A method for predicting crash configurations using counterfactual simulations and real-world data

(article starts on next page)

A method for predicting crash configurations using counterfactual simulations and real-world data

Alexandros Leledakis [a,b,*], Magdalena Lindman [b], Jonas Östh [a,b], Linus Wågström [b], Johan Davidsson [a], Lotta Jakobsson [a,b]

[a] *Chalmers University of Technology, SE-412 96, Gothenburg, Sweden*
[b] *Volvo Car Corporation, SE-405 31, Gothenburg, Sweden*

ABSTRACT

Traffic safety technologies revolve around two principle ideas; crash avoidance and injury mitigation for inevitable crashes. The development of relevant vehicle injury mitigating technologies should consider the interaction of those two technologies, ensuring that the inevitable crashes can be adequately managed by the occupant and vulnerable road user (VRU) protection systems. A step towards that is the accurate description of the expected crashes remaining when crash-avoiding technologies are available in vehicles.

With the overall objective of facilitating the assessment of future traffic safety, this study develops a method for predicting crash configurations when introducing crash-avoiding countermeasures. The predicted crash configurations are one important factor for prioritizing the evaluation and development of future occupant and VRU protection systems.

By using real-world traffic accident data to form the baseline and performing counterfactual model-in-the-loop (MIL) pre-crash simulations, the change in traffic situations (vehicle crashes) provided by vehicles with crash-avoiding technologies can be predicted. The method is built on a novel crash configuration definition, which supports further analysis of the in-crash phase. By clustering and grouping the remaining crashes, a limited number of crash configurations can be identified, still representing and covering the real-world variation.

The developed method was applied using Swedish national- and in-depth accident data related to urban intersections and highway driving, and a conceptual Autonomous Emergency Braking system (AEB) computational model. Based on national crash data analysis, the conflict situations Same-Direction rear-end frontal (SD-ref) representing 53 % of highway vehicle-to-vehicle (v2v) crashes, and Straight Crossing Path (SCP) with 21 % of urban v2v intersection crashes were selected for this study. Pre-crash baselines, for SD-ref (n = 1010) and SCP (n = 4814), were prepared based on in-depth accident data and variations of these. Pre-crash simulations identified the crashes not avoided by the conceptual AEB, and the clustering of these revealed 5 and 52 representative crash configurations for the highway SD-ref and urban intersection SCP conflict situations, respectively, to be used in future crashworthiness studies. The results demonstrated a feasible way of identifying, in a predictive way, relevant crash configurations for in-crash testing of injury prevention capabilities.

1. Introduction

While crash-avoiding systems have shown potential for reducing the number of crashes, crashes will still occur, and there is a need for further enhancement of occupant protection systems. Despite the fact that the effects of crash-avoiding technologies have been investigated in numerous studies, both retrospectively (Cicchino, 2017; Isaksson-Hellman and Lindman, 2016a, 2016b; Kusano and Gabler, 2015; Yue et al.,

2018) and prospectively (Alvarez et al., 2017; Jeppsson et al., 2018; Kovaceva et al., 2020) the results are rarely detailed enough to serve as boundary conditions for future virtual or physical crash testing. Previous studies have targeted the prediction of the combined effects of the crash- and injury mitigation capabilities of future vehicles (Edwards et al., 2014; Lubbe et al., 2018; Östling et al., 2019b), however enhancing protection systems is difficult without access to an accurate description of the expected crash configurations. The crash configuration is defined

* Corresponding author at: Chalmers University of Technology, Hörselgången 4, SAGA Building, Floor 4, SE-417 56, Gothenburg, Sweden.
 E-mail address: alexandros.leledakis@chalmers.se (A. Leledakis).

as the information needed to perform physical or virtual crash testing.

The development of vehicles is a continuous process; therefore, it is important to estimate the effects of traffic safety improvements in a predictive way (Lindman et al., 2017). One of the most significant challenges within this area arises in circumstances when a crash is not avoided despite a crash-avoiding technology intervention. The crash configurations consist of many parameters, such as vehicle speeds, impact location, and relative direction. Multiple parameters are frequently affected during an intervention. Hence, one-dimensional analysis (i.e., when only changes in speed are considered) is inadequate for a thorough vehicle safety assessment in these situations. Identifying the expected crash configurations for cars equipped with crash-avoiding technologies can assist in prioritizing the test-case selection for the in-crash occupant or VRU crash testing. Hence, it is of importance to establish a method for predicting crash configurations.

Crashworthiness test-setups are ultimately targeting to represent real-world crashes. One way of identifying such representative test setups is a statistical clustering of crashes in real-world accident databases. However, the clustering must be performed carefully as the clustering method selected, as well as the parameters used, may considerably affect the final results (Costa et al., 2004; Rodriguez et al., 2019). The need for reproducibility and comparability of results drives the development of automated clustering methods.

With the overall aim of guiding crashworthiness evaluation of future vehicles, the objective of this study was to develop a method for predicting expected and representative crash configurations taking the performance of crash-avoiding technologies into account. A secondary objective involved applying the developed method to some selected operational design domains (ODDs) in order to test and demonstrate the method.

2. The method

This method aims to predict relevant crash configurations by combining knowledge from real-world traffic accident databases and computational models of vehicle crash-avoiding technologies. The

Fig. 1. Overview of the four steps of the method. Step 1: Identification of the traffic safety problem; Step 2: Baseline preparation; Step 3: Treatment simulation; Step 4: Identification of the representative crashes.

method includes four steps, Fig. 1. Initially, the most relevant conflict situations are identified. Crashes resulting from these conflict situations are then selected from an in-depth accident database and used to establish the baseline. Counterfactual pre-crash simulations are then performed, investigating what would have happened if a treatment mechanism, in this case, a crash-avoiding technology, was active in the same situation. The counterfactual treatment simulation results are then compared with the baseline, revealing crashes not avoided. The parameters describing the circumstances of the crash at the time of impact define the crash configurations. These remaining crash configurations are clustered in order to identify the most representative crashes given the crash-avoiding technology studied and the selected ODDs. The four steps are further described in the subsequent sections (Sections 2.1-2.4).

2.1. Step 1: Identification of the traffic safety problem

Traffic safety problems can be retrospectively identified using crash databases. A national or regional crash database typically contains a representative number of crashes that summarize traffic safety challenges. By filtering for the Operational Design Domain (ODD) of the crash-avoiding technology under test, the conflict situations that should be addressed using those system(s) can be identified.

The conflict situation classification, Appendix A, describes the movement of the involved road users in relation to each other before the crash or near-crash event. A conflict situation does not include information on the event circumstances (such as traction lost, driver's distraction, and light conditions) or the crash configuration.

The in-crash phase is greatly affected by the type of road users involved in the crash. Hence, it is essential to categorize crash opponents and analyze each category separately, considering the goal of car crashworthiness evaluation.

2.2. Step 2: baseline preparation

For the identified conflict situations from Step 1 (Section 2.1), in order to predict the safety benefits, baselines are established for the crash-avoiding technology pre-crash simulation. The information on crash details in national accident databases is usually limited or in a format that cannot be directly simulated. Therefore, in-depth accident databases containing details such as trajectories and actions of all involved road users prior to impact can be used. In order to account for uncertainties regarding the precise conditions of a crash, the characteristics of the cases in the original accident dataset are varied in a controlled way, and a large sample of variants, described as synthetic crashes, are created. The collection of the synthetic crashes forms the baseline and help to ensure the robustness of the Advanced Driver Assistant System (ADAS) testing.

2.3. Step 3: treatment simulation

After establishing the baseline, treatment pre-crash simulations can be performed. A prediction can be made about which crashes would be expected to be avoided, mitigated, or not affected by recreating real-world crashes based on the hypothesis that a crash-avoiding technology would be active during the pre-crash sequence.

The performance is evaluated on a complete vehicle level, using computational models that capture the behavior of the involved vehicle subsystems. The subsystems that describe the motion of the vehicle, such as vehicle dynamics and brakes, are typically essential for those evaluations. The inclusion of driver- or road-user behavior (Bärgman et al., 2017) models, along with the systems under test (sensors and crash-avoiding algorithms), can enhance the accuracy of the effectiveness evaluation.

The comparison between the baseline and the treatment simulation output, i.e., the virtual reconstructions of real-world pre-crash situations with and without the crash-avoiding technology, provides a relative
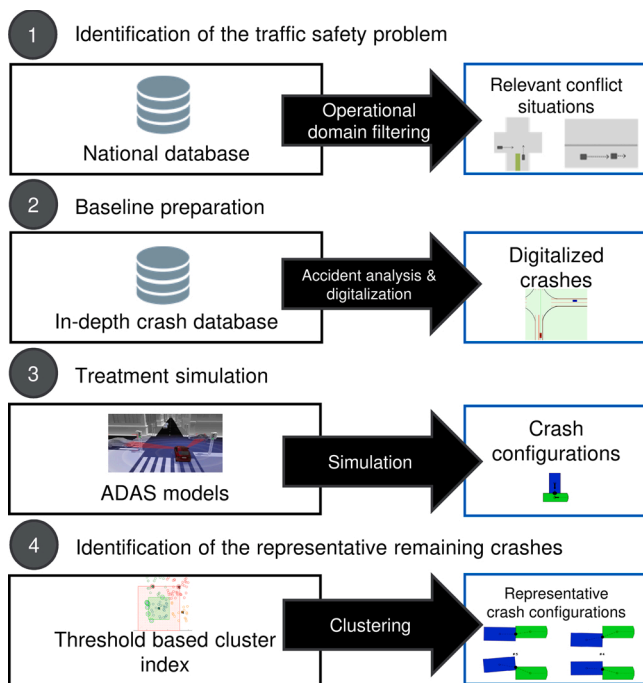
estimation of the effectiveness in that specific situation. The remaining (mitigated and not affected) crashes hold information that can be used to assess and further develop injury prevention systems.

Similar to Step 2 (Section 2.2), many pre-crash parameters may affect the treatment simulation outcome. Those uncertainties can be managed by performing multiple simulations in which alternative assumptions are considered in each setup. The variations include (but are not limited to) which vehicle(s) are equipped with crash-avoiding technology or even specific parameters describing the performance of each system, such as the braking response or sensor detection range.

2.3.1. Crash configuration

The simulation output of interest for this method's purpose is the crash configuration from the remaining crashes. The crash configuration is defined as the necessary parameters needed to set up a virtual or physical crash test, defined in detail by Wågström et al. (2019). It is a detailed, yet compact, description of the circumstances of the crash at the time of impact, using a set of three impact angles and two speeds, Fig. 2. The Host Collision Point Angle (HCPA) describes the first contact point on the host vehicle in terms of a counter-clockwise angle around the geometrical center of the vehicle using the vehicle's centerline as the reference. Likewise, the Opponent Collision Point Angle (OCPA) describes the first contact point on the opponent vehicle. The Opponent Yaw Angle (OYA) expresses the heading angle of the opponent's vehicle in the horizontal plane relative to the heading angle of the host vehicle. The two speeds complete the crash configuration definition. An example of a Straight Crossing Path (SCP) crash configuration is illustrated in Fig. 2.

In order to ensure comparability between vehicles of different dimensions, the vehicles' dimensions are normalized to a square unit car before calculating the angles. After the normalization, the collision angles on the corners of the vehicle are classified with the same numerical value, even for vehicles with dissimilar width-to-length ratio. The normalization, Fig. 3, is performed by scaling the impact point on the vehicle's reference system, according to Eqn (1) and (2).

$$x_n = x_0/VehicleLength \qquad (1)$$

$$y_n = y_0/VehicleWidth \qquad (2)$$
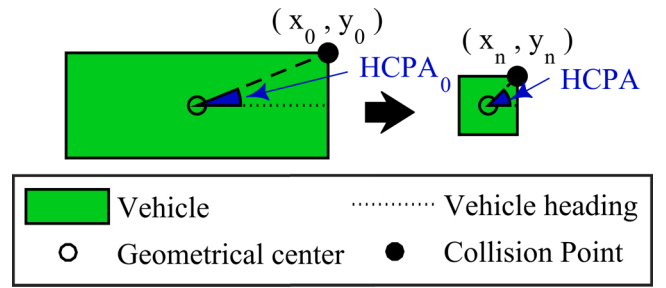
Fig. 3. Vehicle size normalization. The vehicle is scaled to a unit-size, allowing for comparison of crash configurations between vehicles of different dimensions. The first point of contact (x_0,y_0) is scaled to generate the normalized coordinates (x_n, y_n), which, in return, defines the HCPA relative to the vehicle's heading.

Crash configurations of multiple conflict situations can be plotted in the same graph (Fig. 4), creating a crash configuration map that provides a perspective of the complexity of vehicle crashes.

2.3.2. Crash configuration duality

An aspect to consider for v2v crashes is that the occupant protection assessment can be performed separately for each vehicle and can, therefore, be considered as two different crash configurations. In case the host and opponent vehicles are of the same type, the two crash configurations can be merged in the same dataset.

An SCP crash is used as an example to visualize the two possible crash configurations that can be considered, Fig. 5. The numerical relation of the two crash configurations is expressed in Eqn (3), (4) and (5).

$$HCPA_{dual} = OCPA_{original} \qquad (3)$$

$$OCPA_{dual} = HCPA_{original} \qquad (4)$$

$$OYA_{dual} = -OYA_{original} \qquad (5)$$

2.4. Step 4: Identification of the representative remaining crashes

The output of the crash-avoiding technology simulations in terms of crash configurations for remaining crashes can be directly used for the enhancement and evaluation of occupant protection. However, when a large number remains, it would be difficult to account for the computational needs of in-crash simulations in relation to resource limitations. Clustering is a possible solution to this challenge, as it can be used to group similar crash configurations into a limited number of crash configurations suitable for representing the real-world variation.

The clustering algorithm suggested for this method applies empirical knowledge of the acceptable domain-specific cluster spread and automates the process of cluster number selection.

Clustering is performed in two stages. Initially, the geometric parameters of the crash configurations (HCPA, OCPA, and OYA) are clustered. In the second stage for every identified cluster, clustering is performed again in order to select the most representative impact speeds. Clustering in two stages eliminates the need for data normalization since the variables to be clustered in each stage take values from the same range and express similar physical quantities. For both stages of the clustering, the k-medoid clustering algorithm, as described by (Kaufman and Rousseuw, 2005), is selected because of its attribute of using existing data points for the cluster centers. This attribute is essential for this method considering that some HCPA, OCPA, and OYA combinations could potentially result in overlapping vehicles (with more than one intersecting point) at the time of impact, which would make those crash configurations invalid, as previously seen in Wågström et al. (2019).
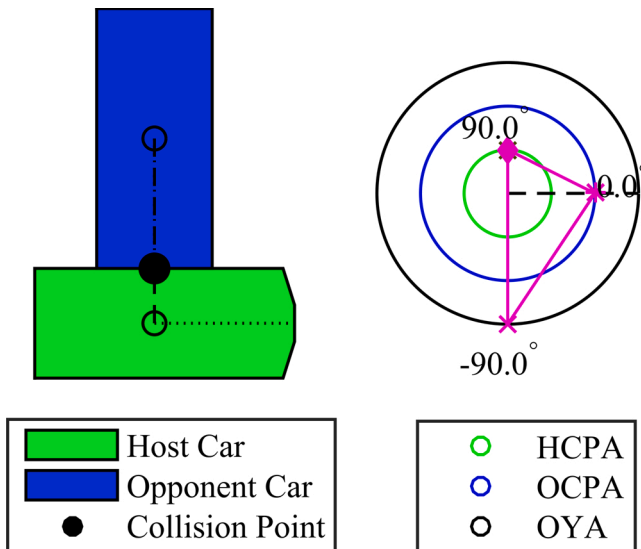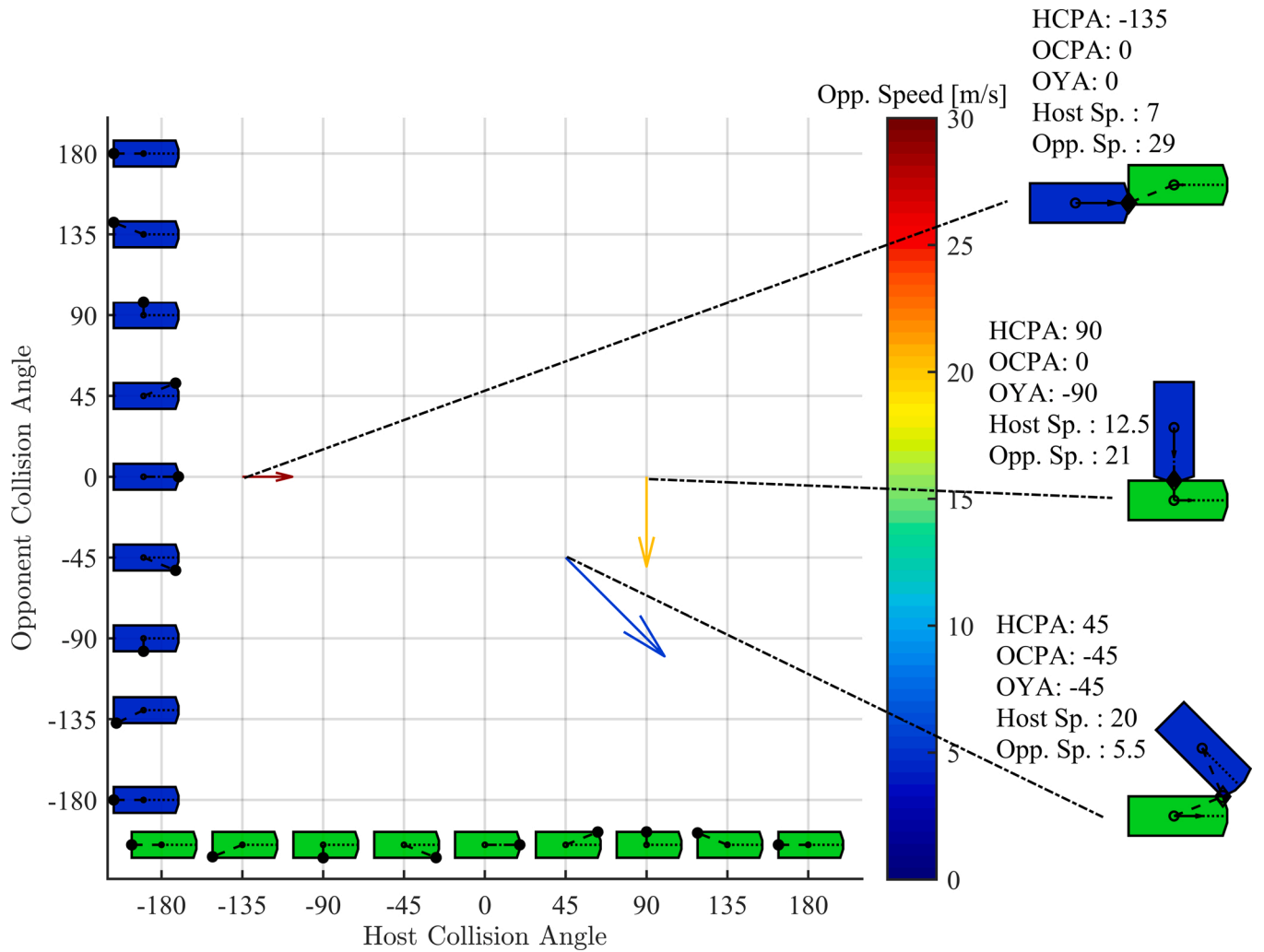

Fig. 2. Example from a Straight Crossing Path (SCP) crash: The host car is impacted on the left side from the perpendicularly moving opponent vehicle. The crash configuration is visualized using three concentric circles; one for every angle of the crash configuration definition (HCPA = 90°; OCPA = 0°; OYA=-90°).

Fig. 4. Example of a Crash configuration map, where multiple crash configurations can be plotted on the same graph allowing comparisons and providing an overview of the crashes in a dataset. Every crash configuration is represented by an arrow. The origin of the arrow is placed in (HCPA, OCPA), and its direction aligns with OYA. The host and opponent vehicle speeds can be visualized by scaling and coloring the arrow.
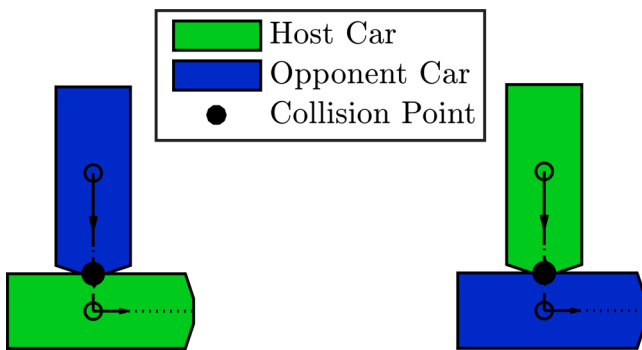
Fig. 5. The same crash configuration from two different perspectives. Crashworthiness is evaluated for the Host Car. A side impact (on the left) can also be considered as a frontal impact (on the right).

The number of clusters is selected using a heuristic optimization algorithm, where the cluster validity index is calculated for different numbers of clusters in order to select the best value. Three cluster validity indices (CVi); Silhouette, Davies-Bouldin, and Calinski-Harabasz, as described by (Arbelaitz et al., 2013) were tested, but neither was selected for the purpose of this method as they were found to be sensitive

to the number of observations and did not consider the absolute data scaling (see Appendix B). Instead, a cluster validity index, hereafter called Threshold Based Cluster index (TBCi), was developed and used in the optimization process. It uses thresholds for cohesion and separation to characterize the quality of the clusters. The thresholds set in the cohesion part of the TBCi index represent the maximum desired spread of the observations assigned to the same cluster. The thresholds set for the separation define the minimum desirable distance of two cluster centers. A detailed explanation of TBCi, as well as the relevant equations, are included in Appendix B. Maximizing the objective function, in this case, TBCi, will yield the number of clusters that best match the selected criteria.

When clustering data containing angles in the $[-\pi, \pi]$ range, special care should be given to ensure that the discontinuity between $-\pi$ and π is resolved. The approach suggested in this method is to perform unit-circle decomposition, where all the angles are replaced with their co-ordinates on a unit circle. That ensures that angles around π are numerically similar.

3. Applying the method

In a feasibility study, the relevant crash configurations for occupant safety testing were investigated to demonstrate the method. A conceptual Autonomous Emergency Braking (AEB) system, representing a

generic crash-avoiding technology, was applied using Swedish data as the baseline. Two ODDs, highway driving, and urban intersections were considered. The ODDs were pre-selected by the OSCCAR[1] consortium and were in-line with the aim of the project, where harmonization analysis methods were applied.

3.1. Step 1: Identification of the traffic safety problem

Regardless of injury severity, passenger car crashes from the Swedish Traffic Accident Data Acquisition (STRADA) database (Howard and Linder, 2014) were included in the analysis. The analysis identified all conflict situations in the ODDs where the AEB system under test was supposed to work. In an attempt to isolate the effects of infrastructure and legislation changes, only crashes between 2010 and 2017 with at least one car were included in the filtering. Two example ODDs: urban intersections (n = 10949) and highway driving (n = 5008) were used to filter crashes from STRADA by using the variables that can be seen in Table 1. Slippery road conditions (snow and ice) were excluded. The conflict situations from STRADA were classified using the *AccidentTypeCode* variable when possible. As it was not possible to automatically classify all conflict situations, a review of the text description of parts of the dataset was performed, see Appendix A.

The analysis of the highway crashes showed that Same-Direction (SD) crashes were the most common conflict situation representing 53 % of the highway crashes, Fig. 6. The urban intersection analysis revealed a more diverse dataset, in which crashes between cars and VRUs were frequent (40 %). The focus of this study was v2v crashes, in which SD and SCP accounted for 35 % and 21 %, respectively, Fig. 6.

3.2. Step 2: baseline preparation

The Volvo Cars Traffic Accident Database (VCTAD) was used for preparing the baseline for the upcoming pre-crash simulations. The VCTAD contains crashes involving Volvo cars in Sweden (Isaksson-Hellman and Norin, 2005). For this study, crashes between 2007 and 2017 were included. Crashes involving the conflict situations identified in step 1: Same-Direction rear-end frontal (SD-ref) in highways, and Straight Crossing Path (SCP) in urban intersections were selected. Again, slippery road conditions (ice, snow) were excluded in order to match with the selected ODD. Specifically, for the highway driving sub-set, crashes that occurred on roads without median separation were excluded for the same reason. Car vs. car and car vs. heavy vehicle were analyzed in two separate categories.

The pre-crash phase of each case was digitized using an in-house crash digitization tool. By this, information on pre-crash vehicle and driver behavior, road geometries, and sight obstructions were converted in a numerical, time history data format for each case.

When compensating for uncertainties of the reconstructed crashes, synthetic cases were generated where the vehicle impact speeds, position in the lane, and braking behavior were altered. Impact speeds were varied uniformly in the interval $[V_{min}, V_{max}]$. The interval was calculated using a speed-dependent function, as shown in Eqn (6), (7) and (8). The speed distributions of the original digitized, and the synthetic crashes of the SCP (car vs. car) crashes, can be seen in Fig. 7.

$$SpeedChangeFactor = 0.1 + 0.9e^{-V_{crash}/20} \tag{6}$$

$$V_{max} = (1 + SpeedChangeFactor)*V_{crash} \tag{7}$$

$$V_{min} = (1 - SpeedChangeFactor)*V_{crash} \tag{8}$$

For the SCP crashes, the accuracy of the witness' information on driver braking maneuvers in the database was considered low; therefore, synthetic cases with multiple brake profiles were generated. If the

vehicle was originally coded as braking in the database, then it was assigned an 80 % chance of braking in the synthetic crash; otherwise, the probability of braking was set to 20 %. The maximum braking level and duration were varied, according to Eqn (9), (10) and (11). The initial speed was adapted in order to preserve the coded impact speed, and the initial speed was also maintained in the range that is described by Eq. (6)–(8). Additionally, the speed reduction was always maintained within ±50 % of the originally coded speed reduction if the crash was originally digitized as braking. An example of how the synthetic brake profiles were generated can be found in Appendix A.

$$P_{Brake} = \begin{cases} 0.8, \ A.(\textit{if originally digitized as } \text{braking}) \\ 0.2, \ B. \ (\textit{if originally digitized as } \text{not braking}) \end{cases} \tag{9}$$

$$Brake_{Level} = \begin{cases} min\big(U(0.5, 1.5) * Brake_{Level_{original}}, \mu * g\big), \ A. \\ U(0, \mu * g), \ B. \end{cases} \tag{10}$$

$$Brake_{duration} = \begin{cases} U(0.5, 1.5) * Brake_{duration_{original}}, \ A. \\ U(0, 3), \ B. \end{cases} \tag{11}$$

Finally, for the SD crashes, the lateral offset relative to lane centerline was varied using a uniform distribution $[-0.3m, \ 0.3m]$ around the coded trajectory for both vehicles.

The process of creating synthetic variants from the originally digitized crashes resulted in 1010 (based on 36 original crashes) and 4814 (163 original crashes) synthetic crashes for SD and SCP, respectively. For every synthetic crash case, a pre-crash kinematic sequence of 15 s, followed by a 5-second post-crash extrapolation, was generated using a sampling rate of 1 kHz. The post-crash extrapolation is generated assuming that both vehicles would continue along their desired path, as-if a crash never occurred. That is necessary for counterfactual simulations in order to ignore the original outcome and estimate the potential benefits of a crash-avoiding technology.

3.3. Step 3: treatment simulation

In order to forecast crashes not avoided by the conceptual AEB intervention, the baseline crashes generated in Step 2 (Section 3.2) were simulated with and without the addition of the treatment technology. The simulations were performed in a planar (2-D) environment, which was sampled at 200 Hz. The conceptual AEB was simulated using a MIL approach where the vehicle(s) equipped with AEB followed the prescribed path from the reconstructed crash until the AEB triggering criteria were met, and the system issued a brake intervention. During the intervention, the driver's reaction was not considered, and it was assumed that the vehicle would continue along the same path.

The vehicle sensory system consisted of a conceptual LIDAR sensor, responsible for the detection as well as the classification of the target, which was implemented using vision rays, as illustrated in Fig. 8. The target was detected at the intersection with the vision ray if it was unobstructed in the field of vision (FoV). As a simplified target classification model, a 150 ms time-window was used, during which the target had to remain unobstructed in the FoV in order to be classified. The LIDAR was positioned on the centerline of the vehicle, 2 m behind the front bumper. It sampled the environment, using a vertical resolution of $0.1°$, at 40 Hz, and had a range of 200 m with 180 °FoV.

The AEB calculated Time-to-collision (TTC) using the relative velocity vector and the distance to all detected points of the classified target. When the TTC of any detection point was below the threshold of 1.2 s, an AEB was triggered, requesting maximum deceleration of the vehicle. The brake subsystem was modeled using a constant brake delay of $200ms$ and a brake gradient of $35m/s^3$, while the maximum brake level was determined by the friction coefficient (μ) of the reconstructed crash, which was set to 0.8g for dry/moist tarmac and 0.7g for wet tarmac. The settings used for the AEB simulations can be found in Appendix C. The impact detection was performed using an intersection search between the vehicle outlines. The simulation was terminated if an

Table 1
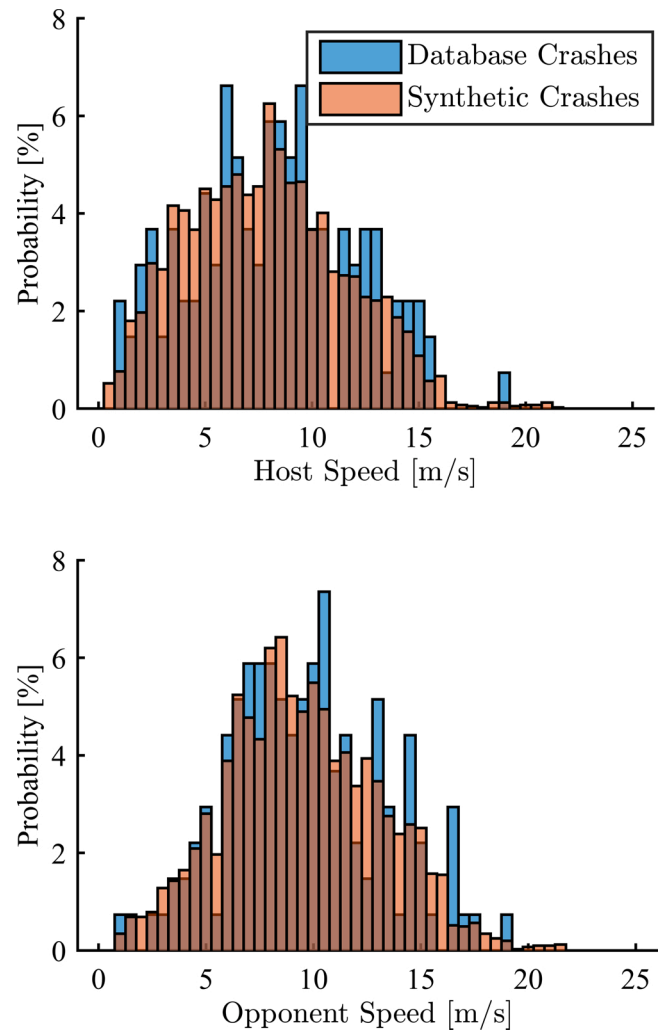Inclusion criteria for STRADA (n = 228408) filtering.

Variable	Highway driving		Urban intersections	
AccDate	20[2010,2017]	n = 92839, 40 %	[2010,2017]	n = 92839, 40 %
TrafficEnvironment	Not densely populated area (2)	n = 42873, 19 %	Densely populated area (1)	n = 43588, 19 %
PlaceType	Intersection (1), Straight road (2)	n = 40812, 18 %	Intersection (1)	n = 18022, 8 %
TypeOfRoadA	Motorway (1)	n = 7197, 3 %	All	n = 18022, 8 %
SpeedRestrictionA	60,70,80,90,100,110,120	n = 6636, 3 %	20,30,40,50	n = 13184, 6 %
RoadState	Dry road (1), Wet road (2)	n = 5008, 2 %	Dry road (1), Wet road (2)	n = 10949, 5 %
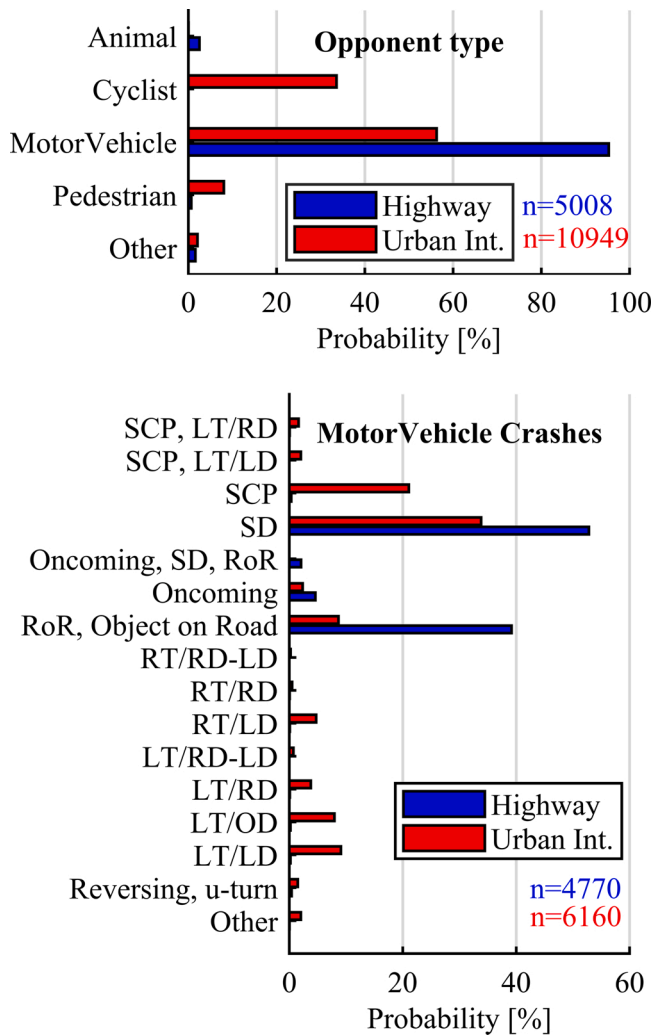
Fig. 6. Highway and urban intersection accident statistics from STRADA. On top, motor vehicle crashes make up approximately 95 % of the highway crashes, while the equivalent number for urban intersection is 56 %. Same-Direction (SD) represents 52 % of all highway crashes between motor vehicles. SD and Straight Crossing Path (SCP) account for 35 % and 21 %, respectively, of the urban intersection crashes between motor vehicles. The abbreviations for and detailed description of the conflict situations can be found in Appendix A.


Fig. 7. Impact speed of the host and opponent vehicles for the original digitized crashes (n = 136) and the synthetically generated crashes (n = 4064).

impact was detected, and the centroid of the vehicles' overlapping area was established as the first impact point.

For the SCP conflict situation, all possible combinations of the two vehicles with or without AEB were simulated. More specifically, that resulted in three distinct treatment simulations; one simulation was performed with AEB enabled in the host vehicle, a second simulation was performed with AEB enabled in the opponent vehicle, and a third simulation was performed with AEB enabled in both vehicles. On the other hand, for the SD conflict situation, only one simulation was

performed, whereby the vehicle following was equipped with the AEB since the concept AEB described above is not designed to intervene when the lead vehicle is about to be impacted from behind.

The output of the pre-crash simulations was used to calculate the collision avoidance rate, Fig. 9. For almost all simulations, there was an intervention that altered the crash configuration to some extent.

Specifically, for the SCP conflict situation, the simulation results suggested that the outcome may be substantially affected by which vehicle(s) (host, opponent, or both) is equipped with the conceptual AEB. Since all cases in the baseline tested could represent real-world traffic situations, the results were merged and further analyzed as a single dataset, assuming the risk of each crash situation is equal. As expected, the avoidance rate is highest when both vehicles are equipped with the conceptual AEB, reaching up to 92 % avoidance. For SCP car to
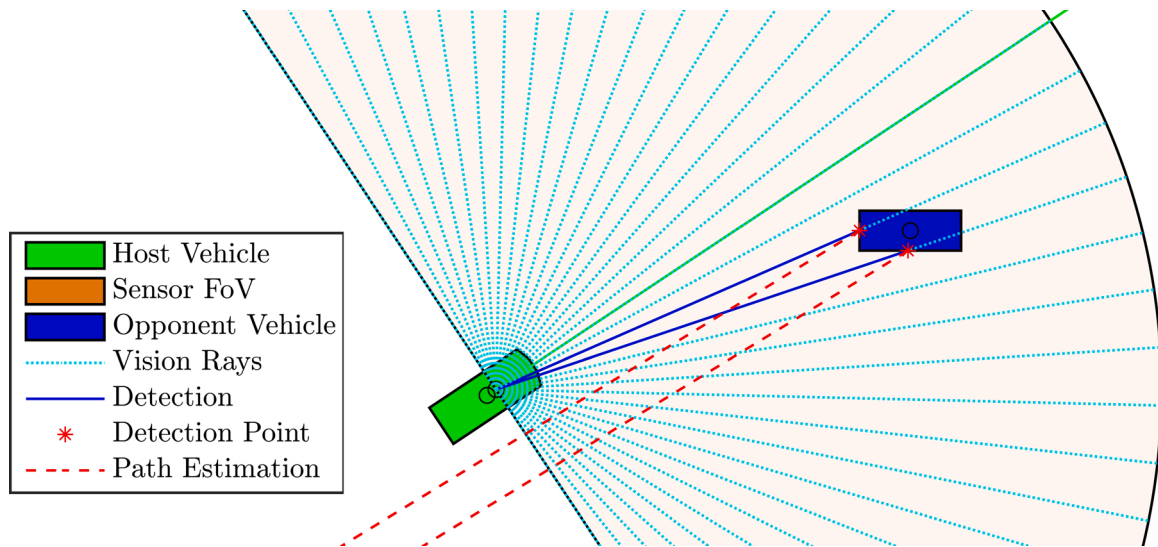
Fig. 8. Illustration of the conceptual LIDAR sensor functionality. When the vision rays detect the target, the relative speed vector is projected from the detection point(s). If the relative speed vector(s) intersects with the host vehicle, a collision is expected, and TTC is calculated.
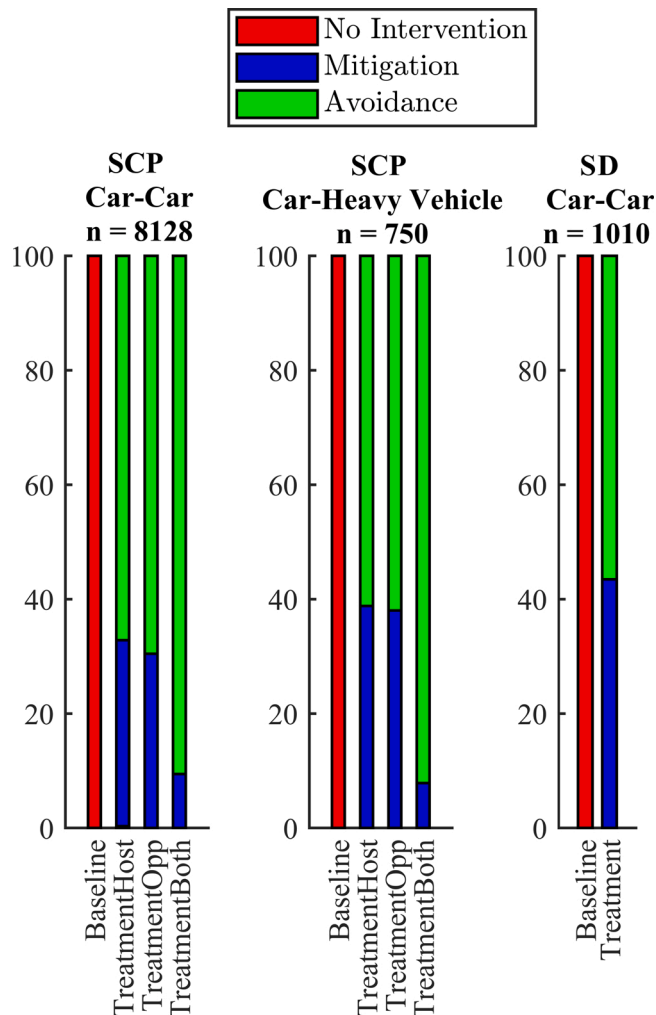
Fig. 9. The outcome of the SD-ref and SCP conflict situations for all possible combinations of vehicles equipped with AEB. For the urban SCP crashes, the avoidance rate varied between 61 % and 92 %, depending on which vehicle(s) were equipped with AEB functionality. Additionally, 56 % of the highway SD crashes were avoided.

car crashes, a crashworthiness evaluation of the host and opponent vehicle were considered equally important. Therefore, both crash configurations were merged in the same dataset for further analysis, doubling the sample size.

The results suggest that the number of car crashes within the baseline sample for this study will decrease (Fig. 9) and that the remaining of the SCP and SD crashes would have altered crash configurations.

3.4. Step 4: Identification of the representative remaining crashes

The pre-crash simulation output in terms of crash configurations for the remaining crashes was clustered with the k-medoid algorithm, using the TBCi criteria to select the number of clusters. The parameters selected for the TBCi can be seen in Table 2, and the sensitivity of the clustering parameters was investigated and is presented in Appendix B.

Applying the method, predicted 5 (car vs. car) and 52 (36 car vs. car and 16 car vs. heavy vehicle) crash configurations for SD-ref and SCP, respectively. The crash configuration map (as defined in Section 2.3.1) of the remaining urban intersections can be seen in Fig. 10. The crash configuration maps of the SD-ref crashes are available in Appendix D, providing an overview of all pre-crash simulated crashes and the predicted crash configurations. The results predicted that for urban intersections crashes, the introduction of a crash avoidance technology could shift the impact point towards the vehicles' corners.

As an example, the crash configuration of the predicted most frequent remaining urban intersection crash is shown in Fig. 11. All clusters are available in Appendix E, where the predicted remaining crash configurations are presented.
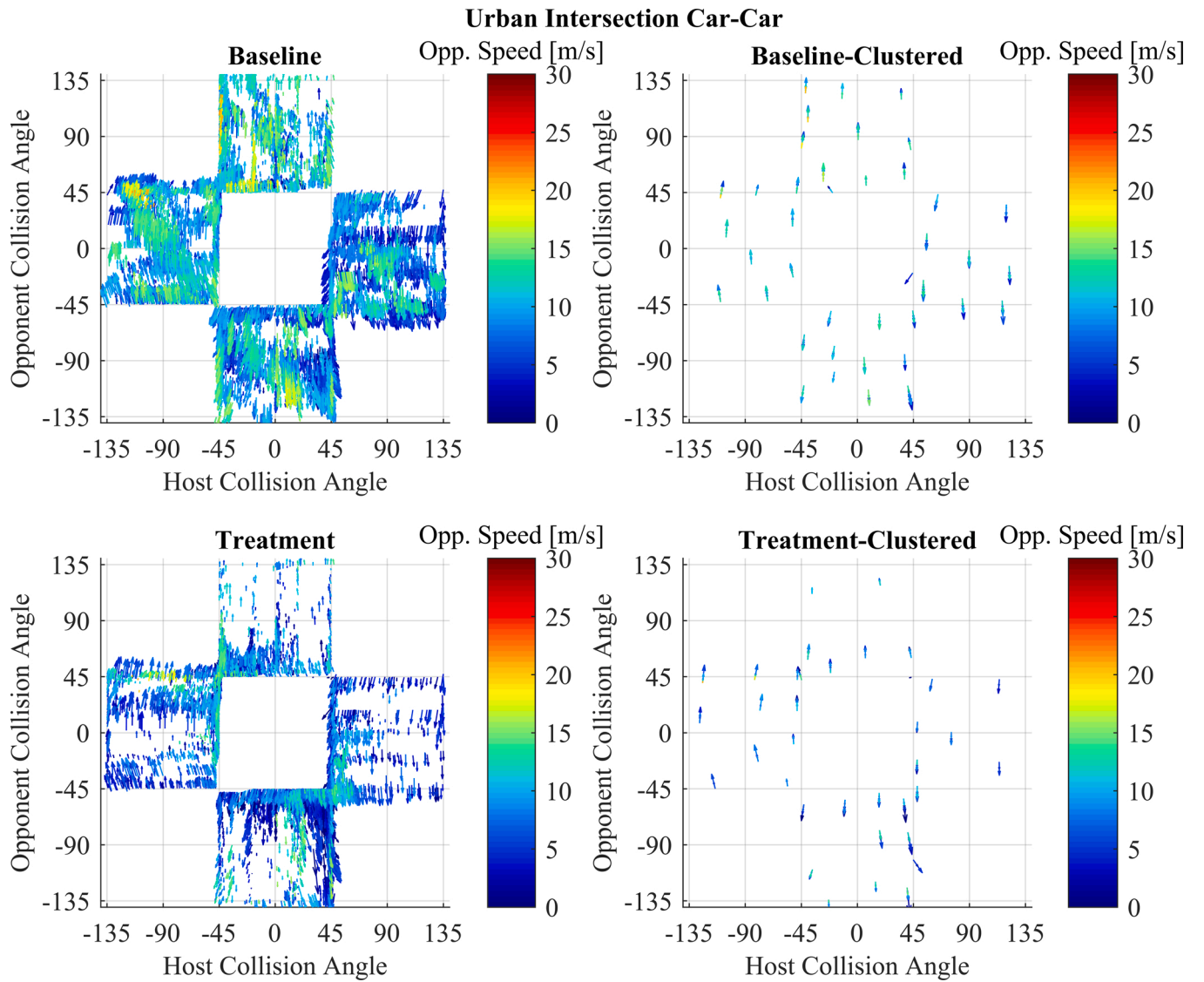
4. Discussion

The method presented in this study derives predicted crash

Table 2
TBCi parameters.

	Stage I: Angle			Stage II: Speed	
	HCPA	OCPA	OYA	Host	Opponent
Th^{coh}	10°			2.5m/s	
Th^{sep}	20°			5m/s	
W_{coh}	−1				
W_{sep}	0.5				

Fig. 10. Crash configuration maps of the baseline and remaining urban intersection crashes. On the left, the crash configuration of the remaining urban intersection crashes can be seen. On the right, the most representative crash configurations, identified by the clustering process, are visualized. The impact speed of the remaining crashes after braking interventions is reduced (illustrated by shorter/blue arrows). Furthermore, the majority of the remaining crashes are distributed around the vehicle corners (Host/Opponent collision angle ±45°).

configurations for vehicles equipped with crash-avoiding technologies, based on real-world data and MIL simulations. The identified crash configurations can be used to support the development of occupant protection systems in a traffic environment with crash-avoiding technology equipped vehicles. The crash configurations serve as a link between collision avoidance performance estimations and occupant protection performance evaluation. They can be considered part of a larger toolchain, providing useful insights for holistic vehicle safety assessment and enhancement. The need to predict the applicable crash configurations is highlighted by the conceptual AEB simulation results since, for most of the simulations, there was an intervention that altered the crash configuration to some extent.

The strength of this method is that it takes models of crash-avoiding technology and situations in real-world crash databases into account in order to predict the expected crash configurations. In addition to previous studies, this method facilitates the evaluation of crash-avoiding technology by including computational models for enabling the transition of the pre-crash simulation results to in-crash injury evaluations. The clustering method that was developed and used in this study is able to maintain the diversity of the expected crashes, and in that way,

enhance the robustness in the development of protection systems.

The conceptual AEB was designed for longitudinal interventions and applied to study the car-to vehicle conflict situations SCP and SD-ref. The tested conflict situations were chosen based on the statistical analysis of pre-selected ODDs. As previously shown by Sander and Lubbe (2018a), SCP represents one of the most challenging conflict situations with many pre-crash parameters affecting the ADAS performance, as well as many parameters that are needed to describe the in-crash response. For SD crashes, the in-crash response is typically easier to predict, and the variations mainly concern vehicle speeds. This shows that the method can address diverse conflict situations. Similarly, simulations could be performed with various crash-avoiding technologies (more simplified or more advanced; see, e.g. (Rothoff et al., 2019)) and multiple road user types, such as cars, heavy vehicles, and VRUs. The method can be considered modular, and each one of the steps described can be adjusted to answer a broad range of questions and to account for diverse aspects of vehicle safety. The tested ODDs were considered sufficient as a proof of concept, although it would be interesting to apply the method further in more conflict situations to obtain a holistic understanding of the crash configuration distributions and cover a larger
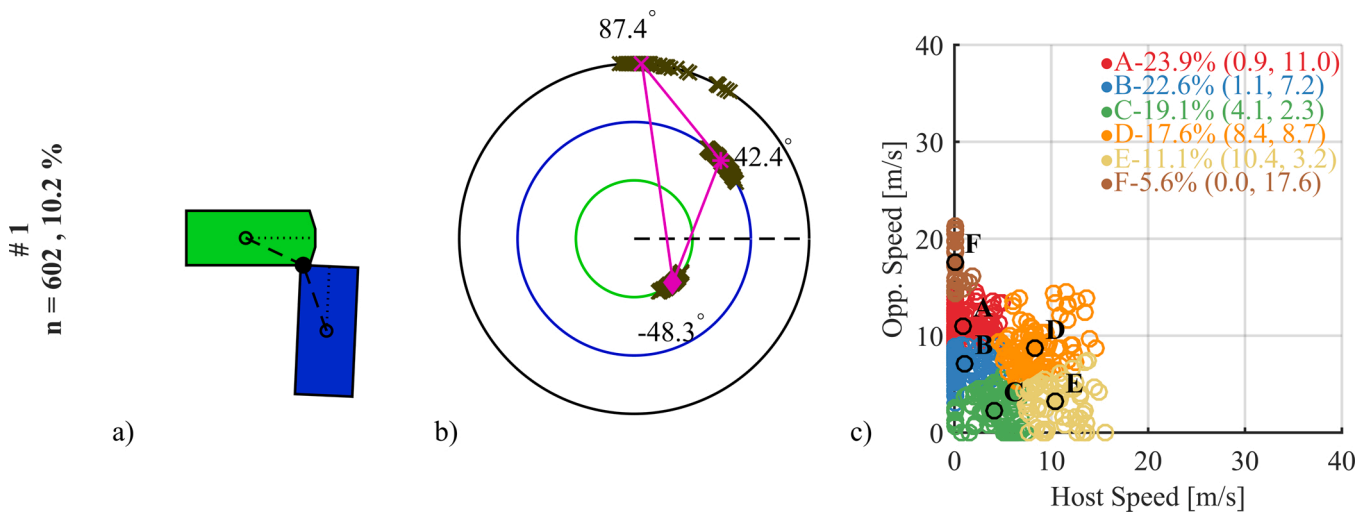
Fig. 11. Example of the first and second stage clustering results for the most representative crash configuration; a) a graphical representation b) visualization using three concentric circles (green=Host Collision Angle, blue = Opponent Collision Angle, and brown = Opponent Yaw Angle), including the variation in the data depicted by crosses and c) plots of vehicle speeds for the clusters with medoids specified in the legend.

share of crashes that were excluded because of the ODD selection. The ODDs that were used in this study covered approximately 7 % of the STRADA database (see Table 1). Additionally, a similar method has previously been used on a different crash database (Östling et al., 2019a) as part of the OSCCAR project collaboration.

Pre-crash simulation techniques might need further development in order to be able to confidently predict the performance of certain types of crash-avoiding technologies (for example, precautionary safety mechanisms as in automated driving concepts, (Rothoff et al., 2019)). Alternative methods for simulating crash-avoiding technology benefits could also be applied, i.e., multi-agent traffic simulations (Kitajima et al., 2019), which would allow the estimation of road infrastructure changes. Counterfactual analysis methods act as a link between what has happened in the past and what could be expected in the future by investigating the effects of an intervention under the same conditions. Likewise, including market penetration estimations (Sander and Lubbe, 2018b) and user acceptance information of the crash-avoiding technology could improve the accuracy of crash configuration distribution predictions.

As with any simulation-based study, it is crucial to highlight and understand the impact of the assumptions made. In this study, a conceptual AEB was used alongside ideal sensors, combined with simplified brake and vehicle dynamics models. The ideal sensor used was not affected by environmental conditions, which could alter object detection and, consequently, in some crash situations, the AEB performance. The simplified brake and vehicle dynamics models, on the other hand, are considered to be adequate for the selected ODDs since highly non-linear effects originating from slippery road conditions were not present. In addition, the 2-D planar simulation environment could also affect the sensor performance by not considering the effects of pitch-roll motions on sensor performance.

The treatment simulations were performed based on the assumption that the driver does not react during an ADAS intervention. Even though it is believed that this is a conservative assumption, it could also be argued that some drivers might override the intervention in certain situations and thereby reduce the safety benefits of the ADAS systems. In order to cover for uncertainties, it would be advantageous to include quantitative driver behavior models (Engström et al., 2018) in the simulations.

When applying this method for assessing future vehicles, it is important to consider the inaccuracies that can be introduced because of the used modeling methods and input data (Wimmer et al., 2019). The process of digitizing real-world crashes is critical to ensure the

representability of the simulation results. Generating synthetic crashes from real-world crashes can account for uncertainties; hence the variables and the methods selected for creating the variations must be carefully selected. Comparing the distributions of databases crashes with synthetic crashes acts as quality control, ensuring that the connection to the real-world is maintained. Considering the validity of the database variables plays a key role in the process of generating synthetic cases. Variables that are considered to be captured accurately in the crash database, such as impact speeds, should maintain a similar distribution when comparing datasets with and without synthetic cases. In contrast, variables that have higher uncertainty should be varied to examine their potential influence. As an example, the braking profile of the vehicles involved in the urban intersection crashes was varied while the initial speeds and speed reduction distribution of the vehicles were maintained in a reasonable range, Appendix A - Fig. A.12. In the future, crash databases containing representative samples of crashes with objective data (such as logged data from Event Data Recorder (EDR) devices) can give valuable information on parameters like these, and using statistical models assist in the generation of higher fidelity synthetic data when needed.

In the event where a collision is not avoided despite an ADAS intervention, there is a need to link the pre-crash kinematics with the in-crash boundary conditions. That link is established using the crash configuration definition proposed in this study. This definition was shown to be useful for accurately describing the impact at the time of the first contact, using a limited number of variables. However, it might be sensitive to objects with long straight edges, for which similar crash configurations can be described using different numerical values. This challenge could be resolved by using more realistic car shapes (compared to the "box-cars" used in this study) in future pre-crash simulations.

Retrospective studies (Cicchino and Zuby, 2019; Isaksson-Hellman and Lindman, 2016a, 2016b) have found that during SD crashes, when the striking vehicle is equipped with AEB, it is more likely to have an impact closer to the vehicle's corner. That was partially associated with the AEB system design (e.g., designed to disengage if the vehicle is turning) and could also be connected to the system's effectiveness during more complex kinematic scenarios. A 4.5 % increase in corner impacts was observed in this study compared to 2.8 % in (Cicchino and Zuby, 2019) based on crash data and 12.5 % in (Isaksson-Hellman and Lindman, 2016a, 2016b) based on insurance claims data. Likewise, for SCP crashes, the impact points were shifted towards the vehicles' corners (Fig. 10). The corner-to-corner impacts were overrepresented in the

remaining crash configurations, which could be grouped in larger clusters compared to the baseline. The seven most frequent (#1-#7) clusters of the treatment simulations described corner-to-corner impacts and covered 47 % of all predicted remaining crashes. In contrast, the seven most frequent baseline clusters accounted for 31 % of all crashes and described front-to-side impacts. Those findings could have implications for the relevance of existing regulatory and assessment crash test protocols (such as the side barrier FMVSS214 and UN R95), which are targeting front-to-side impacts. The predicted remaining crash-configurations could be less severe in terms of intrusions to the passenger compartment, but corner-corner impacts could pose challenges for occupant retention.

Selecting a clustering algorithm such as the k-medoid, has the built-in assumption that an "average observation" can represent all observations of the cluster. While using a worst-case scenario might be considered as a more rational approach, it can be argued that it is not always possible to identify the "worst-case". Selecting TBCi parameters that discourage clusters to span across "large" intervals and therefore maintaining representativity is a viable option since the TBCi parameters will affect the clustering results (Appendix B). The acceptable cluster spread is related to the robustness of the vehicle's structural performance and should be considered when selecting the parameters. The vehicle response will also be affected by the type and properties of the opponent vehicle. Using representative opponent vehicles is an important next step when generating crash-response corresponding to real-world crashes.

The clustering results may produce more test cases than it is feasible to evaluate using virtual/physical crash tests considering testing capacity. Should such circumstances occur, expert judgment is recommended for prioritizing the selection of test cases for the development and assessment of safety systems. Increasing the clustering threshold is not advisable since it could result in clusters with reduced representativity, Appendix B - Fig. B.4.

The multidimensionality of crash configurations makes clustering a challenge when encountering limited samples. This challenge is expected to grow in the future as the overall number of crashes is expected to be reduced over time. In cases where the crash occurs between the same type of vehicle, considering both vehicles for the in-crash safety evaluation can be beneficial since it can effectively double the sample size of the database.

The pre-crash phase was not considered in the clustering stage of this study. By using the pre-crash kinematics in occupant crash simulations, the injury prediction relevance (Östh et al., 2012), could improve by including occupant movement during the pre-crash phase. The pre-crash vehicle kinematics could be described with parametric equations and included in the second stage of the clustering process of future studies.

Furthermore, the k-medoid algorithm considers Euclidian distance to cluster the data, which differs from the TBCi criteria since the k-medoid algorithm is generating hyper-spheres, while TBCi is considering hyper-rectangles. Using a customized clustering method integrated with the TBCi may enhance the clustering performance, although, for this study, the clustering results were considered adequate.

Finally, the expected crash configurations provide only part of the necessary information needed for the assessment and development of future vehicles. Combining that information with seating positions, sitting postures, and diverse human anthropometries, assists the task of performing an overall safety assessment and enables the development of safe automated driving vehicles.

5. Conclusions

This study presents a method for predicting representative crash configurations in vehicles with crash-avoiding technologies to account for the influence of those technologies in the crashworthiness evaluation. In a feasibility study, the influence of a conceptual AEB system on crash configuration distributions was predicted using treatment pre-

crash MIL simulations applied to real-world data baselines. The treatment simulations demonstrated the difference between the expected future crashes compared to the crashes from available real-world databases. The results suggest that the large number of crashes not avoided calls for further enhancement of occupant protection measures. Specifically, for straight crossing path crashes, the conceptual AEB system was found to shift many of the crashes closer to the vehicle's corners. The need for new test setups for occupant in-crash protection evaluation is obvious.

The results showed that thousands of future crashes could be represented by a reduced number of clusters. This demonstrates the ability of the proposed method to limit the testing efforts originating from a magnitude of diverse crashes into a manageable amount of well-defined test-cases for crashworthiness while maintaining the representativity and the diversity of real-world crashes.

Declaration of Competing Interest

Leledakis, Lindman, Östh, Wågström, and Jakobsson are employees of Volvo Car Corporation.

CRediT authorship contribution statement

Alexandros Leledakis: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Magdalena Lindman:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Jonas Östh:** Supervision, Writing - review & editing. **Linus Wågström:** Conceptualization, Methodology, Writing - review & editing. **Johan Davidsson:** Supervision, Writing - review & editing. **Lotta Jakobsson:** Supervision, Writing - review & editing, Project administration, Funding acquisition.

Appendix A. [{(Appendix A–E}] Supplementary data)

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2020.105932.

References

Alvarez, S., Page, Y., Sander, U., Fahrenkrog, F., Helmer, T., Jung, O., Hermitte, T., Düering, M., Döering, S., Op den Camp, O., 2017. Prospective effectiveness assessment of ADAS and active safety systems via virtual simulation: a review of the current practices. Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Paper No 17-0346.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recognit. 46 (1), 243–256. https://doi.org/10.1016/j.patcog.2012.07.021.

Bärgman, J., Boda, C.N., Dozza, M., 2017. Counterfactual simulations applied to SHRP2 crashes: the effect of driver behavior models on safety benefit estimations of intelligent safety systems. Accid. Anal. Prev. 102, 165–180. https://doi.org/10.1016/j.aap.2017.03.003.

Cicchino, J.B., 2017. Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. Accid. Anal. Prev. 99, 142–152. https://doi.org/10.1016/j.aap.2016.11.009. A.

Cicchino, J.B., Zuby, D.S., 2019. Characteristics of rear-end crashes involving passenger vehicles with automatic emergency braking. Traffic Inj. Prev. 20 (sup1), 112–118. https://doi.org/10.1080/15389588.2019.1576172.

Costa, I.G., de Carvalho, F., de, A.T., de Souto, M.C.P., 2004. Comparative analysis of clustering methods for gene expression time course data. Genet. Mol. Biol. 27 (4), 623–631. https://doi.org/10.1590/S1415-47572004000400025.

Edwards, M., Nathanson, A., Wisch, M., 2014. Estimate of potential benefit for Europe of fitting autonomous emergency braking (AEB) systems for pedestrian protection to passenger cars. Traffic Inj. Prev. 15 (sup1), 172–183. https://doi.org/10.1080/15389588.2014.931579.

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G.B., Victor, T., 2018. Great expectations: a predictive processing account of automobile driving. Theor. Issues Ergon. Sci. 19 (2), 156–194. https://doi.org/10.1080/1463922X.2017.1306148.

Howard, C., Linder, A., 2014. Review of Swedish Experiences Concerning Analysis of People Injured in Traffic Accidents, Tech. Rep. No. VTI Notat 7A-2014. URN: urn: nbn:se:vti:diva-6811.

Isaksson-Hellman, I, Lindman, M., 2016a. Using insurance claims data to evaluate the collision-avoidance and crash-mitigating effects of collision warning and brake support combined with adaptive cruise control. Proceedings of the IEEE Intelligent Vehicles Symposium (IV) 1173–1178. https://doi.org/10.1109/IVS.2016.7535538.

Isaksson-Hellman, Irene, Lindman, M., 2016b. Evaluation of the crash mitigation effect of low-speed automated emergency braking systems based on insurance claims data. Traffic Inj. Prev. 17 (sup1), 42–47. https://doi.org/10.1080/15389588.2016.1186802.

Isaksson-Hellman, I., Norin, H., 2005. How thirty years of focused safety development has influenced injury outcome in volvo cars. In: Annual Proceedings. Association for the Advancement of Automotive Medicine, 49, pp. 63–77. PMID: 16179140.

Jeppsson, H., Östling, M., Lubbe, N., 2018. Real life safety benefits of increasing brake deceleration in car-to-pedestrian accidents: simulation of Vacuum Emergency Braking. Accid. Anal. Prev. 111, 311–320. https://doi.org/10.1016/j.aap.2017.12.001.

Kaufman, L., Rousseeuw, P.J., 2005. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Series in Probability and Statistics. ISBN: 978-0-470-31748-8.

Kitajima, S., Shimono, K., Tajima, J., Antona-Makoshi, J., Uchida, N., 2019. Multi-agent traffic simulations to estimate the impact of automated technologies on safety. Traffic Inj. Prev. 20 (sup1), 58–64. https://doi.org/10.1080/15389588.2019.1625335.

Kovaceva, J., Bálint, A., Schindler, R., Schneider, A., 2020. Safety benefit assessment of autonomous emergency braking and steering systems for the protection of cyclists and pedestrians based on a combination of computer simulation and real-world test results. Accid. Anal. Prev. 136 https://doi.org/10.1016/j.aap.2019.105352.

Kusano, K.D., Gabler, H.C., 2015. Comparison of expected crash and injury reduction from production forward collision and lane departure warning systems. Traffic Inj. Prev. 16 (sup2), 109–114. https://doi.org/10.1080/15389588.2015.1063619.

Lindman, M., Isaksson-Hellman, I., Strandroth, J., 2017. Basic numbers needed to understand the traffic safety effect of automated cars. Proceedings of the International Research Council on the Biomechanics of Injury (IRCOBI) 2017, Paper No IRC-17-10.

Lubbe, N., Jeppsson, H., Ranjbar, A., Fredriksson, J., Bärgman, J., Östling, M., 2018. Predicted road traffic fatalities in Germany: the potential and limitations of vehicle safety technologies from passive safety to highly automated driving. Proceedings of the International Research Council on the Biomechanics of Injury (IRCOBI) 2018, Paper No IRC-18-11.

Östh, J., Brolin, K., Carlsson, S., Wismans, J., Davidsson, J., 2012. The occupant response to autonomous braking: a modeling approach that accounts for active musculature. Traffic Inj. Prev. 13 (3), 265–277. https://doi.org/10.1080/15389588.2011.649437.

Östling, M., Jeppsson, H., Lubbe, N., 2019a. Predicted crash configurations for autonomous driving vehicles in mixed German traffic for the evaluation of occupant restraint system. Proceedings of VDI Fahrzeugsicherheit 2019 Conference.

Östling, M., Lubbe, N., Jeppsson, H., Puthan, P., 2019b. Passenger car safety beyond ADAS: defining remaining accident configurations as future priorities. Proceedings of the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Paper No 19-0091.

Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.D.F., Rodrigues, F.A., 2019. Clustering algorithms: a comparative approach. PLoS One 14 (1). https://doi.org/10.1371/journal.pone.0210236.

Rothoff, M., Zarghampour, H., Tivesten, E., Victor, T., Ödblom, A., Bergh, T., Lindenberg, B., 2019. Autonomous Driving Effects on Sustainable Transportation (ADEST) & Autonomous Driving Fuel Economy (ADFE). Tech. Rep Url. https://www.trafikverket.se/contentassets/8f6c28909584476485dd15973a6af62c/adest_adfe_public_vinnova_report_191031.pdf.

Sander, U., Lubbe, N., 2018a. The potential of clustering methods to define intersection test scenarios: assessing real-life performance of AEB. Accid. Anal. Prev. 113, 1–11. https://doi.org/10.1016/j.aap.2018.01.010.

Sander, U., Lubbe, N., 2018b. Market penetration of intersection AEB: characterizing avoided and residual straight crossing path accidents. Accid. Anal. Prev. 115, 178–188. https://doi.org/10.1016/j.aap.2018.03.025.

Wågström, L., Leledakis, A., Östh, J., Lindman, M., Jakobsson, L., 2019. Integrated safety: establishing Links for a comprehensive virtual tool chain. Proceedings of the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Paper No 19-0177.

Wimmer, P., Düring, M., Chajmowicz, H., Granum, F., King, J., Kolk, H., Op den Camp, O., Scognamiglio, P., Wagner, M., 2019. Toward harmonizing prospective effectiveness assessment for road safety: comparing tools in standard test case simulations. Traffic Inj. Prev. https://doi.org/10.1080/15389588.2019.1616086.

Yue, L., Abdel-Aty, M., Wu, Y., Wang, L., 2018. Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems. Accid. Anal. Prev. 117, 55–64. https://doi.org/10.1016/j.aap.2018.04.002.