Set #1	0.958	0.945	0.958	0.042	0.972
Radial basis	Set #2	0.959	0.947	0.959	0.041	0.973
	Set #3	0.639	0.904	0.639	0.361	0.966

Overall, big differences do not exist between the linear and RBF models, with the latter having marginally better indicators, except for recall in the full-test dataset.

Input variable Set #3 produces significantly worse results, compared to the others, meaning that some variables, whose effects are indeed significant, are not included in that set. In practical terms, this means that despite the speed variables being the most important (see

Section 3.3.4), they cannot, on their own, produce a reliable model, which confirms that the percentage of heavy-duty trucks and total traffic volumes are indeed significant, as intuitively expected.

Set #1 and Set #2 return almost identical results, confirming that *HVol_L* is irrelevant, as expected. By focusing on the full-test dataset indicators, the two best models appear to be Linear-Set#1 and RBF-Set#2. The first has a high recall of 98.3%, the second a quite lower 94.7%; nevertheless, the latter is better in all other indicators, with a 95.9% specificity and accuracy and 0.973 AUC.

In addition to showing great potential from a safety point of view, their operational performance is remarkable. In RBF-Set#1, the false alarm rate is only 4.1%; therefore, a false alarm is triggered only once out of 24.4 five-minute intervals or, in other words, fewer than 12 times per day.

3.3.6.2 Contributing factors

SVM is often regarded as a "black-box" technique because the effects of the explanatory variables cannot be directly assessed. To overcome this issue and investigate the factors contributing to the risk of unsafe situations, we applied the sensitivity analysis developed by Fish and Blodgett (2003) and applied in several other road safety works (Dong et al., 2015; Gu et al., 2020; J. Li et al., 2020; Yu & Abdel-Aty, 2013b). According to this procedure, each variable is changed by a pre-specified unit while keeping the others unchanged; then SVMs are recalibrated with the modified datasets, and comparisons are made on the mean predicted probabilities of unsafe situations occurring. This allows retrieval of positive and negative dependence for each input variable and unsafe situation, as well as their functional relationship, but "direct comparison among [input] variables is precluded" (Fish & Blodgett, 2003).

In this work, we considered 10% of the mean value of each of the 12 input variables as changing unit; SVMs were trained with both linear and radial-basis kernel function. As expected, the kernel function type did not have an effect on the interpretation of contributing factors. Results of this analysis are reported in Table 21, highlighting which relationships exist between input variables and unsafe situations, with the same style adopted by Gu et al. (2020) and Yu and Abdel-Aty (2014).

A positive relationship exists between unsafe situations that may lead to rear-end collisions and traffic flow; this result is expected because higher flows imply more interactions between vehicles, and therefore an increased propensity for rear-end conflicts (Dimitriou et al., 2018). Average speed is negatively correlated with traffic density; therefore, it is not surprising that a higher chance of unsafe situations exists as average speed decreases (L. Wang et al., 2017; Yu & Abdel-Aty, 2013b). Standard deviation of speed has a positive relationship with unsafe situations, with higher differences in speeds among vehicles resulting in an increase in risk. This has been observed in several works (Theofilatos et al.,

2019; L. Wang et al., 2017; Z. Zheng et al., 2010), and it is particularly relevant in the case of rear-end conflicts (Dimitriou et al., 2018). Finally, an increase in the percentage of heavy-duty vehicles is linked to an increase in risk. The presence of such vehicles is correlated to an increase in congestion (especially in the rightmost lane), a decrease of average speed, and an increase of standard deviation of speed in the central lane (see Figure 28), which are all factors that contribute to an increase of rear-end conflicts.

Table 21. Sensitivity analysis results - positive relationship: if the input variable increases, the risk of unsafe situation increases; negative relationship: if the input variable decreases, the risk of unsafe situation increases.

Variable	Relationship
Vol_R	positive
Vol_C	positive
Vol_L	positive
HVol_R	positive
HVol_C	positive
HVol_L	positive
Sp_R	negative
Sp_C	negative
Sp_L	negative
SdSp_R	positive
SdSp_C	positive
SdSp_L	positive

3.4 Case study: real-time crash prediction

A traditional RTCPM was developed, using the same traffic data of the RTCPM described in Section 3.3.1.

3.4.1 Crash data

A 6-year crash database was available for the same motorway. Data were collected by the administrative department of the motorway from 1st January 2011 to 31st December 2016. During this period, 5,362 crashes occurred; 1,575 of them were rear-end collisions.

In the RTCPM developed for this work only rear-end collisions which occurred in 2013 within 2.5 kilometers of distance (either upstream or downstream) from any of the radar-equipped cross-sections were considered. One hundred and twenty-seven crashes satisfied these conditions. The decision to focus exclusively on rear-end crashes was motivated by the need to compare this RTCPM directly to an RTConfPM which can only predict rear-end collisions, as explained in Section 3.3.

The selected crashes were matched with the traffic variables recorded at the closest cross-section in the first complete 5-minute interval before the crash. The decision to refer crashes to the closest detector is common in RTCPM literature; however more sophisticated RTCPMs also considered traffic data from one or more upstream and downstream detectors (Hossain et al., 2019). In the present case study, this choice was forced by the irregular disposition and spacing of detectors on the motorway.

Not all the crashes could be matched with traffic conditions because some sensors were not active at the time of the crash: 39 collision events had to be discarded for this reason, resulting in a dataset of 88 usable crashes.

As regards the training and test datasets, the same procedure depicted in Figure 27 was followed. The full test dataset of RTCPM and RTConfPM are different. In the first case, the full dataset's partition was performed so that a full training dataset and full test dataset had the same ratio of crash and non-crash records; in the second case, they had the same ratio of safe and unsafe records (where unsafe situations were defined in accordance with Section 3.3.2).

3.4.2 Variable selection

The same procedure described in Section 3.3.4 was used for the RTCPM.

Because different records are assigned to the two response classes (safe versus unsafe) in RTCPM and RTConfPM, random forest variable selection could, in principle, and from a statistical point of view only, produce different results. However, as explained in Section 1, if it is assumed that traffic conditions are precursors to conflicts, which in turn are

precursors to crashes, then expecting variables that are "important" to predicting conflicts to be "important" to predicting crashes is also reasonable.

Figure 30 presents variable importance as the total decrease in Gini node impurities and, as the reader can observe when comparing this figure to Figure 29, the order of importance and the relative importance between variables is very similar for the two approaches. The random forest used for this analysis had the same parameters (500 trees and three variables randomly sampled at each split) as that of Section 3.3.4.

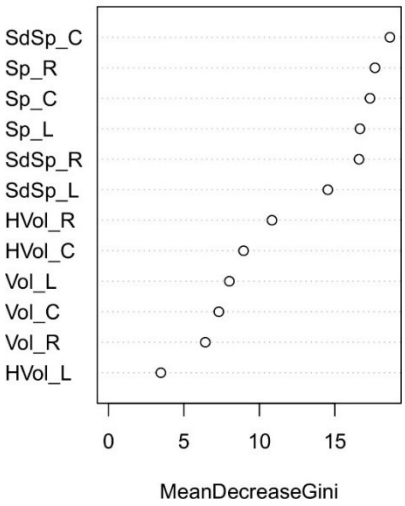

Figure 30. Change in Gini impurity index to determine variable importance in the RTCPM.

3.4.3 Model development

The number of crash records in the RTCPM is much lower than the unsafe records of RTConfPM, leading to an even more unbalanced dataset (crash/non-crash ratio is 1:39,567); therefore, SMOTE is also applied in this case. k = 5 and a 1:1 ratio were chosen for RTConfPM, following a similar sensitivity analysis.

The same six combinations of input variables and kernel functions were analyzed. For each of them, a Monte Carlo cross-validation procedure was applied, the SMOTEd dataset was partitioned 100 times, and 100 independent SVMs were trained for each combination.

3.4.4 Results

Accuracy, recall, specificity and AUC values were averaged over the 100 repetitions for training, test, and full-test datasets (Table 22).

RBF models have significantly higher performance in training than linear models do, but, contrarily to the RTConfPM case (see Section 3.3.6.1), they seem to suffer from overfitting because test and full-test performance is much lower.

The removal of an irrelevant (according to random forest variable selection and intuition) variable (i.e., *HVol_L*) provided an increase, albeit small, of all indicators of the linear models. Considering only the six most important variables according to random forest did not result in good performance: a steep increase in recall was indeed matched by a great decrease in all other indicators, producing unsatisfactory overall accuracy values.

Table 22. Performance indicator values, averaged on the 100 Monte Carlo repetitions, for training, test, and full-test dataset of the RTCPM.

Training dataset

Kernel function	**Input variables**	**Accuracy**	**Recall**	**Specificity**	**False alarm rate**	**AUC**
	Set #1	0.652	0.490	0.814	0.186	0.702
Linear	Set #2	0.662	0.493	0.831	0.169	0.700
	Set #3	0.500	0.809	0.190	0.810	0.604
	Set #1	0.967	0.940	0.995	0.005	1.000
Radial basis	Set #2	0.963	0.944	0.982	0.018	0.998
	Set #3	0.806	0.985	0.627	0.373	0.997

Test dataset

Kernel function	**Input variables**	**Accuracy**	**Recall**	**Specificity**	**False alarm rate**	**AUC**
	Set #1	0.628	0.465	0.791	0.209	0.656
Linear	Set #2	0.636	0.469	0.806	0.194	0.651
	Set #3	0.494	0.800	0.183	0.817	0.580
	Set #1	0.856	0.818	0.894	0.106	0.942
Radial basis	Set #2	0.831	0.774	0.887	0.113	0.931
	Set #3	0.676	0.802	0.594	0.406	0.850

Full-test dataset

Kernel function	**Input variables**	**Accuracy**	**Recall**	**Specificity**	**False alarm rate**	**AUC**
	Set #1	0.788	0.395	0.788	0.212	0.612
Linear	Set #2	0.795	0.529	0.795	0.205	0.682
	Set #3	0.175	0.824	0.175	0.825	0.583
	Set #1	0.899	0.219	0.899	0.101	0.615
Radial basis	Set #2	0.875	0.177	0.875	0.125	0.569
	Set #3	0.485	0.353	0.485	0.515	0.537

Overall, among the six combinations, the best seems to be the linear model which used input variable Set #2. Considering the full-test dataset, it has lower specificity than the RBF-Set#2 model, but the latter greatly struggles with recall. With a 0.68 AUC, 52.9% recall, and 79.5% specificity (corresponding to a 20.5% false alarm rate), this model is not the best in the literature, but it is in line with others (Theofilatos et al., 2019; Yuan et al., 2019). Moreover, it must be taken into account that it was calibrated starting from a full dataset that only included 88 crashes. For comparison, in Hossain et al.'s (2019) systematic review, 30 studies out of 78 had a crash sample higher than 500.

Moreover, as underlined by Basso et al. (2018), many RTCPMs in the literature are only evaluated on artificially balanced datasets, instead of a real-world dataset, and that could provide overoptimistic results for models that are actually not very good. In addition, the results presented in this dissertation are robust because they were obtained with Monte Carlo cross-validation, and therefore not influenced by particularly favorable dataset partitions, which again may produce overoptimistic results (Santos et al., 2018).

Regarding the analysis of the contributing factors on the risk of crash occurrence, the same procedure described in Section 3.3.6.2 was applied. As expected, the same positive and negative relationships were found (see Table 21).

3.5 Case study: discussion and conclusion

3.5.1 Comparison between RTConfPM and RTCPM

The previous sections showed how the RTConfPM is able to predict unsafe situations with significantly more accuracy than the RTCPM can. In the two approaches, unsafe situations are defined in different terms; nevertheless, these are very important results because conflicts can be considered precursors to crashes, thus predicting them and intervening to prevent them from happening would in turn prevent crashes from happening.

The scenario analyzed in this work allows us to perform a further direct comparison that is of great practical interest and answers the following question: how many real-world observed crashes would the RTConfPM predict as unsafe situations?

To answer this question, the same full-test dataset used for RTCPM validation was considered. Records in that dataset are divided into two unbalanced classes: crashes and non-crashes. Instead of making predictions on that dataset using the RTCPM as in Section 3.4.4, the RTConfPM is now applied to determine whether real crashes would be predicted as "unsafe situations" with the conflict-based approach.

Linear-Set#1 was applied (with 100 Monte Carlo repetitions) and produced the following averaged indicators:

- 93.3% accuracy
- 64.7% recall
- 93.3% specificity
- 0.775 AUC

These indicators should be compared with those of the best RTCPM model (Linear-Set#2; see Table 22), and they show that the RTConfPM is able to predict more crashes correctly than the RTCPM can (+11.8% in recall indicator), and do it with a lower false alarm rate (+13.8% in specificity). Overall, the model also has better accuracy (+13.8%) and AUC (+0.093).

This leads to several meaningful considerations. First, a model trained exclusively with conflicts was able to predict more crashes correctly than a model trained with crashes was, confirming what was suggested in theoretical terms at the beginning of Section 3.

Second, this increase in recall is accompanied by an even bigger decrease in the false alarm rate, meaning the conflict-based approach can provide not only safety but also operational benefits.

Finally, there is still a fairly high percentage of crashes not predicted by the RTConfPM: not all crashes are caused by traffic conditions and therefore the model cannot predict all of them. This can be partially improved in the future by using additional input information

(e.g., weather conditions), but a share of crashes will always be primarily caused by unprovoked human error, which would be very challenging to predict from an exclusively cross-sectional point of view.

3.5.2 Advantages and disadvantages of RTConfPMs

The results of this work allow as to identify some potential general advantages and disadvantages of this real-time conflict-based approach, which should be confirmed by carrying out further studies in different scenarios.

As regards the advantages, an RTConfPM can be used in all those situations where crash data are unavailable or do not have the required precision in terms of spatial and/or temporal precision.

In addition, conflicts are more frequent than crashes, and this has several positive consequences: data collection period may be shorter; it is possible to train a RTConfPM using more recent data; given the same data collection time, more data are available for training, which may result in a more reliable model.

Finally, many crashes are caused by factors (e.g., driver distraction) which cannot be used as input in a traditional crash-based approach. Using those crashes to train a RTCPM may reduce predictive performance, because an unsafe situation may be incorrectly linked to traffic/weather conditions which are not responsible for that situation. On the other hand, considering unsafe a time interval in which several conflicts take place (as described in Section 3.3.2) reduces the number of unsafe situations which are in truth unrelated to traffic/weather conditions, and this produces better predictive power.

There are however a number of cases in which adopting a conflict-based approach is not feasible or not ideal, and therefore a traditional RTCPM remains preferable.

First, it is not always possible to have access to disaggregate vehicle-by-vehicle data and therefore to identify traffic conflicts, although there can be some workaround, as shown by Katrakazas et al. (2018), who were able to derive conflicts from aggregate traffic data using a simulation-based approach.

Moreover, contrary to RTCPMs, which are able to make predictions on specific types of accidents, but also regardless of the type, in RTConfPMs the prediction is always on a specific type of crash, because there is not a unique surrogate measure of safety able to provide information on all crash types. For example, the TTC computed as in (26) is a good surrogate for rear-end crashes, but it is useless to investigate side-impact crashes.

Finally, it may be complicate or expensive to obtain reliable values of certain surrogate measures of safety. It is not the case of TTC in motorways, which may be easily computed with data from radars or loop detectors, but it may be in other situations, for example

intersections, where automatic video processing (Saunier & Sayed, 2007) or LiDAR (Tarko et al., 2016; Ural et al., 2015) are needed. However recent developments in sensing technologies and ITS have reduced the costs and improved the reliability of these data, and it is expected that this trend will continue in the future (Tarko, 2019).

3.5.3 Summary and final remarks

To sum up, this work introduced a model to predict conflict-prone situations in real-time. Although the concept is not entirely new (Caleffi et al., 2017; Formosa et al., 2020; Katrakazas et al., 2018), this work represents the first attempt at developing an RTConfPM based on real-world observed surrogate measures of safety and comparing it to a traditional RTCPM.

Both models were developed using a similar and robust procedure. The results were promising, with the best RTConfPM achieving accuracy, recall, and specificity values above 93%, significantly outperforming the crash-based approach.

This work has some limitations that future studies will address.

The RTConfPM was compared with an RTCPM that had below-average performance due to a limited number of observed rear-end collisions and the detectors' layout. Comparing the two approaches in case studies in which the RTCPM provides better performance would be interesting.

This work can be considered a starting point for several research developments that can overcome the limitations of this work and extract the full potential of RTConfPM:

- Using alternative definitions of unsafe situations, possibly considering different levels of risk.
- Collecting traffic and surrogate safety data using different technologies (e.g., V2X).
- Applying the model to other types of infrastructures (e.g., intersections).
- Including in RTConfPM other types of input (e.g., weather conditions).
- Estimating how short the observation period for an RTConfPM can be while still providing good performance.
- Studying transferability and seasonality issues.
- Testing different time intervals for data aggregation and/or different prediction horizons.
- Analyzing real-time intervention strategies to prevent conflicts from happening.

4. Conclusion

This dissertation presented the application of conflict-based approaches for both long-term and real-time road safety analysis.

With respect to the main research questions declared in the Introduction, it was possible to conclude that:

- Extreme value theory is able to model the probabilistic relationship between conflicts and crashes in a satisfactory way, providing reliable long-term prediction of crash rates.
- It is possible to predict crashes in real time using a conflict-based approach; in addition, it was shown that the conflict-based approach has the potential to provide more reliable predictions than the traditional crash-based approach.

4.1 Main findings

Regarding the application of EVT for long-term road safety analysis, this dissertation contributed significantly to the literature and several observations can be made in light of the results presented in this work.

First, existing literature on EVT applications to road safety reached contradictory conclusions regarding which is the better approach between the block maxima and the peak-over-threshold (see Section 2.1). In the proof-of-concept scenarios presented in Sections 2.3 and 2.4, it was possible to highlight how the two approaches (and their bivariate extensions) can, in fact, provide predictions with comparable precision. This was also confirmed by the large-scale real-world case study in Section 2.5. Therefore, in general, neither approach can clear-cut outperform the other; it is however true, as observed in Section 2.5, that the peak-over-threshold approach is able to extract information by a significantly higher amount of data, and therefore it may be preferable to use it when the observation period is relatively short.

Second, the proof-of-concept on the bivariate extreme value theory (Section 2.4) showed that combining information from two surrogate measures of safety datasets improved the performance of the model. This was previously observed in other works; however, here we highlighted the fact that this improvement is significant only as long as the two datasets are somewhat correlated: if the two are uncorrelated, then developing two independent univariate models (instead of a single bivariate one) would produce similar results.

Finally, in Section 2.5 we showed how large-scale EVT applications (in fact, the biggest in terms of observation period that can be found in the literature) can provide very reliable estimates of annual crash rates, suggesting that uncertain/suboptimal results in previous works were probably caused by insufficient length in the observation period. It was also proved that the EVT models estimated in the case study are transferable to other locations with similar characteristics, while transferability in time can be more delicate, as the seasonality effect may play a role (see Appendix 1).

Regarding the conflict-based approach to real-time road safety, this dissertation represents one of the first works in the literature to attempt to train a real-time road safety model with the use of traffic conflict data, in place of crash data.

In particular, it significantly contributes to the literature by highlighting the importance of first defining what a conflict is, rather than just using fixed thresholds from the literature (Formosa et al., 2020; Katrakazas et al., 2018) or qualitatively identifying conflicts (Caleffi et al., 2017). Here, an objective and theoretically robust method was proposed, which uses the results from an EVT model, estimated for the specific case study, to retrieve the threshold used to separate normal traffic interactions from proper conflicts.

In addition, it is the first work that explicitly and directly compares the performance of a conflict-based model with a crash-based one, showing that the former can even outperform the latter, especially in case studies such as the one analyzed, where the amount of crash data available for training is limited.

4.2 Future research and practical implications

This dissertation presented a methodology for evaluating road safety without the use of crash data; such approach provides several positive practical implications.

First, it is possible to apply both long-term and real-time prediction models to all those scenarios in which crash data are unavailable or unreliable; this can be the case of rural roads, third world countries or newly-built infrastructures.

Furthermore, since the frequency of traffic conflicts is much higher than that of crashes (in the case study of Chapter 3 more than 100,000 conflicts were recorded in one year, versus "only" 88 crashes; see Sections 3.3.2 and 3.4.1), it is possible to retrieve information with shorter observation times or at fewer locations. In particular, by reducing the time needed to collect data for calibration, it is possible to obtain models much more resilient,

as they can quickly "adapt" themselves to the potentially changing characteristics of traffic and infrastructure.

As a further consequence, contrary to the crash-based approach, which is reactive, since it needs a certain number of crashes to happen before being able to provide safety estimations, the proposed approach is proactive, allowing one to intervene more quickly and avoid more crashes from happening.

There are some methodological aspects that can be investigated in the future.

Regarding the application of EVT to road safety, in light of the potential shown by the bivariate approach (see Section 2.4), it could be interesting to investigate an extension to multivariate models. This has recently been addressed by Fu et al. (2020), but further research is needed in order to have more insights on the benefits of that approach. For real-world applications, it could be interesting to test the method with different types of data source (e.g., V2X communications) or different types of conflicts analyzed.

These last application-oriented aspects are relevant also for real-time conflict prediction models. In addition, it could be interesting to extend the pool of candidate input variables, including traffic characteristics collected at upstream/downstream cross-sections or weather characteristics, as they could provide further improvements in terms of model performance, and to test the method also in different infrastructures, such as intersections. Also, it would be crucial for practical implementations to quantify the effect of the duration of the observation period length and of seasonality, and to investigate the transferability to highway segments with different characteristics.

Finally, the long-term methodology and real-time methodology proposed in this work could also be tested in scenarios in which vulnerable road users are involved.

Future research will also need to further investigate practical applications to be put in place in order to avoid crashes from happening. The present dissertation focused on methodologies whose objective is to *predict* crashes, but it will be fundamental to then study approaches to *prevent* crashes, given the predictions.

For long-term road safety, EVT can be used to perform before-after analyses. There are some early examples of this in the literature, with the works of L. Zheng, Sayed, et al. (2018) and L. Zheng & Sayed (2019b), respectively applying block maxima and peak-over-threshold approaches to evaluate the safety effects of left-turn bay extensions at several intersections. It would be very relevant to test these models to quantify the safety of innovative infrastructure interventions (e.g., shared spaces), without waiting to collect enough crash data. A further step would be to perform these before-after analyses *before* the physical implementation of the infrastructural solution; this could be carried out with the use of driving simulators or micro-simulation tools (after appropriate validation), and

it would provide relevant benefits not only in terms of safety, but also in terms of costs, as it could prevent expensive infrastructural solutions which could be potentially ineffective. Furthermore, this could facilitate the testing of innovative and creative solutions.

For real-time road safety, some intervention strategies have been investigated in the past: variable message signs and ramp metering (Abdel-Aty et al., 2007; Pande & Abdel-Aty, 2005); yet, both approaches have several downsides.

The effectiveness of variable message signs is strongly dependent on user compliance, which may not be high enough, especially if drivers are repeatedly given instructions during their trip; in addition, such signs must be positioned at short and regular space intervals, and this could be rather expensive to implement. The ramp metering strategy could provide safety benefits, but at the expense of operational performance (e.g., creating congestions at highway entry points); in addition, the ramp metering system exists only in a limited number of countries, and it could be unrealistic to envision large-scale worldwide adoption.

However, advances in telecommunications and automated driving technologies could provide viable and more effective solutions, which deserve to be studied. One possibility is to provide in-vehicle messages to drivers, thanks to V2I communications, but the effectiveness of the intervention would still depend on user compliance. A more advanced solution would be to interact in real-time with automated or semi-automated vehicles, modifying their kinematic variables, and influencing the overall traffic variables of the stream (composed of both automated and manually-driven vehicles). In the literature, there are some early studies investigating the possibility of interacting with advanced vehicle cruise control systems in real time, to optimize the capacity of the infrastructure (Manolis et al., 2020). A similar approach could be followed with the aim of minimizing the risk of crashes.

Appendix 1. Transferability and seasonality in EVT applications to road safety[6]

This Appendix reports an analysis of the transferability and of seasonality effect of an EVT application in real-world.

A1.1 Data collection

The analyses reported in this Appendix were carried out on data from 14 of the cross-sections analyzed in Section 2.5. These cross-sections showed similar traffic characteristics across the sections: Annual Average Daily Traffic (AADT) was computed for each of them (AASHTO, 2009), and each of them had an AADT ranging between 34,370 and 38,642 (see Table A-1).

[6] The work presented in this Appendix is part of the following publication:

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2020). Transferability and seasonality in extreme value theory applications to road safety: a case study in an Italian motorway. *Advances in transportation studies*, (SI 2), 33-46. (Orsini et al., 2020b)

Table A-1. Annual Average Daily Traffic for each motorway section

Section ID	AADT
1	38,642
2	37,879
3	37,489
4	37,965
5	37,404
6	36,670
7	38,523
8	37,744
9	35,904
10	35,828
11	34,370
12	34,460
13	38,230
14	35,996

A1.2 Analysis and results

Traffic data were used to estimate GP distributions with the POT approach. In Section 2.5.2. the TTC values derived from these traffic data were aggregated for each section and used to assess the rear-end collision risk for each of them, by estimating separate EVT models for each section.

In this case the main aim was to evaluate the transferability of this EVT approach across similar sections, i.e. if it is possible to assess the collision risk of a road section using an EVT model calibrated using TTC values collected in other road sections.

Moreover, it was interesting to investigate the seasonality effect, since traffic flows and weather conditions significantly change during the year, as well as the observed number of rear-end collisions.

In this Appendix two different sets of models were estimated:

1. Set #1. TTC values were aggregated across all sections, for each month. A total of 12 GP distributions were fitted, one for each month.
2. Set #2. For each section and each month, only TTC collected in all the other sections in the same month were considered. A total of 168 GP distributions were fitted (12 months * 14 road sections).

Each GP distribution of the two set of models described above was fitted with the procedure described in Section 2.5.2.2.

In order to apply the POT approach, a threshold has to be selected. As mentioned in Section 2.2.1.2, the traditional method to choose the threshold is by performing graphical diagnostics, which is however practically unfeasible when the number of models to estimate is high, as in this case. For this reason the ATSM developed by Thompson et al. (2009) was adopted.

In the ATSM, the analyst must choose a set of candidate thresholds. In this case the set was the largest possible, ranging from the lowest value of negated-TTC (i.e., all datasets values are extremes) and the 30th highest value of negated-TTC. The latter condition was adopted because both literature (Farah & Azevedo, 2017) and analysts' experience suggest that model estimations with less than 30 values are usually unreliable.

The ATSM returns a range of suitable threshold. The largest value of this range was chosen, in analogy to the graphical diagnostics method, where the highest possible threshold is usually chosen.

For the sake of brevity, Table A-2 only presents the thresholds chosen for the GP distributions of Set #1.

Table A-2. Thresholds, associated TTC values, and GP parameters estimates for Set#1 models.

Month	Threshold (TTC value)	Shape (std. error)	Scale (std. error)
January	-0.94 (0.94)	-0.2770 (0.0133)	0.2834 (0.0063)
February	-0.48 (0.48)	-0.2114 (0.0497)	0.1457 (0.0110)
March	-0.82 (0.82)	-0.2733 (0.0145)	0.2451 (0.0061)
April	-0.48 (0.48)	-0.2302 (0.0406)	0.1418 (0.0091)
May	-0.46 (0.46)	-0.3542 (0.0807)	0.1946 (0.0228)
June	-0.44 (0.44)	-0.3147 (0.0976)	0.1823 (0.0248)
July	-0.43 (0.43)	-0.2057 (0.0638)	0.1189 (0.0113)
August	-0.43 (0.43)	-0.2839 (0.0497)	0.1494 (0.0109)
September	-0.50 (0.50)	-0.1245 (0.0519)	0.1204 (0.0091)
October	-0.47 (0.47)	-0.2093 (0.0497)	0.1325 (0.0098)
November	-0.40 (0.40)	-0.2214 (0.0794)	0.1332 (0.0139)
December	-0.43 (0.43)	-0.1522 (0.0792)	0.1064 (0.0131)

After selecting the threshold, the final estimation of GP parameters is performed with Maximum Likelihood (ML) method. Table A-2 presents GP parameters for the

distributions of Set #1. Again, parameter estimates of the models of Set#2 are not reported.

Once the threshold is selected and the parameters estimated, the GP cumulative distribution function G(-TTC) can be defined (see equation 2). A collision between two vehicles occurs when -TTC=0, therefore the collision probability, conditional to the two vehicles being in a conflict situation (negated-TTC higher than the threshold), can be defined as in equation 31.

Given observation time *t*, the number of predicted collisions in that time period is obtained by multiplying the collision probability $P_{inc,t}$ (produced estimating the GP model using only TTC values collected during *t*), by the number of exceedances $N_{exc,t}$, i.e. the number of observed situations in which the negated-TTC is higher than the threshold, during *t*, resulting in equation 32.

Here, specific considerations have to be made for models of Set #1 and Set #2.

In Set #1, given a month *t*, there is only one fitted GP distributions and therefore only one value of collision probability $P_{inc,t}$. When this probability is multiplied by $N_{exc,t}$, (i.e., the total number of exceedances observed in all road sections), the result is the predicted total number of collisions in the 14 road sections. In addition to this, to any section *s* corresponds a value of $N_{exc,t,s}$, (i.e., the number of exceedances observed in section *s*), and therefore it is possible to predict the number of collisions for each section as:

$$Inc_{t,s} = P_{inc,t} * N_{exc,t,s} \qquad (A.1)$$

In Set #2, there is one different GP distribution for each month *t* and section *s*, which results in a different collision probability for each month and each section. We call this probability $P_{inc,t,\sim s}$ to emphasize the fact that this probability has been obtained from a GP distribution fitted on TTC values collected during month *t* in all sections but *s*. Again, to any section *s* corresponds a value of $N_{exc,t,s}$. The predicted number of collisions for month *t* and section *s* is thus defined as:

$$Inc_{t,s} = P_{inc,t,\sim s} * N_{exc,t,s} \qquad (A.2)$$

Table A-3 contains all predicted number of collisions obtained from models of Set#1, and Table A-4 for Set#2.

Predicted annual number of collisions for each section (Inc_s), and predicted total number of collisions for each month (Inc_t) can be defined as follows:

$$Inc_s = \sum_{t=1}^{N_t} Inc_{t,s} \qquad (A.3)$$

$$Inc_t = \sum_{s=1}^{N_s} Inc_{t,s} \qquad (A.4)$$

Where N_t is the total number of months and N_s the total number of sections.

Table A-3. Predicted number of collisions for each month and each section, considering Set#1 models

Section ID	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Tot. section
1	0.03	0.05	0.03	0.05	0.06	0.01	0.02	0.07	0.10	0.07	0.11	0.01	0.60
2	0.04	0.07	0.03	0.06	0.08	0.04	0.04	0.12	0.13	0.07	0.14	0.01	0.82
3	0.02	0.05	0.02	0.05	0.02	0.00	0.02	0.03	0.04	0.03	0.08	0.01	0.36
4	0.02	0.07	0.02	0.02	0.03	0.00	0.01	0.05	0.08	0.03	0.09	0.01	0.44
5	0.02	0.07	0.02	0.03	0.05	0.03	0.01	0.05	0.04	0.03	0.04	0.00	0.38
6	0.02	0.10	0.01	0.03	0.06	0.00	0.02	0.03	0.07	0.03	0.15	0.04	0.55
7	0.03	0.11	0.03	0.06	0.07	0.02	0.04	0.07	0.08	0.04	0.19	0.02	0.74
8	0.04	0.09	0.03	0.05	0.06	0.03	0.02	0.09	0.09	0.04	0.15	0.02	0.69
9	0.03	0.09	0.02	0.04	0.06	0.02	0.01	0.06	0.06	0.03	0.09	0.02	0.53
10	0.03	0.11	0.03	0.05	0.04	0.01	0.02	0.06	0.10	0.04	0.11	0.02	0.61
11	0.02	0.06	0.02	0.03	0.04	0.01	0.02	0.05	0.03	0.03	0.09	0.03	0.41
12	0.02	0.08	0.01	0.03	0.05	0.01	0.01	0.05	0.05	0.02	0.07	0.01	0.39
13	0.01	0.03	0.01	0.01	0.01	0.00	0.01	0.02	0.03	0.01	0.06	0.01	0.21
14	0.03	0.08	0.03	0.05	0.10	0.03	0.02	0.06	0.08	0.05	0.12	0.01	0.64
Tot month	0.34	1.06	0.28	0.55	0.72	0.22	0.25	0.81	0.95	0.50	1.49	0.21	7.37

Table A-4. Predicted number of collisions for each month and each section, considering Set#2 models

Section ID	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Tot. section
1	0.05	0.06	0.02	0.06	0.04	0.01	0.03	0.04	0.13	0.14	0.13	0.01	0.70
2	0.01	0.08	0.00	0.07	0.09	0.04	0.00	0.07	0.13	0.06	0.17	0.01	0.73
3	0.01	0.04	0.00	0.05	0.03	0.00	0.00	0.05	0.02	0.04	0.09	0.01	0.32
4	0.01	0.08	0.02	0.03	0.03	0.01	0.02	0.03	0.09	0.01	0.02	0.01	0.36
5	0.03	0.08	0.01	0.03	0.05	0.03	0.05	0.07	0.03	0.03	0.06	0.00	0.46
6	0.02	0.06	0.00	0.01	0.07	0.01	0.02	0.03	0.03	0.02	0.02	0.05	0.33
7	0.07	0.11	0.00	0.04	0.07	0.02	0.04	0.06	0.08	0.02	0.21	0.01	0.72
8	0.10	0.10	0.01	0.02	0.11	0.03	0.00	0.08	0.07	0.01	0.06	0.00	0.59
9	0.02	0.10	0.01	0.05	0.05	0.03	0.01	0.04	0.03	0.01	0.06	0.02	0.40
10	0.01	0.16	0.03	0.12	0.05	0.01	0.02	0.09	0.11	0.10	0.07	0.12	0.87
11	0.01	0.08	0.02	0.03	0.03	0.03	0.01	0.01	0.02	0.01	0.19	0.03	0.46
12	0.02	0.09	0.00	0.03	0.05	0.02	0.01	0.02	0.05	0.02	0.06	0.01	0.35
13	0.01	0.03	0.00	0.01	0.02	0.01	0.01	0.03	0.02	0.01	0.07	0.01	0.22
14	0.06	0.03	0.03	0.05	0.14	0.00	0.01	0.07	0.06	0.06	0.15	0.01	0.67
Tot month	0.43	1.08	0.16	0.58	0.80	0.24	0.22	0.67	0.86	0.52	1.33	0.31	7.19

Table A-5. For each road section: predicted annual number of collisions and related 95% confidence intervals for Set#1, Set#2 and the benchmark model; observed average, minimum and maximum annual number of collisions

	Set#1		Set#2		Benchmark		Observed	
Section ID	Pred.	95% C.I.	Pred.	95% C.I.	Pred.	95% C.I.	Mean	[Min;Max]
1	0.60	[0.01;2.06]	0.70	[0.03;2.35]	0.68	[0.00;2.67]	0.50	[0;2]
2	0.82	[0.02;2.86]	0.73	[0.04;2.55]	1.32	[0.06;3.82]	0.50	[0;1]
3	0.36	[0.01;1.25]	0.32	[0.01;1.07]	0.66	[0.02;2.29]	0.00	[0;0]
4	0.44	[0.01;1.46]	0.36	[0.01;1.27]	1.84	[0.01;4.88]	0.33	[0;1]
5	0.38	[0.01;1.32]	0.46	[0.03;1.39]	0.18	[0.00;1.11]	0.50	[0;1]
6	0.55	[0.01;1.86]	0.33	[0.00;1.15]	0.25	[0.00;1.44]	0.50	[0;2]
7	0.74	[0.01;2.52]	0.72	[0.04;2.47]	0.43	[0.07;1.23]	0.67	[0;1]
8	0.69	[0.01;2.38]	0.59	[0.03;2.10]	0.36	[0.01;1.36]	0.33	[0;1]
9	0.53	[0.01;1.81]	0.40	[0.01;1.45]	1.60	[0.15;4.21]	0.17	[0;1]
10	0.61	[0.02;2.05]	0.87	[0.08;2.67]	0.37	[0.00;1.78]	0.83	[0;2]
11	0.41	[0.01;1.44]	0.46	[0.05;1.34]	0.18	[0.01;0.65]	0.33	[0;1]
12	0.39	[0.01;1.31]	0.35	[0.02;1.16]	0.06	[0.00;0.40]	0.33	[0;2]
13	0.21	[0.00;0.73]	0.22	[0.00;0.74]	0.42	[0.00;2.06]	0.50	[0;1]
14	0.64	[0.01;2.19]	0.67	[0.00;2.34]	0.51	[0.01;1.88]	0.83	[0;1]

Table A-6. For each month: predicted total number of collisions and related 95% confidence intervals for Set#1 and Set#2; observed average, minimum and maximum annual number of collisions.

	Set#1		Set#2		Observed	
Month	Pred.	95% C.I.	Pred.	95% C.I.	Mean	[Min;Max]
January	0.34	[0.01;1.28]	0.43	[0.03;1.61]	0.33	[0;1]
February	1.06	[0.06;2.96]	1.08	[0.05;2.91]	0.50	[0;1]
March	0.28	[0.01;1.12]	0.16	[0.00;1.12]	0.17	[0;1]
April	0.55	[0.01;1.99]	0.58	[0.03;2.14]	1.17	[0;2]
May	0.72	[0.00;2.41]	0.80	[0.01;2.78]	0.83	[0;2]
June	0.22	[0.00;1.18]	0.24	[0.00;0.75]	0.50	[0;1]
July	0.25	[0.00;1.39]	0.22	[0.01;1.09]	0.33	[0;1]
August	0.81	[0.00;2.79]	0.67	[0.05;2.15]	0.17	[0;1]
September	0.95	[0.06;2.76]	0.86	[0.06;2.61]	1.00	[0;2]
October	0.50	[0.00;2.02]	0.52	[0.01;1.99]	0.17	[0;1]
November	1.49	[0.01;4.18]	1.33	[0.08;3.63]	0.33	[0;1]
December	0.21	[0.00;1.15]	0.31	[0.03;1.28]	0.83	[0;2]

A1.3 Discussion

It is now interesting to compare the predicted number of collisions with the observed number of collisions, to assess model transferability and the effect of seasonality.

Transferability

Model transferability refers to situations in which a model is transferred and applied to a different context from where it was originally estimated. Performing this operation allows to reduce efforts in model development and the need for large data collection in the application context (Rossi, Meneguzzer, et al., 2013), which is of high interest in road safety.

In this Appendix we first evaluate the performance of the models when aggregating input data from multiple road sections with similar characteristics (model Set#1). Then, we assess if it is possible to transfer such models to other sections with similar characteristics (model Set#2).

Models are evaluated by comparing collision predictions (and respective confidence intervals) and observed number of collisions located within 250 meters around the road section. This distance of 250 meters was chosen consistently with the analysis Section 2.5, where multiple distances were evaluated. Benchmark collision predictions obtained with

models estimated for each section with data collected only in that section (see Table A-5) were used. These benchmark predictions were computed in Section 2.5. Note that in that case the estimation dataset included, for each section, data collected during the whole year.

Prediction performance of benchmark models can be considered good, when compared to other EVT applications to road safety (Songchitruksa & Tarko, 2006; L. Zheng et al., 2014a; L. Zheng, Ismail, et al., 2018). Set#1 is able to produce even better predictions, in terms of mean error (ME), mean absolute error (MAE) and root mean square error (RMSE) (see Table A-7). This improvement was made possible by estimating different GP distributions for each month, instead of using data from the whole year. Indeed, as can be observed in Table A-5, and as it is further explored in the next paragraph, there is seasonality effect to consider, and estimating a model using data coming from the whole year, regardless of the season, appears to be less effective.

Set#2 performance is in line with Set#1, and this is an interesting result, because it indicates that it is possible to transfer an EVT model estimated using TTC values collected in a set of sections to another section with similar characteristics.

All three models are able to include 13 out 14 observed collision values within the predicted confidence interval. The only section that cannot be predicted correctly is road section #3, where no collision was recorded during the 6-year period (see Table A-6). It can be noted, however, that the value 0 is only marginally outside the confidence intervals of the three models.

Table A-7. Performance comparison between benchmark models, Set#1 and Set#2.

Model	ME	MAE	RMSE	Obs. within CI
Benchmark	-0.180	0.480	0.661	13/14
Set#1	-0.074	0.192	0.226	13/14
Set#2	-0.061	0.154	0.184	13/14

Seasonality

Many studies in the past have shown a relationship between road safety and traffic/weather conditions (Theofilatos & Yannis, 2014; Chao Wang et al., 2013; Y. T. Wang & Liu, 2017). If characteristics of traffic and weather cannot be assumed constant throughout the year, then seasonality effect must be taken into account.

The analysis in this paragraph focuses on results from models of Set#1; however, since results, in terms of collision predictions, are similar between Set#1 and Set#2 (see Table A-6), the main findings of this analysis are valid also for Set#2.

Table A-6 shows that both predicted and observed number of collisions change significantly during the year. Figure A-1 illustrates the trends of predicted and observed collision during the year. Three main peaks emerge from the observed incident trend: April, September and December. Set#1 is able to model very precisely September's peak, it completely includes the interval of observed number of collisions within its predicted collisions confidence interval in April, and, even though it is less performing in December, still includes the observed number of collisions within the predicted confidence interval. In the other months observed and predicted number of collisions are distributed consistently, except for November, where Set#1 predicts a peak that is not observed. It has to be noted that the number of collisions observed for 6 years near the 14 road sections analyzed in this work is relatively small (38 collisions in total), therefore the considerations made in this paragraph about the observed collision seasonal trend must be weighted accordingly.

The proposed analysis highlights the relevance of accounting for seasonality effect when extending predictions based on data collected during a limited span of time. Road safety EVT literature (e.g., L. Zheng et al., 2014a) usually defines the predicted number of accidents in time T as the accident risk in the period of observation t multiplied by the ratio T/t. This is equivalent to the formula for Inc_s (equation A.3) only if $Inc_{t,s}$ is constant throughout T; however, this assumption does not always hold, as can be clearly observed in Table A-4 and Table A-5.

This case study used data collected in a motorway that may be particularly sensitive to the seasonal effect, as that area of Northern Italy experiences four very different seasons during the year, with cold, dry and foggy winters, relatively wetter springs and autumns, and hot, dry (but with occasional storms) summers (see Table A- 8). In other areas seasonality effect may have a lesser impact, in terms of predicted collisions.

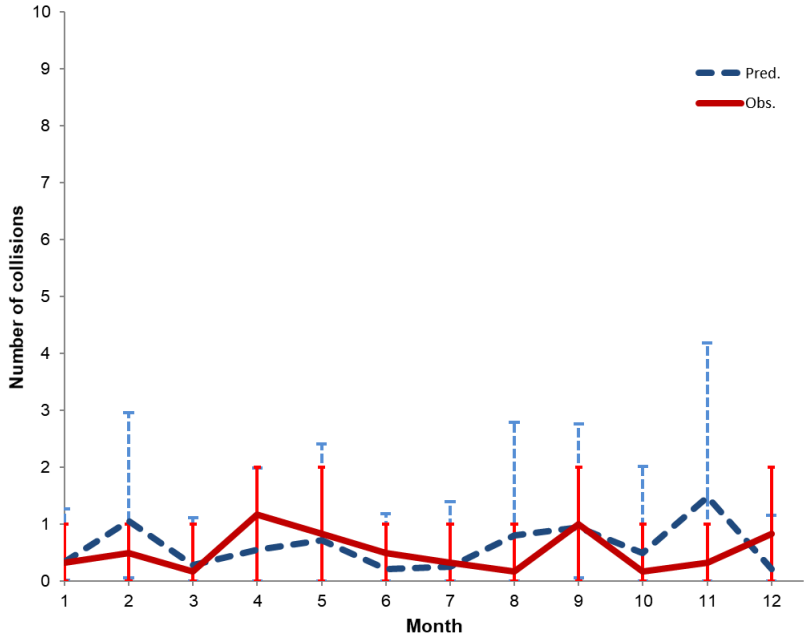

Figure A-1. Comparison between Set #1 predicted number of collisions (blue dotted line) and observed annual collisions (red solid line). Vertical bars represent respectively 95% confidence intervals and observed interval of annual collisions.

Table A- 8. Climate data (1971-2000) for a city located on the motorway path

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Daily mean [°C]	2.5	4.5	8.4	12.0	17.2	20.8	23.6	23.3	19.0	13.3	7.1	3.1
Avg. precipitations [mm]	50.9	43.3	48.7	70.4	74.2	87.2	62.6	81.7	76.2	91.0	64.8	52.5
Avg. precipitation days (>1mm)	6.8	5.1	6.0	8.9	8.6	8.6	5.5	5.8	6.0	7.4	7.1	6.2
Avg. relative humidity (%)	85	78	73	75	73	73	73	74	76	81	84	84
Avg. monthly sunshine hours	94	102	156	180	241	255	304	262	199	158	72	81

Several interesting findings emerged from this study.

First of all, it was shown that it is possible to aggregate data from several road sections, provided that they have similar geometric, traffic and weather characteristics, to estimate EVT models, whose performance is even better than traditional EVT models estimated with data collected in single road sections. Moreover, this kind of model is transferable with very good results to other road sections. This shows once more the potential of EVT

in practical road safety application: it is possible to estimate models using data collected in a small number of sections to produce collision probabilities which can be used for road safety estimation in other locations with similar characteristics.

Secondly, different EVT models were estimated for different months, producing significantly different results between each other, but consistent with the seasonal variation of the observed number of collisions. This analysis stresses the importance to take into account seasonality effect in practical real-world applications: the collision predictions obtained from EVT models estimated with data collected during an observation period t are valid only for that period; trying to extend those predictions to a larger time period T can be risky and may lead to large overestimation or underestimation of collision risk. Therefore, if the objective of a road safety analysis is to calculate the overall annual collision risk, in order to obtain more reliable predictions, it is recommended to collect data in different period of the year and, if possible, in different weather conditions.

Future developments will involve the use of covariates (Méndez et al., 2006), in order to transfer collision predictions to sections with different traffic, weather and geometrical characteristics. Moreover, the effect of the lane on the collision risk could be analyzed.

Appendix 2. Comparative study of machine learning classifiers for real-time conflict prediction[7]

This Appendix reports a comparative study on the performance of several machine learning classifiers within a RTConfPM.

A2. 1 Machine-learning classifiers

K-Nearest Neighbors

K-Nearest Neighbors (KNN) is one of the simplest classification methods in machine learning. The basic idea is that a new sample is classified by looking at the classes of the known samples closest to it. The two main parameters which have to be chosen are: k, i.e. the number of neighbor samples to be considered; the distance function (e.g. Euclidean). KNN can be summarized according to the following procedure:

The k-nearest neighbors to the new sample (according to the distance function) are selected;

For each class, it is counted how many of these k samples belong to it;

The new sample is classified into the class which includes the highest number of the k neighbors.

KNN has been applied in several RTCPM/RTConfPM (Katrakazas et al., 2018; Osman et al., 2019; Theofilatos et al., 2019).

[7] The work presented in this Appendix is part of the following publication:

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2021). Real-time conflict prediction: a comparative study of machine learning classifiers. *Transportation Research Procedia*, *52*, 292-299. (Orsini, Gecchele, Gastaldi, et al., 2021)

Decision tree

Decision trees (DT) are another simple and commonly used classification method in machine learning. A tree is built by splitting the original full dataset (i.e. the root node) into two branches, then the process is repeated for the two new subsets, and so on, until all the data of the subset at a node have the same target variable values, or when splitting no longer adds value to the predictions.

Tree growing is performed by recursively partitioning the target variable to maximize "purity" in the child node: indeed, the splits are performed so that each child node is more homogeneous than its parent node. The most commonly used split criterion in DTs is based on Gini's diversity index, calculated as:

$$Gini = 1 - \sum_{1}^{n} \rho_i^2 \qquad (A.5)$$

Where i is the class (e.g., crash/non-crash), n the number of classes, ρ_i the percentage of observations in class i.

DT has been used in several RTCPMs (Hossain & Muromachi, 2011; Theofilatos et al., 2019), and often as a variable-selection method (Yu & Abdel-Aty, 2013b).

Naïve Bayes

While KNN and DT/RF do not make any assumptions on the distribution of the underlying data, with Naïve Bayes (NB) technique, it is assumed that data come from a certain distribution, and therefore it is possible to treat data as a statistical sample. In NB classification it is assumed that observations in each response class are samples from probability distributions and, in particular, a separate distribution for each class. Knowing these probability distributions, it is possible to determine the probability of a new observation occurring at a given location in the feature space, under the assumption that it belongs to a specific class. Bayes theorem allows to "reverse" this, and to calculate the probability that the new observation comes from a given class:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} \qquad (A.6)$$

Where: $P(y|x)$ is the posterior probability of response class y, given predictor x; $P(x|y)$ the likelihood, i.e. the probability of the predictor given the class; $P(y)$ the prior probability of the class; $P(x)$ the prior probability of the predictor. The "naïve" assumption is that each predictor variable is independent in each response class, and this greatly simplifies the calculation, since performing a prediction only requires determining the probability of the observation and applying (2). Then, since it is assumed that observations are independent in each predictor, the probability is simply the product of the probabilities in each variable.

In the end, given a new sample, it is possible to calculate the probability of it belonging to each class, and subsequently to classify it to the most likely.

A benefit of this approach is that probabilities give some indications on how "clear" the classification is. Moreover, since predictions are based on the statistical distribution of all the observations, rather than on individual observations, they are robust to noise in the training data and to outliers.

NB has been applied in a number of real-time road safety applications (Osman et al., 2019; Theofilatos et al., 2019).

Discriminant analysis

Similar to NB, in discriminant analysis (DA) it is assumed that observations in each prediction class can be modeled with a normal probability distribution. However, there is no assumption of independence in each predictor, and a multivariate normal (MVN) distribution is fitted to each class.

Training data is used to fit MVN distributions for each response class and, subsequently, to find the location of the boundary between these classes. The boundary between two classes is defined as the locus of points in the feature space where there is equal probability of belonging to either class, and it depends on the parameters of the MVN distributions (means and variance-covariance matrices). If homoscedasticity is assumed, then the boundary is linear; alternatively, it is quadratic. As NB, DA is robust to noise and to outliers.

DA has been applied in few RTCPM/RTConfPM works (Caleffi et al., 2017; C. Xu et al., 2012).

Support vector machine

A description of SVM classifier is given in Section 3.2.3

A2.2 Results and discussion

The SMOTEd training dataset, obtained after data preparation procedures (see Section 3.3), was used to train several different classifiers: KNN, DT, DA, NB and SVM. As mentioned in Section A2.1, the training of each of these classifiers depends on some specific parameters of the classifier itself; therefore, each classifier was trained several times, using different combination of its parameters. For the sake of brevity, it was not feasible to report the results of all parameter combinations of each classifier; in this Appendix, only the results of the best performing combination are presented.

The following list contains the values of the parameters of the best combination, for each classifier.

- KNN: k=5; Euclidean distance function.
- DT: Gini split criterion.
- NB: normal distribution.
- DA: quadratic discriminant analysis.
- SVM: Linear kernel function, cost=1, gamma=1.

For robustness, Monte Carlo cross-validation with 100 repetitions was adopted. In practice, the 80%-20% split of SMOTEd dataset (see Section 3.3.3 and Figure 27) was repeated 100 times and each time classifiers were re-trained, and their performance evaluated. The values of Recall, Specificity and AUC presented in Table A-9 are the averages across the 100 repetitions.

The results reported in Table A-9, show generally very high performance, compared to many RTCPM in the literature (Theofilatos et al., 2019; Yuan et al., 2019). Training and Test indicators are almost identical for all classifiers, except for DT, which suffers from a decrease (around 0.03-0.04) in all indicators, suggesting that there may be some slight overfitting problem.

The most relevant indicators, in order to evaluate the performance of the classifiers are those related to the Full-test dataset, i.e. the real-world unbalanced dataset. Here DT approach shows another decrease in all indicators, proving to be the least reliable of the five classifiers compared in this work. Also the NB approach is not completely satisfying, presenting the lowest Specificity of all classifiers, whereas DA overperforms in all indicators. SVM and KNN results are in general very good, with excellent Recall and Specificity and AUC only lightly worse than DA.

Overall, considering that Recall is the most important indicator under a safety point of view, SVM and KNN appear to be the best classifiers, since they both are able to correctly predict more than 98% of the real-world unsafe situations; on the other hand, DA, despite having slightly better Specificity and AUC, fails to predict more than 8% of unsafe situations (6% more than SVM and KNN).

In view of practical applications, it is important to have low False Alarm Rate (1-Specificity), because a system which triggers false alarms too often would be unfeasible to apply and/or would compromise the effectiveness of intervention strategies. Under this point of view, SVM and KNN provide good results, with rates lower than 7%; therefore, a false alarm is triggered only once out of 18~19 5-minutes time intervals.

Another important aspect to consider is consistency, i.e. how much the predicting performance depends on the specific partition of training/test datasets. In order to evaluate consistency, we analyzed the 95% confidence intervals (across the 100 repetition of the Monte Carlo cross-validation) of the indicators in the full-test dataset (see Figure A-2). In

particular, considering Recall indicator, it is possible to appreciate that both SVM and KNN not only have the best average value, but also the narrowest confidence interval, underlining once more that they are the best performing classifiers in this case study.

Table A-9. Comparison of classifiers performance.

Dataset	Training			Test			Full-test		
Indicator	Recall	Specificity	AUC	Recall	Specificity	AUC	Recall	Specificity	AUC
KNN	0.994	0.945	0.999	0.993	0.936	0.970	0.983	0.936	0.962
DT	0.990	0.977	0.997	0.955	0.940	0.959	0.861	0.940	0.915
NB	0.930	0.909	0.968	0.929	0.909	0.968	0.903	0.909	0.961
DA	0.955	0.957	0.974	0.953	0.956	0.973	0.919	0.956	0.964
SVM	0.982	0.934	0.952	0.982	0.934	0.953	0.983	0.933	0.948

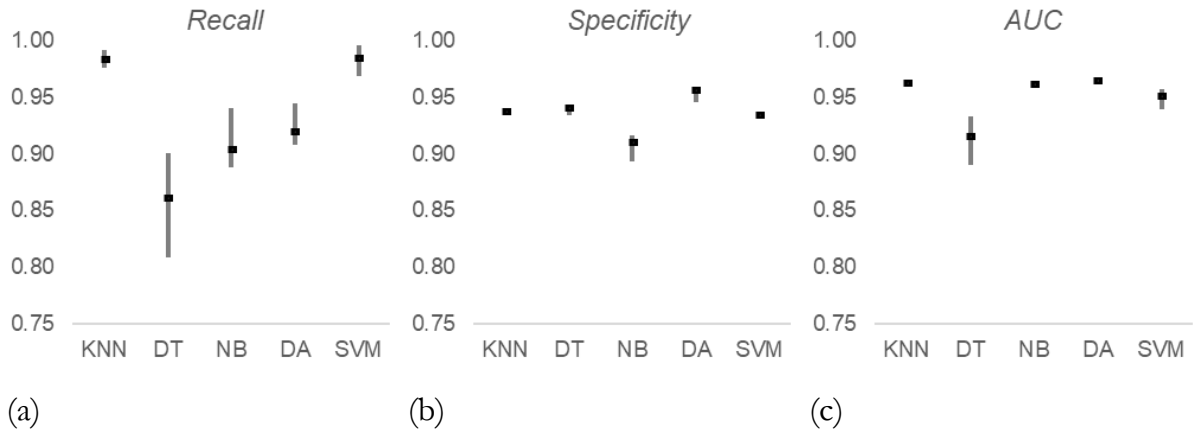

Figure A-2. Full-test dataset. Average values (black square) and 95% confidence intervals (grey bars), for each classifier, across 100 repetitions of Monte Carlo cross-validation for: (a) Recall; (b) Specificity; (c) AUC.

This work compared several machine learning classifiers for real-time conflict prediction, by applying them for predicting rear-end unsafe situations in a real-world motorway case study. According to our findings, SVM and KNN appear to be the most effective, providing by far the best and most consistent values of Recall indicator, which is the most relevant under a safety point of view. Moreover, having a high Specificity (and therefore low False Alarm Rate) they show good potential for practical applications.

A possible explanation of why they are the best classifiers in this application is that they do not make assumptions on the probability distribution of training data, contrarily to NB and DA. In NB, it is assumed that each class (safe vs unsafe) has a different distribution and that each predictor variable is independent in each response class. DA, which relaxes the second assumption, provides better results, hinting that indeed their lower performance could be attributed to these assumptions. DT, like SVM and KNN does not make assumption on the distribution of data; however, the DT model implemented in this case study appears to suffer from overfitting, which indeed is something that often affects this approach. A possible solution could be to apply an ensemble learning approach and consider multiple decision trees: one of the most popular of such approaches is Random Forest.

In the future, other types of approaches will be investigated: (i) statistical models, such as classic and Bayesian logistics regressions, which, in addition of providing yes-no predictions, could offer more insights on the effects of the various input variables; (ii) ensemble learning and deep learning approaches, which may be able to improve performance by exploiting more complex relations between data.

References

AASHTO. (2009). *AASHTO guidelines for traffic data programs*.

Abdel-Aty, M., & Haleem, K. (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. *Accident Analysis and Prevention*, *43*(1), 461–470. https://doi.org/10.1016/j.aap.2010.10.002

Abdel-Aty, M., Pande, A., Das, A., & Knibbe, W. J. (2008). Assessing safety on Dutch freeways with data from infrastructure-based intelligent transportation systems. *Transportation Research Record*, *2083*(1), 153–161. https://doi.org/10.3141/2083-18

Abdel-Aty, M., Pande, A., Lee, C., Gayah, V., & Santos, C. Dos. (2007). Crash risk assessment using intelligent transportation systems data and real-time intervention strategies to improve safety on freeways. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. https://doi.org/10.1080/15472450701410395

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M. F., & Hsia, L. (2004). Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record*, *1897*(1), 88–95. https://doi.org/10.3141/1897-12

Abdel-Aty, M., & Wang, L. (2017). Implementation of Variable Speed Limits to Improve Safety of Congested Expressway Weaving Segments in Microsimulation. *Transportation Research Procedia*, *27*, 577–584. https://doi.org/10.1016/j.trpro.2017.12.061

Ahmed, M., Abdel-Aty, M., & Yu, R. (2012). Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. *Transportation Research Record*, *2280*(1), 51–59. https://doi.org/10.3141/2280-06

Allen, B. L., Shin, B. T., & Cooper, P. (1978). Analysis of Traffic Conflicts and Collisions. *Transportation Research Record*, *667*, 67–74.

Arun, A., Haque, M. M., Bhaskar, A., Washington, S., & Sayed, T. (2021). A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accident Analysis & Prevention*, *153*, 106016. https://doi.org/10.1016/J.AAP.2021.106016

Astarita, V., & Giofré, V. P. (2019). From traffic conflict simulation to traffic crash simulation: Introducing traffic safety indicators based on the explicit simulation of potential driver errors. *Simulation Modelling Practice and Theory*. https://doi.org/10.1016/j.simpat.2019.03.003

Basso, F., Basso, L. J., Bravo, F., & Pezoa, R. (2018). Real-time crash prediction in an urban expressway using disaggregated data. *Transportation Research Part C: Emerging Technologies*, *86*, 202–219. https://doi.org/10.1016/j.trc.2017.11.014

Basso, F., Basso, L. J., & Pezoa, R. (2020). The importance of flow composition in real-time crash prediction. *Accident Analysis & Prevention*, *137*, 105436. https://doi.org/https://doi.org/10.1016/j.aap.2020.105436

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D., & Ferro, C. (2004). Statistics of extremes: Theory and applications. In *Statistics of Extremes: Theory and Applications*. https://doi.org/10.1002/0470012382

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*(1), 106. https://doi.org/10.1186/1471-2105-14-106

Breiman L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burdett, B., Bill, A. R., & Noyce, D. A. (2017). Evaluation of Roundabout-Related Single-Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2637*, 17–26. https://doi.org/10.3141/2637-03

Caleffi, F., Anzanello, M. J., & Cybis, H. B. B. (2017). A multivariate-based conflict prediction model for a Brazilian freeway. *Accident Analysis and Prevention*, *98*, 295–302. https://doi.org/10.1016/j.aap.2016.10.025

Campbell, K. L., Joksch, H. C., & Green, P. E. (1996). *A bridging analysis for estimating the benefits of active safety technologies*.

Cavadas, J., Azevedo, C. L., Farah, H., & Ferreira, A. (2020). Road safety of passing maneuvers: a bivariate extreme value theory approach under non-stationary conditions. *Accident Analysis & Prevention*, *134*, 105315. https://doi.org/10.1016/j.aap.2019.105315

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Coles, S. G. (2001). An introduction to statistical modeling of extreme values. In *Springer Series in Statistics*. https://doi.org/10.1007/978-1-4471-3675-0

Coles, S. G., & Tawn, J. A. (1994). Statistical Methods for Multivariate Extremes: An Application to Structural Design. *Applied Statistics*, *43*(1), 1–48. https://doi.org/10.2307/2986112

Collet, L., Beevers, L., & Prudhomme, C. (2017). Assessing the impact of climate change and extreme value uncertainty to extreme flows across Great Britain. *Water (Switzerland)*, *9*(2). https://doi.org/10.3390/w9020103

Cooper, D. F., & Ferguson, N. (1976). Traffic studies at T-junctions. 2. A conflict simulation model. *Traffic Engineering & Control*, *17*(7), 306–309. https://trid.trb.org/view/66554

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1007/bf00994018

Daganzo, C. F. (1997). Fundamentals of Transportation and Traffic Operations. In *Fundamentals of Transportation and Traffic Operations*. https://doi.org/10.1108/9780585475301

Das, B., & Ghosh, S. (2016). Detecting tail behavior: mean excess plots with confidence bounds. *Extremes*, *19*(2), 325–349. https://doi.org/10.1007/s10687-015-0238-9

Das, S., & Maurya, A. K. (2020). Bivariate modeling of time headways in mixed traffic streams: a copula approach. *Transportation Letters*, *12*(2), 138–148. https://doi.org/10.1080/19427867.2018.1537209

Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. *Accident Analysis and Prevention*, *43*(6), 1907–1919. https://doi.org/10.1016/j.aap.2011.05.001

Dimitriou, L., Stylianou, K., & Abdel-Aty, M. (2018). Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. *Accident Analysis and Prevention*, *118*, 221–235. https://doi.org/10.1016/j.aap.2018.02.024

Dong, N., Huang, H., & Zheng, L. (2015). Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. *Accident Analysis and Prevention*, *82*, 192–198. https://doi.org/10.1016/j.aap.2015.05.018

El-Basyouny, K., & Sayed, T. (2013). Safety performance functions using traffic conflicts. *Safety Science*, *51*(1), 160–164. https://doi.org/10.1016/J.SSCI.2012.04.015

Eltrass, A., & Khalil, M. (2018). Automotive radar system for multiple-vehicle detection and tracking in urban environments. *IET Intelligent Transport Systems*, *12*(8), 783–792. https://doi.org/10.1049/iet-its.2017.0370

Elvik, R. (2003). Effects on Road Safety of Converting Intersections to Roundabouts: Review of Evidence from Non-U.S. Studies. *Transportation Research Record: Journal of the Transportation Research Board*, *1847*, 1–10. https://doi.org/10.3141/1847-01

Elvik, R. (2017). Road safety effects of roundabouts: A meta-analysis. *Accident Analysis and Prevention*, *99*, 364–371. https://doi.org/10.1016/j.aap.2016.12.018

Essa, M., & Sayed, T. (2018). Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. *Transportation Research Part C: Emerging Technologies*, *89*, 289–302. https://doi.org/10.1016/j.trc.2018.02.014

ETSC. (2016). *Reducing casualties involving young drivers and riders in Europe*. http://etsc.eu/wp-content/uploads/2017_01_26_young_drivers_report.pdf

Farah, H., & Azevedo, C. L. (2017). Safety analysis of passing maneuvers using extreme value theory. *IATSS Research*, *41*(1), 12–21. https://doi.org/10.1016/j.iatssr.2016.07.001

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fish, K. E., & Blodgett, J. G. (2003). A visual method for determining variable importance in an artificial neural network model: An empirical benchmark study. *Journal of Targeting, Measurement and Analysis for Marketing*, *11*(3), 244–254. https://doi.org/10.1057/palgrave.jt.5740081

Flach, P., Hernández-Orallo, J., & Ferri, C. (2011). A coherent interpretation of AUC as a measure of aggregated classification performance. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*.

Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., & Yuan, J. (2020). Predicting real-time traffic conflicts using deep learning. *Accident Analysis and Prevention*, *136*, 105429. https://doi.org/10.1016/j.aap.2019.105429

Fu, C., Sayed, T., & Zheng, L. (2020). Multivariate Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. *Analytic Methods in Accident Research*, *28*, 100135. https://doi.org/10.1016/j.amar.2020.100135

Gastaldi, M., Orsini, F., Gecchele, G., & Rossi, R. (2021). Safety analysis of unsignalized intersections: a bivariate extreme value approach. *Transportation Letters*, *13*(3), 209–218. https://doi.org/10.1080/19427867.2020.1861503

Gecchele, G., Orsini, F., Gastaldi, M., & Rossi, R. (2019). Freeway rear-end collision risk estimation with extreme value theory approach. A case study. *Transportation Research Procedia*, *37*, 195–202. https://doi.org/10.1016/j.trpro.2018.12.183

Gettman, D., & Head, L. (2003). Surrogate Safety Measures from Traffic Simulation Models. *Transportation Research Record*, *1840*(1), 104–115. https://doi.org/10.3141/1840-12

Gilat, A. (2014). *MATLAB: An Introduction with Applications* (5th editio). Wiley.

Gilli, M., & Këllezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*. https://doi.org/10.1007/s10614-006-9025-7

Glauz, W. D., & Migletz, D. J. (1980). APPLICATION OF TRAFFIC CONFLICT ANALYSIS AT INTERSECTIONS. *NCHRP Report*, *219*.

Gu, X., Abdel-Aty, M., Lee, J., Xiang, Q., & Ma, Y. (2020). Identification of contributing factors for interchange crashes based on a quasi-induced exposure method. *Journal of Transportation Safety and Security*. https://doi.org/10.1080/19439962.2020.1812783

Guo, F., Klauer, S. G., Hankey, J. M., & Dingus, T. A. (2010). Near Crashes as Crash Surrogate for Naturalistic Driving Studies: *Https://Doi.Org/10.3141/2147-09*, *2147*, 66–74. https://doi.org/10.3141/2147-09

Güttinger, V. A. (1984). Conflict Observation in Theory and in Practice. In *International Calibration Study of Traffic Conflict Techniques*. https://doi.org/10.1007/978-3-642-82109-7_3

Ha, D.-H., Aron, M., & Cohen, S. (2012). Time headway variable and probabilistic modeling. *Transportation Research Part C: Emerging Technologies*, *25*, 181–201. https://doi.org/https://doi.org/10.1016/j.trc.2012.06.002

Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis and Prevention*, *14*(5), 359–364. https://doi.org/10.1016/0001-4575(82)90014-8

Hayward J.C. (1972). Near miss determination through use of a scale of danger. In *Highway Research Record* (Vol. 384, Issue 384, pp. 24–34). https://doi.org/TTSC 7115

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hossain, M., Abdel-Aty, M., Quddus, M. A., Muromachi, Y., & Sadeek, S. N. (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accident Analysis and Prevention*, *124*, 66–84. https://doi.org/10.1016/j.aap.2018.12.022

Hossain, M., & Muromachi, Y. (2011). Understanding Crash Mechanisms and Selecting Interventions to Mitigate Real-Time Hazards on Urban Expressways. *Transportation Research Record: Journal of the Transportation Research Board*, *2213*, 53–62. https://doi.org/10.3141/2213-08

Hossain, M., & Muromachi, Y. (2012). A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis and Prevention*, *45*, 373–381. https://doi.org/10.1016/j.aap.2011.08.004

Hughes, R. G., & Council, F. M. (1999). Exploratory Investigation of Relationship Between Congestion and Safety in Terms of Real Time Traffic Performance. *78th Annual*

Meeting of the Transportation Research Board, Washington, DC.

Hydén, C. (1987). The development of a method for traffic safety evaluation:the Swedish traffic conflicts technique. *Bulletin Lund Institute Od Technology*, *70*.

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. *Accident Analysis and Prevention*, *130*, 84–90. https://doi.org/10.1016/j.aap.2017.02.022

Ito, Y., Yamada, Y., Minemura, A., Tsuchida, J., & Shimizu, M. (2017). *ESTIMATED TIME-TO-COLLISION (TTC) CALCULATION APPARATUS AND ESTIMATED TTC CALCULATION METHOD* (Issue 20170210360). http://www.freepatentsonline.com/y2017/0210360.html

Johnsson, C., Laureshyn, A., & De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. *Transport Reviews*, *38*(6), 1–21. https://doi.org/10.1080/01441647.2018.1442888

Jonasson, J. K., & Rootzén, H. (2014). Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach. *Accident Analysis and Prevention*, *62*, 102–109. https://doi.org/10.1016/j.aap.2013.09.013

Katrakazas, C., Antoniou, C., & Yannis, G. (2019). Time Series Classification Using Imbalanced Learning for Real-Time Safety Assessment. *Transportation Research Board 98th Annual Meeting*.

Katrakazas, C., Quddus, M., & Chen, W. H. (2017). A simulation study of predicting conflict-prone traffic conditions in Real-time. *96th Annual Meeting of Transportation Research Board (TRB)*.

Katrakazas, C., Quddus, M., & Chen, W. H. (2018). A Simulation Study of Predicting Real-Time Conflict-Prone Traffic Conditions. *IEEE Transactions on Intelligent Transportation Systems*, *19*(10), 3196–3207. https://doi.org/10.1109/TITS.2017.2769158

Ke, J., Zhang, S., Yang, H., & Chen, X. (2019). PCA-based missing information imputation for real-time crash likelihood prediction under imbalanced data. *Transportmetrica A: Transport Science*, *15*(2), 872–895. https://doi.org/10.1080/23249935.2018.1542414

Knoop, V., Hoogendoorn, S. P., & Van Zuylen, H. (2009). Empirical differences between time mean speed and space mean speed. *Traffic and Granular Flow 2007*, 351–356. https://doi.org/10.1007/978-3-540-77074-9_36

Kumar, N., & Kalyani, S. (2017). Modeling the Behavior of Peaks of OFDM Signal Using 'Peaks Over Threshold' Approach. *IEEE Transactions on Wireless Communications*, *16*(6), 3590–3600. https://doi.org/10.1109/TWC.2017.2685499

Kusano, K. D., & Gabler, H. (2011). Method for Estimating Time to Collision at Braking in Real-World, Lead Vehicle Stopped Rear-End Crashes for Use in Pre-Crash System Design. *SAE International Journal of Passenger Cars - Mechanical Systems*, *4*(1), 435–443. https://doi.org/10.4271/2011-01-0576

Laureshyn, A., Svensson, Å., & Hydén, C. (2010). Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. *Accident Analysis and Prevention*, *42*(6), 1637–1646. https://doi.org/10.1016/j.aap.2010.03.021

Learn, S., Ma, J., Raboy, K., Zhou, F., & Guo, Y. (2018). Freeway speed harmonisation

experiment using connected and automated vehicles. *IET Intelligent Transport Systems*, *12*(5), 319–326. https://doi.org/10.1049/iet-its.2017.0149

Lee, C., Saccomanno, F., & Hellinga, B. (2002). Analysis of crash precursors on instrumented freeways. *Transportation Research Record*, *1784*(1), 1–8. https://doi.org/10.3141/1784-01

Li, J., Guo, J., Wijnands, J. S., Yu, R., Xu, C., & Stevenson, M. (2020). Assessing injury severity of secondary incidents using support vector machines. *Journal of Transportation Safety and Security*. https://doi.org/10.1080/19439962.2020.1754983

Li, P., Abdel-Aty, M., & Yuan, J. (2020). Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis & Prevention*, *135*, 105371.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. https://cran.r-project.org/doc/Rnews/

Lv, Y., Tang, S., & Zhao, H. (2009). Real-time highway traffic accident prediction based on the k-nearest neighbor method. *2009 International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2009*. https://doi.org/10.1109/ICMTMA.2009.657

Lv, Y., Tang, S., Zhao, H., & Li, S. (2009). Real-time highway accident prediction based on support vector machines. *2009 Chinese Control and Decision Conference, CCDC 2009*. https://doi.org/10.1109/CCDC.2009.5192409

Madanat, S., & Liu, P.-C. (1995). *A prototype system for real-time incident likelihood prediction*. United States. Federal Highway Administration.

Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research*, *17*, 1–13. https://doi.org/10.1016/J.AMAR.2017.10.002

Mannering, F., Bhat, C. R., Shankar, V., & Abdel-Aty, M. (2020). Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic Methods in Accident Research*, *25*, 100113. https://doi.org/10.1016/J.AMAR.2020.100113

Manolis, D., Spiliopoulou, A., Vandorou, F., & Papageorgiou, M. (2020). Real time adaptive cruise control strategy for motorways. *Transportation Research Part C: Emerging Technologies*, *115*, 102617. https://doi.org/10.1016/J.TRC.2020.102617

Méndez, F. J., Menéndez, M., Luceño, A., & Losada, I. J. (2006). Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *Journal of Geophysical Research: Oceans*, *111*(C7). https://doi.org/10.1029/2005JC003344

Menon, A. K. (2009). Large-scale support vector machines: algorithms and theory. *Research Exam, University of California, San Diego*.

Montella, A. (2011). Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. *Accident Analysis and Prevention*, *43*(4), 1451–1463. https://doi.org/10.1016/j.aap.2011.02.023

Nadimi, N., Ragland, D. R., & Mohammadian Amiri, A. (2020). An evaluation of time-to-collision as a surrogate safety measure and a proposal of a new method for its application in safety analysis. *Transportation Letters*, *12*(7), 491–500.

https://doi.org/10.1080/19427867.2019.1650430

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2019). Collision prediction in roundabouts: a comparative study of extreme value theory approaches. *Transportmetrica A: Transport Science*, *15*(2), 556–572. https://doi.org/10.1080/23249935.2018.1515271

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2020a). Large-scale road safety evaluation using extreme value theory. *IET Intelligent Transport Systems*, *14*(9), 1004–1012. https://doi.org/10.1049/iet-its.2019.0633

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2020b). Transferability and seasonality in extreme value theory applications to road safety: a case study in an Italian motorway. *Advances in Transportation Studies*, *2*(Special Issue), 33–46.

Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2021). Real-time conflict prediction: A comparative study of machine learning classifiers. *Transportation Research Procedia*, *52*, 292–299. https://doi.org/10.1016/j.trpro.2021.01.034

Orsini, F., Gecchele, G., Rossi, R., & Gastaldi, M. (2021). A conflict-based approach for real-time road safety analysis: Comparative evaluation with crash-based models. *Accident Analysis & Prevention*, *161*, 106382. https://doi.org/10.1016/j.aap.2021.106382

Osman, O. A., Hajij, M., Bakhit, P. R., & Ishak, S. (2019). Prediction of Near-Crashes from Observed Vehicle Kinematics using Machine Learning. *Transportation Research Record*, *2673*(12), 463–473. https://doi.org/10.1177/0361198119862629

Paikari, E., Moshirpour, M., Alhajj, R., & Far, B. H. (2014). Data integration and clustering for real time crash prediction. *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration, IEEE IRI 2014*. https://doi.org/10.1109/IRI.2014.7051936

Pande, A., & Abdel-Aty, M. (2005). A freeway safety strategy for advanced proactive traffic management. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *9*(3), 145–158. https://doi.org/10.1080/15472450500183789

Pande, A., & Abdel-Aty, M. (2006). Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis and Prevention*, *38*(5), 936–948. https://doi.org/10.1016/j.aap.2006.03.004

Park, H., & Haghani, A. (2016). Real-time prediction of secondary incident occurrences using vehicle probe data. *Transportation Research Part C: Emerging Technologies*, *70*, 69–85. https://doi.org/10.1016/j.trc.2015.03.018

Park, H., Haghani, A., Samuel, S., & Knodler, M. A. (2018). Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion. *Accident Analysis and Prevention*, *112*, 39–49. https://doi.org/10.1016/j.aap.2017.11.025

Parsa, A. B., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2019). Real-time accident detection: Coping with imbalanced data. *Accident Analysis and Prevention*, *129*, 202–210. https://doi.org/10.1016/j.aap.2019.05.014

Peng, Y., Li, C., Wang, K., Gao, Z., & Yu, R. (2020). Examining imbalanced classification algorithms in predicting real-time traffic crash risk. *Accident Analysis and Prevention*, *144*, 105610. https://doi.org/10.1016/j.aap.2020.105610

Perkins, S. R., & Harris, J. L. (1968). Traffic conflict characteristics-accident potential at

intersections. *Highway Research Record*, *225*.

Polders, E., Daniels, S., Casters, W., & Brijs, T. (2015). Identifying Crash Patterns on Roundabouts. *Traffic Injury Prevention*, *16*(2), 202–207. https://doi.org/10.1080/15389588.2014.927576

Postacchini, M., Lalli, F., Memmola, F., Bruschi, A., Bellafiore, D., Lisi, I., Zitti, G., & Brocchini, M. (2019). A model chain approach for coastal inundation: Application to the bay of Alghero. *Estuarine, Coastal and Shelf Science*, *219*, 56–70. https://doi.org/10.1016/j.ecss.2019.01.013

Qu, X., Wang, W., Wang, W., & Liu, P. (2013). Real-time freeway sideswipe crash prediction by support vector machine. *IET Intelligent Transport Systems*, *7*(4), 445–453. https://doi.org/10.1049/iet-its.2011.0230

Rodegerdts, L., Blogg, M., Wemple, E., Myers, E., Kyte, M., Dixon, M., List, G., Flannery, A., Troutbeck, R., Brilon, E., Wu, N., Persaud, B. N., Lyon, C., Harkey, D., & Carter, D. (2007). *NCHRP Report 572: Roundabouts in the United States*.

Rossi, R., Gastaldi, M., Biondi, F., & Mulatti, C. (2013). Oppel-Kundt Illusion and lateral optic flow manipulation in affecting perceived speed in approaching roundabouts: experiments witha driving simulator. *Proceedings of the 92nd Transportation Research Board Meeting.*, 1–13.

Rossi, R., Gastaldi, M., Gecchele, G., Biondi, F., & Mulatti, C. (2014). Traffic-calming measures affecting perceived speed in approaching bends: On-field validated virtual environment. *Transportation Research Record*, *2434*, 35–43. https://doi.org/10.3141/2434-05

Rossi, R., Gastaldi, M., & Meneguzzer, C. (2018). Headway distribution effect on gap-acceptance behavior at roundabouts: driving simulator experiments in a case study. *Advances in Transportation Studies*, *46*, 97–110. https://doi.org/10.4399/9788255186418

Rossi, R., Gastaldi, M., Meneguzzer, C., & Gecchele, G. (2011). Gap-Acceptance Behavior at Priority Intersection: Field Observations Versus Experiments with Driving Simulator. *Proceedings of the 90th Transportation Research Board Meeting*, 1–14.

Rossi, R., Meneguzzer, C., & Gastaldi, M. (2013). Transfer and updating of Logit models of gap-acceptance and their operational implications. *Transportation Research Part C: Emerging Technologies*, *28*, 142–154. https://doi.org/10.1016/j.trc.2011.05.019

Rossi, R., Meneguzzer, C., Orsini, F., & Gastaldi, M. (2020). Gap-acceptance behavior at roundabouts: validation of a driving simulator environment using field observations. *Transportation Research Procedia*, *47*, 27–34. https://doi.org/10.1016/j.trpro.2020.03.069

Roy, A., Hossain, M., & Muromachi, Y. (2018). Enhancing the Prediction Performance of Real-Time Crash Prediction Models: A Cell Transmission-Dynamic Bayesian Network Approach. *Transportation Research Record*, *2672*(38), 58–68. https://doi.org/10.1177/0361198118797802

Sadeq, H., & Sayed, T. (2016). Automated Roundabout Safety Analysis: Diagnosis and Remedy of Safety Problems. *Journal of Transportation Engineering*, *142*(1–8), Content ID 04016062. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000887

Saha, P., Roy, R., Sarkar, A. K., & Pal, M. (2019). Preferred time headway of drivers on two-lane highways with heterogeneous traffic. *Transportation Letters*, *11*(4), 200–207.

https://doi.org/10.1080/19427867.2017.1312859

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., & Santos, J. (2018). Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [Research Frontier]. *IEEE Computational Intelligence Magazine*, *13*(4), 59–76. https://doi.org/10.1109/MCI.2018.2866730

Saunier, N., & Sayed, T. (2007). Automated analysis of road safety with video data. *Transportation Research Record*, *2019*(1), 57–64. https://doi.org/10.3141/2019-08

Sayed, T., & Zein, S. (1999). Traffic conflict standards for intersections. *Transportation Planning and Technology*, *22:4*(October 2014), 309–323. https://doi.org/10.1080/03081069908717634

Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. In *Revstat Statistical Journal* (Vol. 10, Issue 1, pp. 33–60).

Shaikh, U., & Thalkar, N. (2019). Vehicle Communication Systems: Technology and Review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3349639

Shekhar Babu, S., & Vedagiri, P. (2018). Proactive safety evaluation of a multilane unsignalized intersection using surrogate measures. *Transportation Letters*, *10*(2), 104–112. https://doi.org/10.1080/19427867.2016.1230172

Shi, Q., & Abdel-Aty, M. (2015). Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2015.02.022

Simon, R. (2007). Resampling strategies for model assessment and selection. In *Fundamentals of Data Mining in Genomics and Proteomics*. https://doi.org/10.1007/978-0-387-47509-7-8

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*(1), 67–90. https://doi.org/10.1093/biomet/72.1.67

Songchitruksa, P., & Tarko, A. P. (2006). The extreme value theory approach to safety estimation. *Accident Analysis and Prevention*, *38*(4), 811–822. https://doi.org/10.1016/j.aap.2006.02.003

St-Aubin, P., Saunier, N., & Miranda-Moreno, L. (2015). Large-scale automated proactive road safety analysis using video data. *Transportation Research Part C: Emerging Technologies*, *58*, 363–379. https://doi.org/10.1016/j.trc.2015.04.007

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307. https://doi.org/10.1186/1471-2105-9-307

Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, *40*(12), 3358–3378. https://doi.org/10.1016/j.patcog.2007.04.009

Tageldin, A., Sayed, T., & Shaaban, K. (2017). Comparison of time-proximity and evasive action conflict measures case studies from five cities. In *Transportation Research Record*. https://doi.org/10.3141/2661-03

Tarko, A. P. (2012). Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis and Prevention*, *45*, 230–240. https://doi.org/10.1016/j.aap.2011.07.008

Tarko, A. P. (2018a). Chapter 17. Surrogate Measures of Safety. In *Safe Mobility: Challenges, Methodology and Solutions* (pp. 383–405). https://doi.org/10.1108/S2044-994120180000011019

Tarko, A. P. (2018b). Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution – A theoretical and numerical exploration. *Accident Analysis and Prevention*, *113*, 63–73. https://doi.org/10.1016/j.aap.2018.01.008

Tarko, A. P. (2019). Measuring road safety with surrogate events. In A. P. Tarko (Ed.), *Measuring Road Safety with Surrogate Events*. Elsevier. https://doi.org/10.1016/C2016-0-00255-3

Tarko, A. P., Ariyur, K. B., Romero, M. A., Bandaru, V. K., & Lizarazo, C. G. (2016). *TScan: Stationary LiDAR for Traffic and Safety Studies—Object Detection and Tracking*. https://doi.org/10.5703/1288284316347

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2018.08.003

Theofilatos, A., Chen, C., & Antoniou, C. (2019). Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction. *Transportation Research Record*, *2673*(8), 169–178.

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis and Prevention*, *72*, 244–256. https://doi.org/10.1016/j.aap.2014.06.017

Thompson, P., Cai, Y., Reeve, D., & Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, *56*(10), 1013–1021. https://doi.org/10.1016/j.coastaleng.2009.06.003

Tolle, J. E. (1971). The Lognormal Headway Distribution Model. *Traffic Engineering & Control*, *13*(1), 22–24. http://trid.trb.org/view.aspx?id=117094

Ueyama, M. (1997). Effects of vehicular behavior at the pre-accident stage on decision-making at intersections without signals. *Third International Symposium on Intersections Without Traffic SignalsTransportation Research Board; Federal Highway Administration; National Center for Advanced Transportation Technology, University of Idaho; Transportation Northwest, University of Washi*.

Ural, S., Shan, J., Romero, M. A., & Tarko, A. (2015). Road and roadside feature extraction using imagery and lidar data for transportation operation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. https://doi.org/10.5194/isprsannals-II-3-W4-239-2015

Van Der Horst, R. (1990). *A Time-based Analysis of Road User Behaviour in Normal and Critical Encounters* [Institute for Perception TNO]. https://doi.org/uuid:8fb40be7-fae1-4481-bc37-12a7411b85c7

Van Der Horst, R., & Hogema, J. (1993). Time-to-collision and collision avoidance systems. *Proceedings of The 6th Workshop of the International*, 1–12. https://doi.org/10.1.1.511.3548

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. In *The Nature of Statistical Learning Theory*. Springer. https://doi.org/10.1007/978-1-4757-2440-0

Wang, Chao, Quddus, M. A., & Ison, S. G. (2013). The effect of traffic and road characteristics on road safety: A review and future research direction. *Safety Science*, *57*, 264–275. https://doi.org/10.1016/j.ssci.2013.02.012

Wang, Chen, Xie, Y., Huang, H., & Liu, P. (2021). A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis & Prevention*, *157*, 106157. https://doi.org/10.1016/J.AAP.2021.106157

Wang, Chen, Xu, C., & Dai, Y. (2019). A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accident Analysis & Prevention*, *123*, 365–373. https://doi.org/https://doi.org/10.1016/j.aap.2018.12.013

Wang, J., Luo, T., & Fu, T. (2019). Crash prediction based on traffic platoon characteristics using floating car trajectory data and the machine learning approach. *Accident Analysis and Prevention*, *133*, 105320. https://doi.org/10.1016/j.aap.2019.105320

Wang, L., Abdel-Aty, M., & Lee, J. (2017). Safety analytics for integrating crash frequency and real-time risk modeling for expressways. *Accident Analysis and Prevention*, *104*, 58–64. https://doi.org/10.1016/j.aap.2017.04.009

Wang, L., Abdel-Aty, M., Lee, J., & Shi, Q. (2019). Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis and Prevention*, *122*, 378–384. https://doi.org/10.1016/j.aap.2017.06.003

Wang, Pengfei, Chen, K., Zhu, S., Wang, P., & Zhang, H. (2020). Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. *Resources, Conservation and Recycling*. https://doi.org/10.1016/j.resconrec.2020.104814

Wang, Pin, & Chan, C. Y. (2017). Vehicle collision prediction at intersections based on comparison of minimal distance between vehicles and dynamic thresholds. *IET Intelligent Transport Systems*, *11*(10), 676–684. https://doi.org/10.1049/iet-its.2017.0065

Wang, R., Xu, Z., Zhao, X., & Hu, J. (2019). V2V-based method for the detection of road traffic congestion. *IET Intelligent Transport Systems*, *13*(5), 880–885. https://doi.org/10.1049/iet-its.2018.5177

Wang, Y. T., & Liu, X. Y. (2017). Seasonal passenger flow model of an inter-city expressway based on ARIMA. *Advances in Transportation Studies*, *3*, 111–120. https://doi.org/10.4399/97888255108292l0

Williams, M. J. (1981). Validity of the traffic conflicts technique. *Accident Analysis & Prevention*, *13*(2), 133–145. https://doi.org/10.1016/0001-4575(81)90025-7

Wu, Y., Abdel-Aty, M., Cai, Q., Lee, J., & Park, J. (2018). Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data. *Transportation Research Part C: Emerging Technologies*, *87*, 11–25. https://doi.org/10.1016/j.trc.2017.12.012

Wu, Y., Abdel-Aty, M., & Lee, J. (2018). Crash risk analysis during fog conditions using real-time traffic data. *Accident Analysis and Prevention*, *114*, 4–11. https://doi.org/10.1016/j.aap.2017.05.004

Xu, C., Liu, P., Wang, W., & Li, Z. (2012). Evaluation of the impacts of traffic states on

crash risks on freeways. *Accident Analysis and Prevention*, *47*, 162–171. https://doi.org/10.1016/j.aap.2012.01.020

Xu, C., Tarko, A. P., Wang, W., & Liu, P. (2013). Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis and Prevention*, *57*, 30–39. https://doi.org/10.1016/j.aap.2013.03.035

Xu, C., Wang, W., & Liu, P. (2013). A genetic programming model for real-time crash prediction on freeways. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 574–586. https://doi.org/10.1109/TITS.2012.2226240

Xu, Q. S., & Liang, Y. Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, *56*(1), 1–11. https://doi.org/10.1016/S0169-7439(00)00122-2

Yasmin, S., Eluru, N., Wang, L., & Abdel-Aty, M. A. (2018). A joint framework for static and real-time crash risk analysis. *Analytic Methods in Accident Research*. https://doi.org/10.1016/j.amar.2018.04.001

Yu, R., & Abdel-Aty, M. (2013a). Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes. *Accident Analysis and Prevention*, *58*, 97–105. https://doi.org/10.1016/j.aap.2013.04.025

Yu, R., & Abdel-Aty, M. (2013b). Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis and Prevention*, *51*, 252–259. https://doi.org/10.1016/j.aap.2012.11.027

Yu, R., & Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, *63*, 50–56. https://doi.org/10.1016/j.ssci.2013.10.012

Yu, R., Abdel-Aty, M., & Ahmed, M. (2013). Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2012.05.011

Yu, R., Abdel-Aty, M., Ahmed, M. M., & Wang, X. (2014). Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. *IEEE Transactions on Intelligent Transportation Systems*, *15*(1), 205–213. https://doi.org/10.1109/TITS.2013.2276089

Yuan, J., & Abdel-Aty, M. (2018). Approach-level real-time crash risk analysis for signalized intersections. *Accident Analysis and Prevention*, *119*, 274–289. https://doi.org/10.1016/j.aap.2018.07.031

Yuan, J., Abdel-Aty, M., Gong, Y., & Cai, Q. (2019). Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. *Transportation Research Record*, *2673*(4), 314–326. https://doi.org/10.1177/0361198119840611

Yuan, J., Abdel-Aty, M., Wang, L., Lee, J., Yu, R., & Wang, X. (2018). Utilizing bluetooth and adaptive signal control data for real-time safety analysis on urban arterials. *Transportation Research Part C: Emerging Technologies*, *97*, 114–127. https://doi.org/10.1016/j.trc.2018.10.009

Zheng, D., Qin, X., Tillman, R., & Noyce, D. A. (2010). Negotiation-based conflict exposure methodology in roundabout crash pattern analysis. *Proceedings of the 89th Transportation Research Board Annual Meeting*.

Zheng, L., Ismail, K., & Meng, X. (2014a). Freeway safety estimation using extreme value

theory approaches: A comparative study. *Accident Analysis and Prevention*, *62*, 32–41. https://doi.org/10.1016/j.aap.2013.09.006

Zheng, L., Ismail, K., & Meng, X. (2014b). Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes. *Safety Science*, *64*, 155–162. https://doi.org/10.1016/j.ssci.2013.12.003

Zheng, L., Ismail, K., Sayed, T., & Fatema, T. (2018). Bivariate extreme value modeling for road safety estimation. *Accident Analysis & Prevention*, *120*, 83–91. https://doi.org/https://doi.org/10.1016/j.aap.2018.08.004

Zheng, L., & Sayed, T. (2019a). From univariate to bivariate extreme value models: Approaches to integrate traffic conflict indicators for crash estimation. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2019.04.015

Zheng, L., & Sayed, T. (2019b). Application of Extreme Value Theory for Before-After Road Safety Analysis. *Transportation Research Record*, 0361198119841555. https://doi.org/10.1177/0361198119841555

Zheng, L., & Sayed, T. (2020). A novel approach for real time crash prediction at signalized intersections. *Transportation Research Part C: Emerging Technologies*, *117*. https://doi.org/10.1016/j.trc.2020.102683

Zheng, L., Sayed, T., & Essa, M. (2019). Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. *Accident Analysis and Prevention*, *123*, 314–323. https://doi.org/10.1016/j.aap.2018.12.007

Zheng, L., Sayed, T., & Mannering, F. (2021). Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Analytic Methods in Accident Research*, *29*, 100142. https://doi.org/10.1016/J.AMAR.2020.100142

Zheng, L., Sayed, T., & Tageldin, A. (2018). Before-after safety analysis using extreme value theory: A case of left-turn bay extension. *Accident Analysis & Prevention*, *121*, 258–267. https://doi.org/https://doi.org/10.1016/j.aap.2018.09.023

Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis and Prevention*, *42*(2), 626–636. https://doi.org/10.1016/j.aap.2009.10.009