LOUGHBOROUGH UNIVERSITY

Loughborough Design School

DOCTORAL THESIS

Safety-critical scenarios and virtual testing procedures for automated cars at road intersections

by

Philippe Nitsche

Submitted in partial fulfilment of the requirements for the award of
Doctor of Philosophy of Loughborough University

August 2018

Abstract

This thesis addresses the problem of road intersection safety with regard to a mixed population of automated vehicles and non-automated road users. The work derives and evaluates safety-critical scenarios at road junctions, which can pose a particular safety problem involving automated cars. A simulation and evaluation framework for car-to-car accidents is presented and demonstrated, which allows examining the safety performance of automated driving systems within those scenarios.

Given the recent advancements in automated driving functions, one of the main challenges is safe and efficient operation in complex traffic situations such as road junctions. There is a need for comprehensive testing, either in virtual testing environments or on real-world test tracks. Since it is unrealistic to cover all possible combinations of traffic situations and environment conditions, the challenge is to find the key driving situations to be evaluated at junctions.

Against this background, a novel method to derive critical pre-crash scenarios from historical car accident data is presented. It employs k-medoids to cluster historical junction crash data into distinct partitions and then applies the association rules algorithm to each cluster to specify the driving scenarios in more detail. The dataset used consists of 1,056 junction crashes in the UK, which were exported from the in-depth "On-the-Spot" database. The study resulted in thirteen crash clusters for T-junctions, and six crash clusters for crossroads. Association rules revealed common crash characteristics, which were the basis for the scenario descriptions.

As a follow-up to the scenario generation, the thesis further presents a novel, modular framework to transfer the derived collision scenarios to a sub-microscopic traffic simulation environment. The software CarMaker is used with MATLAB/Simulink to simulate realistic models of vehicles, sensors and road environments and is combined with an advanced Monte Carlo method to obtain a representative set of parameter combinations. The analysis of different safety performance indicators computed from the simulation outputs reveals collision and near-miss probabilities for selected scenarios. The usefulness and applicability of the simulation and evaluation framework is demonstrated for a selected junction scenario, where the safety performance of different in-vehicle collision avoidance systems is studied. The results show that the number of collisions and conflicts were reduced to a tenth when adding a crossing and turning assistant to a basic forward collision avoidance system.

Due to its modular architecture, the presented framework can be adapted to the individual needs of future users and may be enhanced with customised simulation models. Ultimately, the thesis leads to more efficient workflows when virtually testing automated driving at intersections, as a complement to field operational tests on public roads.

Acknowledgment

The research for this thesis was conducted in the scope of a joint project between Loughborough University and the AIT Austrian Institute of Technology. The accident data in this study was acquired in cooperation with the UK Department for Transport that provided access to the RAIDS database. The Road Accident In Depth Studies (RAIDS) programme and associated database were commissioned by the United Kingdom Department for Transport in 2012 to consolidate data gathered from historical in-depth collision investigation programmes dating back to the year 2000.

I would like to express my gratitude to my supervisors Ruth Welsh and Pete Thomas for their valuable advice and guidance throughout my studies. You helped me to find my way through the forest of possible options to achieve my goals.

Special thanks also to my sponsor AIT, who made this dissertation possible and provided the necessary research infrastructure. I particularly acknowledge the advice in mathematical questions given by Rainer Stütz as well as the support from Alexander Genser in implementing the MATLAB/Simulink models.

Finally, the greatest Thank You is addressed to my family and my partner Maggie, who did not only accept countless hours of absenteeism I spent in my office but also constantly believed in my perseverance.

Table of content

1 Introduction

1.1 Research problem

Current developments and ambitions for the future of the road transport system are now working towards automated driving. Over the past few years, automation of road vehicles has gained an increasing presence on the agendas of companies and public authorities. OEMs and their suppliers, as well as the European Commission, have started to push Automated Driving Systems (ADS) into the forefront of research (Jääskeläinen, 2013). The word "autonomous" stems from the Greek word "autonomos", meaning "having one's own laws"(Oxford Dictionaries, n.d.). With the rise of autonomous vehicles on public roads, it hence seems obvious that they must have their own principles and that they cannot just take over human principles in order to function properly[1]. The introduction of self-driving vehicles of whatever level of automation necessitates a shift of "laws", both in a legal and in a technical sense. This shift also involves a change in road safety policy as well as road infrastructure management and maintenance.

Concerning road safety, it is still not clear what impact automated vehicles can have on crash risk, and what kinds of (new) risks they might cause. In particular, the safety risks coming with a mixed vehicle population, namely traffic with both driverless and driver-operated vehicles are still subject to research. While automated cars and their onboard sensors have better capabilities to recognise their environment compared to the human drivers, they have limitations, e.g. in challenging urban traffic situations, inclement weather conditions or when facing unexpected behaviour of traffic participants.

On spots in a road network, where traffic conflicts are likely to occur, e.g. at intersections, it must be ensured that automated vehicles can operate safely and efficiently, and even more important, that conventional vehicles driven by humans will have the same safety level as they have now. Figure 1 illustrates the transition phase from the safety situation nowadays to the one coming with automated vehicles in public road traffic. While nowadays the nature and distribution of crashes is widely known based on historical accident data, the future situation can only be assessed by assumptions, since automated cars are not yet widely spread on public roads. Hence, it is neither entirely clear what crash risks can be expected nor how much the overlap to the existing situation will be. Which crash types might become negligible in road safety policy, and what new crash types will come with automated driving?

The technical reliability of ADS depends on the functionality under varying road infrastructure and environment conditions as well as on a safe interplay with

[1] A further definition of different automation levels is given in Section 2.1, but it should be clarified at this point that automated driving can still involve a human driver, although he or she does not have to permanently monitor the traffic environment anymore, as this is done by the vehicle.

traditional vehicles and vulnerable road users. Consequently, verification and validation procedures[2] for those systems are paramount. There is a need for comprehensive testing, either in virtual environments or on real-world test tracks. While in some countries testing on public roads has been allowed, all OEMs use virtual testing methods to evaluate their vehicle systems' functionality and reliability. Apart from driving simulators, there are different levels of simulation, ranging from macroscopic to microscopic analysis of traffic. The so-called sub-microscopic simulation level refers to the most detailed scale of investigation, because single vehicles are simulated using physical models, e.g. for tires, suspension, engine or sensors, and their interaction with the surrounding road area can be studied. While microscopic traffic simulation is normally used to evaluate traffic flow for transportation system design, traffic operations and management alternatives evaluation, sub-microscopic simulation is applied to validate vehicle systems and components in particular situations.

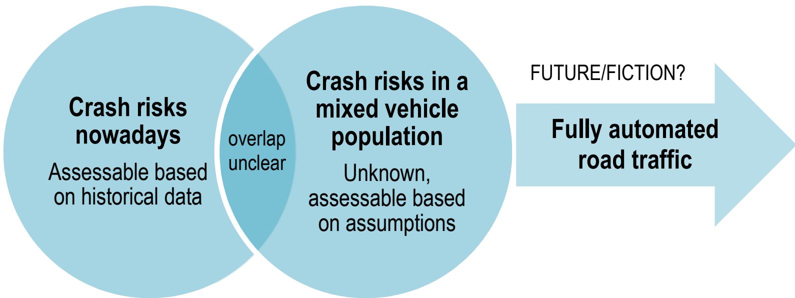


Figure 1: Gap of knowledge on crash risks in a mixed vehicle population

Simulation is an appropriate method for studying complex systems that are inaccessible through direct observation and real-world measurement, because they create abstractions of the system (Lamotte et al., 2010). Those abstractions are made by modelling traffic including vehicle systems, sensors, road environment and driving behaviour. The current problem in this respect is how to ensure realistic and representative behaviour of the models. On the one hand, real-world data can be processed to create models, e.g. pre-crash behaviour from accident studies. On the other hand, valid and representative assumptions must be made for the parameters, where limited data is available.

Maximum reliability of ADS must be achieved even in the rarest events that can occur in traffic. This leads to another challenge, namely to find the key driving situations to be evaluated. The space of possible scenarios is spanned by many dimensions, such as road geometry, the behaviour of the driver and other participants, weather conditions, vehicle characteristics, component faults and others (Tatar, 2016). Since it is unrealistic to cover all possible combinations of traffic situations and environment

---

[2] A simple way to explain the difference between those terms is: Are we building the product right? (Verification) And, are we building the right product? (Validation)

conditions, the most representative "benchmark" scenarios must be known. The automotive industry is still seeking these key scenarios (Lemmer, 2017; Pütz et al., 2017). There may be different testing scenarios depending on which issue is targeted. For example, targeting at maximum casualty reduction for vulnerable road users will require different testing measures than just targeting at the vehicles' full functionality.

As road intersections are locations, where the paths of multiple traffic participants are crossed, they are considered high-risk spots for safety researchers. For automated vehicles, road intersections of whatever type constitute a major point of interest along their routes due to the increased likelihood of conflicts with other road users. Therefore, intersections play a particularly important role in testing assisted and automated driving. Automated vehicles should be capable of safely manoeuvring through an intersection and of avoiding or mitigating a collision. Intersection assistance systems, also denoted as intersection crash avoidance and mitigation systems (ICAMS), can be either vehicle-based or infrastructure-based. The latter type of systems, e.g. roadside-based warnings communicated to the vehicles can assist vehicle-based ICAMS, which have limitations in adverse conditions and particular situations. However, there is still no commercially available intersection assistance system for automated vehicles on the market. There is a need to evaluate those systems on a large variety of intersection layouts and characteristics, taking into account the most critical combinations of collision parameters.

The assessment of testing scenarios is commonly based on safety indicators, which can be used to measure the spatial and temporal proximity of safety-critical events, assuming that they have an established relationship with accidents (Archer, 2005). Hence, the safety evaluation of simulation runs should not only result in collision probabilities, but also in near-miss probabilities. There has been significant research on appropriate indicators to evaluate the severity of road conflicts (Archer, 2005; Bagdadi, 2013; Brown, 2005; De Ceunynck, 2017; Laureshyn et al., 2017, 2010; Mahmud et al., 2017), but only a few studies have investigated indicators for simulated junction scenarios in particular. Since such indicators are the most important elements of a virtual testing procedure, there is a need to establish a profound method to assess the safety performance of ICAMS.

In summary, the research problem addressed by this thesis comprises two issues: First, to ensure that automated vehicles can operate safely and efficiently at road intersections. Second, to identify the key scenarios and virtual testing procedures for evaluating intersection assistance for ADS.

1.2 Objectives and research questions

The objective of the thesis is to provide a novel validation tool for testing the safety of automated vehicles in road junction environments. This tool is realised in the form of a simulation and evaluation framework, which leads to more efficient workflows when virtually testing ICAMS. The work includes the development of a method to generate safety-critical testing scenarios from accident data, a method to virtually reconstruct and evaluate the scenarios as well as the application of indicators to quantify the safety performance of ICAMS.

The thesis primarily targets the automotive and supply industry to test and develop vehicle sensors and automated driving systems in a more efficient way. Furthermore, the thesis enables the systematic testing of roadside-based ICAMS, which helps infrastructure providers or manufacturers to evaluate the effectiveness. The framework is to be designed in a modular architecture, which can be adapted to the individual needs of future users and enhanced with customised models. Accordingly, the following *research questions* are addressed in this thesis.

RQ1. Which technical hurdles and challenges do ADS currently have to overcome?

⇒ Literature review (Chapter 2)

⇒ Expert survey (Chapter 4)

RQ2. What are the state-of-the-art technologies to enable automated driving in junction environments?

⇒ Literature review (Chapter 2)

RQ3. What are the current collision scenarios at three- and four-legged at-grade junctions and how can they be clustered from historical accident data?

⇒ Initial analysis of accident data (Chapter 4)

⇒ Identification of critical scenarios from crash data (Chapter 5)

RQ4. How can those collision scenarios be represented and enhanced for sub-microscopic simulation to evaluate the safety performance of intersection assistance systems?

⇒ Development of a modular simulation and evaluation framework (Chapter 6)

RQ5. What general recommendations can be made for the safety performance indicators to be considered in virtual testing of ADS at junctions?

⇒ Safety performance evaluation (Chapter 6.5)

⇒ Discussion and conclusions (Chapter 7)

1.3 Scope and contributions

This thesis deals with safety aspects of automated road transport. The scope is limited to at-grade road intersections involving all types of three- and four-legged junctions, i.e. excluding roundabouts. This limitation has been chosen due to the following reasons:

- Every third accident occurs at a junction, which can therefore be considered as a relevant risk area for road safety(European Commission, n.d.).

- Junctions have more conflict points between road user paths than other road areas. Hence, passing an intersection constitutes a complex and challenging driving task for ADS (European Commission, 2015).
- Elvik and Vaa (2004) state that the total number of accidents is significantly reduced when converting intersections to roundabouts, e.g. due to fewer conflict points, lower speeds or avoidance of 90 degrees angles etc. As mentioned in the previous section, this thesis presents a framework to test automated vehicle systems in challenging intersection scenarios. Roundabouts were excluded from the analysis, because risks that come with sight obstructions and crossing road users are mitigated to some extent by the safe design of roundabouts. Also, the different design principle of a roundabout limits the number of parameter variations compared to other junctions. For example, the directions of travel, the driving manoeuvres or the yield instructions do normally not vary at roundabouts. However, this exclusion does not restrict the methodology and applicability of this thesis, because the provided framework can be adapted to evaluate roundabouts as well.
- Vehicle sensors and automated driving technologies struggle with complex traffic environments that involve many different road users or sight obstructions (Templeton, n.d.). A profound investigation of intersection safety would contribute to the near-term challenges coming with the deployment of automated vehicles, especially on non-motorway roads or in urban areas.

As the research questions imply, the thesis investigates how collision scenarios derived from accidents in driver-operated traffic can be utilised for virtual, simulation-based testing, where an automated driving system replaces the driver. At the moment, there are limited European regulations on validating the reliability of highly automated road vehicles. Standardized procedures for evaluating automated driving systems are highly relevant to guarantee high safety in a varying environment. To this end, this research will contribute to the development of novel vehicle environment recognition systems (e.g. visual or LIDAR –based systems) and infrastructure-based intersection assistance systems by providing the following main outputs:

1. *Pre-crash scenarios for testing ADS at intersections*, taking into account different intersection characteristics as well as interplay with non-automated vehicles: A novel method to derive pre-crash scenarios from historical car accident data is presented. It employs k-medoids to cluster historical junction crash data into distinct partitions and then applies the association rules algorithm to each cluster to specify the driving scenarios in more detail.
2. *A sub-microscopic simulation framework* to virtually reconstruct and evaluate collision scenarios for automated cars under various conditions and in different junction sceneries: The framework uses physical models for vehicles and their sensors as well as accurate road environment models and applies the

Monte-Carlo method to sample a representative set of parameter combinations.

3. *Safety performance indicators* as metrics to quantify the impact of in-vehicle or roadside-based intersection assistance systems on collision and near-miss risk: Those indicators are a direct output of the simulation framework.

1.4 Thesis structure and chapter summary

The thesis is structured as depicted in Figure 2. The background on the topic as well as existing work and findings are reviewed in Chapter 2, by summarising literature on current challenges and technologies for automated driving, the particular safety problems at intersections, technology-based solutions for avoiding or mitigating intersection accidents as well as current testing and validation procedures. Finally, Chapter 2is concluded with a definition of research gaps to be addressed by this thesis. It addresses the research questions 1 and 2.

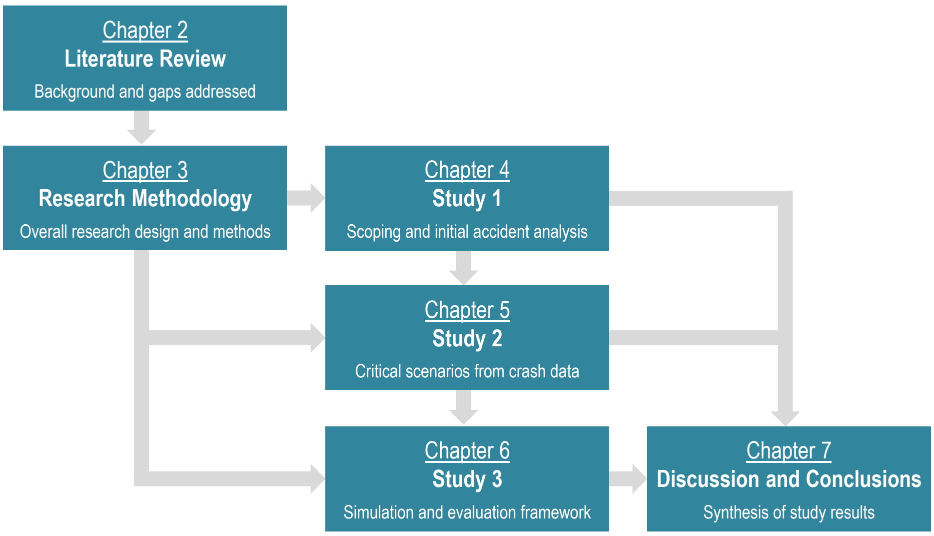


Figure 2: Structure of the thesis chapters

In Chapter 3, the overall research design is described. This is done for all three research studies within the thesis, respectively, and gives a complete picture of the methods used. The first research study is presented in Chapter 4, which sets the scope for the further studies. It presents a web expert survey conducted to identify current challenges and influencing factors for the performance of automated vehicles, followed by a descriptive analysis of historical junction accident data. Hence, the expert survey enhances the literature review and addresses research question 1, while the accident analysis provides initial findings for research question 3.

The conclusions from the scoping study lead to the second study, which is explained in Chapter 5. It deals with the clustering of in-depth crash data to identify safety-critical scenarios at junctions and addresses research question 3.

Chapter 6 follows up on the clustering analysis and presents a novel methodology that aims to transfer the derived collision scenarios to a sub-microscopic traffic simulation environment, where the safety performance of automated driving functions can be evaluated. The usefulness and applicability of the simulation methodology is demonstrated for a selected junction scenery, where the safety performance of two different intersection assistance systems is studied. This third study provides answers to the research question 4 and 5.

Ultimately, the results from all three studies are synthesised in Chapter 7. It covers research question 5 by discussing virtual testing procedures. Limitations of the methods and the generalizability of the results are discussed, before the thesis is concluded with a summary and remarks for future work.

1.5 Publications

Throughout of the duration of the doctoral study, the following topic-related papers were published by the author (see full texts in Appendix H):

Nitsche, P., Thomas, P., Stuetz, R., Welsh, R., 2017. Pre-crash scenarios at road junctions: A clustering method for car crash data. Accident Analysis & Prevention 107, 137–151.

This journal paper was published in Accident Analysis and Prevention in October 2017. It presents the method and results from study 2, i.e. a novel data analysis method including the preparation, analysis and visualisation of car crash data, to identify the critical pre-crash scenarios at three- and four-legged junctions as a basis for testing the safety of automated driving systems.

Nitsche, P., Mocanu, I., Reinthaler, M., 2014a. Requirements on Tomorrow's Road Infrastructure for Highly Automated Driving, in: The 3rd International Conference on Connected Vehicles & Expo (ICCVE 2014). Vienna, Austria.

This paper was submitted to the ICCVE conference 2014 in Vienna. It summarises the web survey results from the scoping study, where ADS was distinguished into lane assistance systems, speed control systems and collision avoidance systems. The paper was selected for oral presentation, which was held by the co-author Martin Reinthaler due to unavailability of the main author. The paper was published in the conference proceedings.

Mocanu, I., Nitsche, P., Saleh, P., 2015. Highly automated driving and its requirements on road planning and design, in: Proceedings of the 25th PIARC World Road Congress. Seoul, Korea.

This paper was submitted to the PIARC World Road Congress 2016 in Seoul. It is an extension to the ICCVE paper from 2014, by summarising the literature review of the scoping study, the web survey results and by giving implications on road planning and road design principles. The latter part of the paper was not part of the PhD activities and was written by Isabela Mocanu and Peter Saleh. The paper was selected for poster presentation, which was done by Philippe Nitsche. Additionally, the authors were invited to present the paper in a special plenum session, which was done by Philippe Nitsche.

2 Literature review

This chapter summarises the literature review carried out and is concluded with research gaps addressed by the thesis. Each literature review section is condensed by listing the most important findings.

2.1 The state of play concerning automated driving

There are different terms used for automated cars, among which are autonomous, self-driving, driver-less or robot cars. However, it is argued that automation and connectivity are strongly linked (ERTRAC, 2017). Hence, automated cars will have to be connected and to receive data from "the outside" such as digital map updates, dynamic traffic or route information or hazard warnings. The car can thus not be called "autonomous" anymore. Therefore, this thesis henceforth uses the term "automated".

2.1.1 Evolution vs. revolution

Automated vehicles were not brought to market directly, but were introduced incrementally due to different reasons concerning legal, technical, economical and other barriers. This step-wise introduction can be called an evolution instead of a revolution (Vanholme et al., 2013). This viewpoint is argued with the fact that Advanced Driver Assistance Systems (ADAS) have built the bridge towards automated driving, with clear evolutionary steps from basic stability assistance such as the Anti-Lock Braking System (ABS) to more sophisticated technologies.

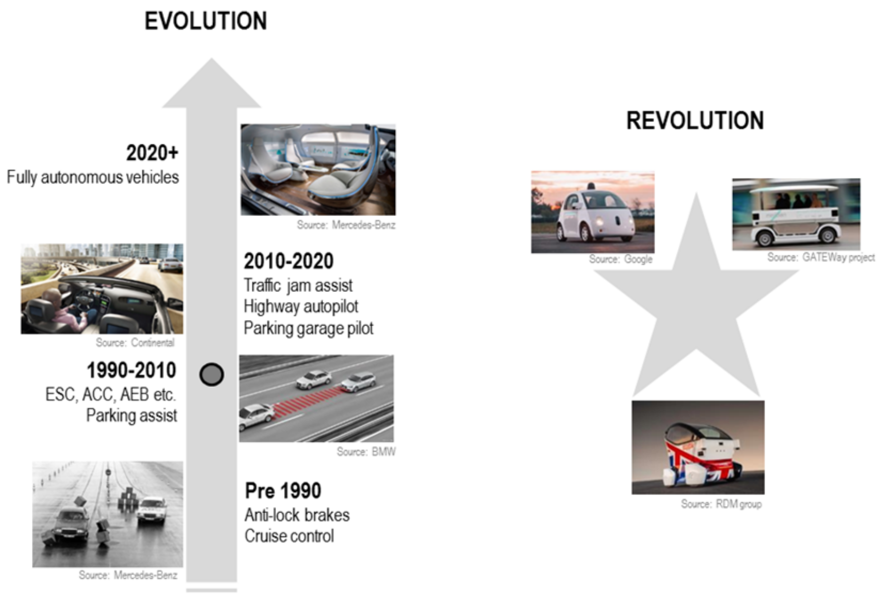


Figure 3: Evolution vs. revolution in automated road transport

As depicted in Figure 3, automated parking, Autonomous Emergency Braking (AEB), Adaptive Cruise Control (ACC) and Lane Keeping Assist (LKA) are examples of technologies that have been brought into the market in the past decade. Among others, these systems have enabled partial automation in certain driving situations with the aim of increasing travel comfort, safety and efficiency. On the other hand, new players such as Google, Tesla or Navya have started to produce automated vehicles, which can be seen as a revolutionary step. Automated vehicles were built from scratch, mostly based on electric drive technologies.

2.1.2 Levels of automation

The American National Highway Traffic Safety Administration (NHTSA, 2013), Germany's Bundesanstalt für Straßenwesen BASt (Gasser, 2012) and the Society of Automotive Engineers (SAE, On-road Automated Vehicle Standards Committee (2014) have introduced certain levels of automation, which differ in the extent of human driver involvement. The European industry has agreed to use the SAE levels from 0 to 5 as common understanding of automated driving (see Figure 4) and therefore, this thesis refers to those levels in the further chapters.

SAE Level	SAE Name	SAE Narrative Definition	Execution of Steering/ Acceleration/ Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System capability (driving modes)	BASt Level	NTHSA Level
Human Driver monitors the driving environment								
0	No Automation	the full-time performance by *the human driver* of all aspects of the *dynamic driving task*	Human Driver	Human Driver	Human Driver	N/A	Driver only	0
1	Driver Assistance	the *driving mode-specific* execution by a driver assistance system of either steering or acceleration/deceleration	Human Driver and Systems	Human Driver	Human Driver	Some Driving Modes	Assisted	1
2	Partial Automation	Part-time or driving mode-dependent execution by **one or more driver assistance systems** of both steering and acceleration/deceleration. Human driver performs all other aspects of the *dynamic driving task*.	**System**	Human Driver	Human Driver	Some Driving Modes	Partially Automated	2
Automated driving system ("system") monitors the driving environment								
3	Conditional Automation	*driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task* - *human driver* **does** respond appropriately to a *request to intervene*	System	**System**	Human Driver	Some Driving Modes	Highly Automated	3
4	High Automation	*driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task* - *human driver* **does not** respond appropriately to a *request to intervene*	System	System	**System**	Some Driving Modes	Fully Automated	3/4
5	Full Automation	**full-time** performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	**System**	Some Driving Modes		

Figure 4: Levels of vehicle automation (adapted from ERTRAC, 2017)

There is a clear shift from level 2 (partial automation) to level 3 (conditional automation), namely that the vehicle system takes over the monitoring of the driving environment. While in level 3, the driver still has to take over vehicle control appropriately if requested, level 4 (high automation) does not require the driver to respond in time, at least for a certain amount of time or travel route. Promising near-future systems of level 4 are the highway pilot or parking garage pilot, as certain car manufacturers name them. Level 5 (full automation) is the only level where the human driver is fully kept out of the loop from source to destination, or might not even be present in the car. The latter level is often called "robot car", e.g. particularly interesting for car sharing companies to relocate their fleet.

While level 1 and 2 systems are already on the market (e.g. adaptive cruise control, stop &go functionality, lane keeping assist), little market research can be found about the timescale of introducing level 3 to 5 on European roads. However, experts estimate fully automated cars to be ready not before 2025, although the area of use must be considered. For example on highways, highly automated vehicles are expected earlier than on urban roads (ERTRAC, 2017).

2.1.3 Current challenges

Many prototypes of automated vehicles have been developed and demonstrated in recent years, originally initiated in the USA by the DARPA Grand Challenges (2004 and 2005) for autonomous land vehicles or the DARPA Urban challenge in 2007 (Buehler et al., 2009). Nowadays, the rapid developments in automated driving technologies encouraged European organisations to establish real-world testing areas for automated vehicles. Examples for such test sites are AstaZero (Chalmers, 2014), the IDIADA testing grounds in Spain, DRIVE-ME (Volvo Car Group, 2013) in Sweden, the A9 Digital Motorway Testbed (BMVI, n.d.) and the Testfeld Niedersachsen (DLR, n.d.) in Germany, AlpLab in Austria ("AlpLab - Die Zukunft des Fahrens," 2017) or several test sites in the UK (Burn-Callander, 2015; Council, n.d.; TRL, 2014). The overall goal of those real-world tests is to reveal challenges and problem areas when driving on public roads. Among the most important known challenges for ADS, as published in various papers (Dokic et al., 2015; ERTRAC, 2017; iMobility Forum, 2013), are:

- *Environment detection, perception and prediction*: The capability to perceive a vehicle's environment with high accuracy and reliability is crucial for ADS. The performance of environment sensors must meet regulatory requirements, and sensor fusion techniques must be further investigated and applied. Furthermore, robustly predicting other road users' trajectory and movements will be a central part of future research.
- *Reliability and safety of technology:* Reliable and safe system performance is a prerequisite for implementing automated driving. This involves a fault-tolerant architecture, but more importantly, the technological readiness must be demonstrated to convince public decision makers.
- *Common validation and testing procedures:* A common validation procedure is required to test ADS under all circumstances including varying road infrastructure and environmental condition or false usage. Issues such as key performance indicators, certification procedures and cost-effectiveness are still subject to research.
- *Interplay with the road infrastructure:* Basic research is needed on what adaptations existing roads require for implementing self-driving vehicles. Or vice versa, what minimum requirements do vehicles need to function properly. This also includes digital infrastructure, such as level of connectivity and data provision.

- *Human factors:* The automation level 3 still involves the driver in certain phases, so that behavioural factors, situation awareness and intuitive usability must be ensured, especially when handing over control. This involves research areas such as human-machine interface (HMI) design or long-term effect studies.
- *Users' and societal acceptance:* Besides affordability, which is, without doubt, a major acceptance hurdle, safety, trust, data security and privacy play an important role. Research is needed to identify the benefits of ADS to raise public awareness about the impacts of ADS on society. However, also the negative effects need to be investigated.
- *Traffic management:* A further challenge is to identify the role and tasks of traffic management systems in the world of automated vehicles. Research is needed on the level of supervision of automation, cooperation and communication with vehicles or new logistics applications.
- *Liability and legal aspects*: Legal issues have to be regulated and harmonised for maintaining Europe's competition in the automotive section compared to the US. It must further be clarified, how insurance companies can adapt their liability models considering the new requirements with ADS.
- *Data security*: On the one hand, data security needs further research on reliable and secure vehicular communication and data storage or accessibility. On the other hand, this also involves issues such as data ownership.

Recent developments regarding sensing, perception and trajectory planning, as summarised by Okuda et al. (2014), show promising improvements. This includes vision-based algorithms for object recognition, Simultaneous Localization and Mapping (SLAM, Lategahn et al., 2011) and sophisticated path and motion planning. Keeping these technologies in mind, the following chapter focuses on the challenges from an engineering point of view. Legal and societal challenges as well as human factors certainly constitute major barriers for ADS introduction on a large scale, but are not addressed by this thesis.

2.1.4 European research projects

The EU projects SCOUT (Safe and Connected Automation in Road Transport) and CARTRE (Coordination of Automated Road Transport Deployment for Europe) are coordination and support actions that have built a joint initiative called CAD (Connected and Automated Driving) to support the development of clearer and more consistent policies. On their website[3], the initiative lists all completed and ongoing research projects and field trials dealing with automated road transport. Table 1 gives all ongoing projects, sorted by the project end date. When looking at the scope of those projects, it can be seen that the focus of current European research points towards 1) real-world testing areas and pilot sites, 2) validation procedures for

[3] https://connectedautomateddriving.eu/, accessed on 11 July 2018.

sensors, in particular in adverse weather conditions and 3) integration of cooperative ITS in vehicle automation.

Table 1: Ongoing European research projects (status: April 2018)

Project name	Countries	Duration	Description
Digital Motorway Test Bed https://www.bmvi.de/SharedDocs/EN/Articles/DG/digital-motorway-test-bed.html	DE	15.9.2015 – 31.12.2022	Test Bed on the A9 motorway in Bavaria, Germany
C-ROADS https://www.c-roads.eu/platform.html	AT, BE, CZ, FI, FR, DE, HU, IE, NL, SI, CH, UK	1.2.2016 – 31.12.2020	Open platform for interoperable C-ITS services for European travellers
TransAID https://www.transaid.eu	DE, BE, FR, GR, NL, UK	30.9.2017- 31.8.2020	Transition Areas for Infrastructure-Assisted Driving, traffic management procedures and protocols
L3Pilot http://www.l3pilot.eu/	AT, BE, FI, FR, DE, GR, IT, NL, NO, SE, CH, UK	1.9.2017 – 31.8.2020	Large-scale piloting of SAE Level 3 functions, with additional assessment of some Level 4 functions
INFRAMIX http://www.inframix.eu	AT, DE, ES, GR	1.6.2017 – 31.5.2020	Road infrastructure to support the coexistence of conventional and automated vehicles
ADAS&ME http://www.adasandme.com/	SE	9.1.2016 – 29.2.2020	ADAS development to incorporate driver state, situational/ environmental context and adaptive interaction
AUTOPILOT http://autopilot-project.eu	FI, FR, IT, NL	1.1.2017 – 31.12.2019	IoT integration into the automotive world to transform connected vehicles into highly and fully automated vehicle
VI-DAS http://vra-net.eu/	ES	1.9.2016 – 31.8.2019	Non-invasive, vision-based sensing technologies to enable contextual driver behaviour modelling
MAVEN http://www.maven-its.eu/	DE, NL, UK	1.9.2016 – 31.8.2019	Infrastructure assisted algorithms for the management of automated vehicles for trajectory and manoeuvre planning
AutoMate	DE	1.9.2016 – 31.8.2019	Novel driver-automation interaction and cooperation concept
PEGASUS http://www.pegasus-projekt.info	DE	1.1.2016 – 30.6.2019	Standardized procedure for the testing and experimenting of automated vehicle systems in simulation, on test stands and in real environments
DENSE http://dense247.eu	BE, FI, FR, DE, NL, ES, SE	1.6.2016 – 31.5.2019	Environmental sensing system for adverse weather conditions
ENABLE-S3 http://www.enable-s3.eu/	AT + 73 partners	1.5.2016 – 30.4.2019	Virtual testing, verification and coverage-oriented test selection methods for automated systems
AUTOC-ITS https://project.inria.fr/autocits/	FR, PT, ES	1.11.2016 – 31.12.2018	Enhancing C-ITS interoperability for autonomous vehicles through 3 pilots
SCOOP@F	AT, FR, PT, ES	1.1.2016 – 31.12.2018	C-ITS pilot deployment for validating hybrid communication (ITS-G5/3-4G) on five test sites
Aurora http://www.liikennevirasto.fi/web/en/e8-	FI	1.1.2016 – 31.12.2018	Test area for verification and validation of intelligent transport automation in real extreme weather conditions

aurora#.WY2CjsYlFaR			
UnCoVerCPS https://cps-vo.org/group/UnCoVerCPS	FR, DE, IT, ES, UK	1.1.2015 – 31.12.2018	Control and verification concepts for cyber-physical systems in unknown environments and unanticipated situations.
Venturer http://www.venturer-cars.com/	UK	1.7.2015 – 1.7.2018	Test site facility for connected and autonomous vehicles in the South West UK.
aFAS https://www.afas-online.de/projektueberblick/	DE	1.8.2014 – 31.7.2018	aFAS pilots driverless road maintenance utility vehicles on public motorways
inLane https://inlane.eu	AT, BE, FR, DE, NL, ES	1.1.2016 – 30.6.2018	Development of a low-cost, lane-level, precise turn-by-turn navigation application through the fusion of EGNSS and computer vision technology
SocialCar http://socialcar-project.eu/	BE, HR, HU, IT, LU, MK, PL, SI, ES, UK	1.6.2015 – 31.5.2018	Intelligent transport system for carpooling in urban and peri-urban areas.
RobustSENSE https://www.robustsense.eu/	AT, FI, DE, IT, ES	1.6.2015 – 31.5.2018	Test bench for sensors of ADS to be validated in harsh environmental conditions
ROADART http://www.roadart.eu/	DE, GR, NL	1.5.2015 – 30.4.2018	Integration of ITS communication units into trucks to boost road safety (truck-to-truck and truck-to-infrastructure systems)
HIGHTS http://hights.eu/	FR, DE, LU, NL, SE	1.5.2015 – 30.4.2018	High precision positioning system with the accuracy of 25cm for C-ITS
SADA http://robotik.dfki-bremen.de/en/research/projects/sada-1.html	DE	1.2.2015 – 31.1.2018	Smart Adaptive Data Aggregation, by linking data from mobile onboard sensors with data from stationary sensor infrastructure.
UKAutodrive http://www.ukautodrive.com/	UK	1.1.2015 – 31.1.2018	Programme for trialling automated vehicles in the UK, including sites in Coventry, Milton Keynes and the HORIBA-MIRA test track.
GATEway https://gateway-project.org.uk/	UK	1.1.2015 – 1.1.2018	Pilot test site for automated shuttle buses in Greenwich, London
Nordic Way http://vejdirektoratet.dk/EN/roadsector/Nordicway	DK, FI, NO, SE	1.1.2015 – 31.12.2017	Pilot to enable vehicles to communicate safety hazards through cellular networks on a road corridor through Scandinavia.
Drive Me http://www.volvocars.com/intl/about/our-innovation-brands/intellisafe/autonomous-driving/drive-me	SE	30.11.2012 – 30.11.2017	Large-scale autonomous driving pilot project in which 100 self-driving Volvo cars will use public roads in Gothenburg, Sweden
HFAuto http://hf-auto.eu/	DE, NL, SE, UK	1.11.2013 – 31.10.2017	Multidisciplinary research and training programme for human factors of automated driving

2.1.5 Findings

In summary, the following findings about automated driving in general can be stated:

- While traditional OEMs are advancing their driving assistance systems step-wise towards automated driving, new players such as Google or Tesla revolutionised the market with their vehicles.
- There is common agreement among the automotive industry to use the five levels of automation published by SAE. Accordingly, there is a clear shift from level 2 to level 3, where the vehicle system takes over the monitoring of the driving environment.
- Highly or fully automated vehicles are far from operating perfectly, as there are still major technical challenges when it comes to robust environment and object perception, interplay with other road users and trajectory prediction.
- There is a large number of ongoing research projects funded by the European Commission that address various challenges of automated road transport. The main focus points towards validation and verification procedures, either on public roads or in virtual environments.

2.2 Current technologies for automated vehicles

Automated vehicles employ numerous technologies to drive safely and efficiently through road traffic. This section summarises state of the art sensors, algorithms and methods for self-driving vehicles by describing their technical capabilities as well as limitations. There are different ways for structuring ADS technologies. For example, one option is to investigate the wide range of ADAS and to distinguish between stability assistance, longitudinal and lateral assistance (Vanholme, 2012).

Stability assistance outperforms the driver in critical situations, e.g. by changing the speed or the trajectory, as done with systems like ABS, Electronic Stability Control (ESC, Liebemann et al., 2004) or Traction Control Systems (TCS, Emig and Schramm, 1989; Maisch et al., 1993). These systems intervene when the limits of vehicle stability are reached, with almost no human controllability except the option to deactivate the function.

Longitudinal assistance works continuously during the trip by controlling speed, braking or gear changing. An early application was the automatic transmission system, followed by cruise control and later Adaptive Cruise Control (ACC, Fancher et al., 2004; Yamamura et al., 2001), which keeps the target speed and safe distance to vehicles ahead, in stop-and-go or free flow traffic. As a next step, Intelligent Speed Adaptation (ISA) was developed to support drivers in keeping the prevailed speed limit, either by advising/warning the driver (e.g. visual or audio signals) or by active intervention, e.g. haptic or dead throttle (Brookhuis and de Waard, 1999; Carsten, 2001; Lai et al., 2012; Partrouche, 2006). ISA requires precise information on actual speed limits and can therefore be based on visual recognition of speeds signs or on updated digital maps, preferably on a combination of both. Regarding automated driving, ISA paved the way towards automated longitudinal control.

The third type of driver assistance, namely *lateral assistance*, delivers support for drivers to stay within the lane or to change to another lane. Automotive manufacturers first introduced a Lane Departure Warning System (LDWS, Visvikis et al., 2013), followed by the Lane Keeping Assistant (LKA), which actively prevents lane departure by intervening in steering or braking. As driving often requires changing a lane or merging to another lane, automotive developers made progress in terms of lane change assistance systems. This increased complexity, since the positions and velocities of surrounding vehicles must be recognised by environment sensors, and the optimal timing of the lane change manoeuvre must be determined by intelligent algorithms (Tideman et al., 2007).

However, since the categorisation into stability assistance, longitudinal and lateral assistance is based on ADAS, i.e. SAE automation level 1 and 2, only a combination of these assistance systems lead to a successful implementation of automated vehicles from level 3 to 5. Furthermore, collision avoidance is a core function of modern vehicles, but not included in the categories above. Therefore, another grouping similar to those reported by Okuda et al. (2014) or Lutin et al. (2013) is chosen for summarising the state of the art (see Table 2):

- *Sensing and perception*: Sensing and perceiving the vehicle's environment is an elementary task of automated vehicles. This includes the recognition of lane edges, other road users and objects mainly based on vision sensors.
- *Vehicular communication*: Connectivity between vehicles (vehicle-to-vehicle, V2V) or between vehicles and roadside units (vehicle-to-infrastructure, V2I/I2V) is a main trend in automotive technologies and therefore considered an important aspect in ADS.
- *Route and motion planning*: This group comprises techniques to find and follow a route to the desired destination as well as methods for short-term manoeuvre planning, e.g. lane changes, evasion manoeuvres or trajectory planning at intersections. Mapping technologies play a significant role in this category, as do collision avoidance and the imitation of human behaviour.

Table 2: Categories of technologies applied to ADS

Sensing and perception	(3D) LIDAR Computer vision (Mono/Stereo vision) Radar Sonar sensors Data fusion methods
Vehicular communication	WiFi (IEEE 802.11) 3G/4G LTE cellular networks Bluetooth
Route and motion planning	GNSS + Inertial sensors Digital maps SLAM Collision avoidance and mitigation systems

It is important to note that these three categories should not be seen as individual functions but as a joint system. For example, motion planning techniques also use information gathered from environment sensors or V2I/I2V or V2V messages.

2.2.1 Sensing and perception

Vehicle sensor technology comprises a variety of hardware elements, while perception includes the software to fuse information gathered by the different sensors. Nowadays, data fusion has become an essential function of ADS developments.

The particular strength of LIDAR is the robust recognition of directly visible elements, independent of ambient light, e.g. during the night. Especially in combination with recorded digital LIDAR maps, this enables real-time vehicle localisation. Another advantage is its robustness against interference, which is the reason why LIDAR is preferred to radar[4] in many applications. However, some drawbacks must be mentioned. First, besides the high costs that are expected to decrease in the upcoming years, most current devices are mounted outside, ideally on the vehicle roof, to function properly. This contradicts most automotive design principles, where all sensors are supposed to be mounted in an unobtrusive way somewhere in the vehicle chassis. Secondly, LIDAR still struggles with adverse weather conditions such as snow, heavy rain and fog, as do other vision sensors (Templeton, n.d.). Here comes the main advantage of radar, which works under conditions where optical sensors normally fail.

Since LIDAR recognises its environment by distance vectors to illuminated targets, it obviously cannot see colour. For example, when approaching a signalised intersection, LIDAR could detect the position of a traffic signal, but not its status (red/green). Due to this fact and due to reasons such as lower costs, higher resolution and higher sight distance, cameras are often used as a supplement to LIDAR. In vehicular applications, computer vision is used to detect traffic signs, lane edges or other objects, persons or vehicles. In theory, computers are trained to extract information from images, similar to the human ability to understand what the eyes see.

Many applications use only one camera to capture the environment, a technique called mono-vision, which is sufficient in many cases as they can recognize objects and estimate distances quite robustly by using state-of-the-art image processing algorithms (e.g. Gornea et al., 2014; Klappstein et al., 2007; Lategahn and Stiller, 2013; Nienhuser et al., 2008). Those algorithms are mainly based on machine learning models trained with a large set of images. However, monovision has some limitations when it comes to measuring the position of and distance to unknown objects, i.e. those, which are not included in the training data.

In stereo-vision, depth-in-field is inferred from two or more cameras, making it possible to estimate relative distances to objects and therefore reconstruct the

[4] Radar was originally an acronym for Radio Detection And Ranging. The term radar has since entered English and other languages as a common noun, losing all capitalisation.

environment. In other words, stereo vision extracts 3D image information from digital images. A major challenge for stereo vision is variations in light conditions such as shadows, light reflections or sun glare (Huh et al., 2008).

As mentioned, LIDAR as well as computer vision rely on optical information captured by light sensors, which makes them susceptible to perception problems in conditions such as rain, snow or fog. That is why modern vehicles still use radar as additional environment sensor, as it transmits radio waves instead of light. Furthermore, sonar sensors are used for vehicle parking assistance to measure the distance to obstacles.

2.2.2 Vehicular communication

Without a doubt, vehicular communication could bring tremendous benefits in road safety, as most collisions could be avoided due to Car-to-X information systems. Examples of promising safety applications are hazardous location warning (e.g. slippery roads), pre-crash sensing, stop sign assist or speed advisory (Baldessari et al., 2007; Ibanez et al., 2011a).

The de facto standard for vehicular communication is the WiFi-based IEEE 802.11p (IEEE, 2007), which is also known as WAVE (Wireless Access in Vehicular Environments) or DSRC (Dedicated Short Range Communication) in the US. The standard was developed to provide the minimum set of specifications required to ensure interoperability between wireless devices in environments that might rapidly change and where delivery must be completed in a very short time period. Examples of applications for WAVE are forward collision warning, lane change warning or traffic signal timing information (Ibanez et al., 2011b).

Vehicular communication can either be realised by V2I/I2V or V2V applications. V2V is also known as VANETs (vehicular ad hoc networks). However, most safety applications based on V2V communication require a high penetration of equipped vehicles in order to bring substantial benefits for road safety. For example, there are intersection collision avoidance systems assuming that all vehicles are equipped with communication devices. As this seems unrealistic, Tsugawa (2005) argues that drivers should not become dependent on such systems and that they must conduct a visual check. Hence, those systems are "meaningless", he states. The development of V2V applications should begin with services that bring benefits even in the absence of high penetration, e.g. incident information from preceding to succeeding vehicles. It can therefore be stated that I2V systems constitute a more promising option to improve intersection safety in the near term.

An alternative for vehicular communication is the use of the 4G/5G network, which has limited applicability for V2V, but more for I2V/V2I. While IEEE 802.11p is designed to operate in a range of up to a few hundred metres, 4G, LTE (Long Term Evolution) or 5G can cover areas of 100 km. That is why it is applied to other vehicular fields such as information systems about road or traffic condition or probe vehicle data collection. Performance studies have shown that 4G meets most of the application requirements regarding reliability, scalability, and mobility support, but has high delays in the presence of higher cellular network load (Abid et al., 2012; Mir and Filali, 2014). Furthermore, Bluetooth can be used for other applications (e.g. parking lot payment), but due to its limited range of 10 metres, it is more practical for stationary or very slow moving vehicles.

2.2.3 Route and motion planning

When a vehicle receives data from its environment sensors or via vehicular communication, it can use the underlying information to perform a certain manoeuvre or to navigate through a critical traffic situation. The third category of ADS technologies, Route and Motion Planning can be described by three different layers (Okuda et al., 2014), see Figure 5. In the highest layer called *route planning*, given road network information is used to create a directed graph to reach the desired destination. GNSS (Global Navigation Satellite Systems) such as GPS combined with inertial sensors, digital maps and route planning methods are the basic technologies for doing this. In areas of low or no satellite coverage or where no digital maps are available, vehicles use SLAM to acquire a map of their environment while simultaneously localising itself relative to the map (Thrun et al., 2006). In the middle layer called *manoeuvre planning*, the self-driving vehicle needs to compute and perform the actual basic manoeuvres along the route, such as lane-change, giving way according to traffic laws or other safety decisions. And the third and lowest layer of planning can be called *motion planning*, when the vehicle has to compute a path to avoid an obstacle or a collision. Along with the interpretation of the vehicle's environment, as described before, algorithms must generate a dynamically feasible trajectory, which is a crucial part of future collision avoidance systems. For ensuring safe automated driving, relevant types of collisions must be evaluated and varying realistic circumstances (e.g. road layout, behaviour of other traffic participants) must be taken into account. The latter issue of collision avoidance is the core topic of this thesis and is described in Section 2.4.

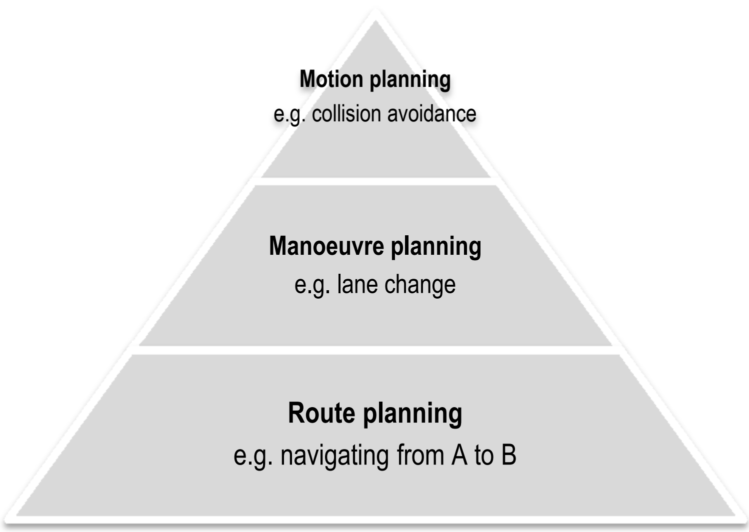


Figure 5: Three layers of automated route and motion planning

2.2.4 Findings

The previous sections can be concluded by the following findings:

- LIDAR, radar and camera systems are the most commonly used sensors in automated vehicles. As they all have particular strengths and weaknesses, sensor fusion techniques are applied.
- Vehicular communication can be distinguished into infrastructure-to-vehicle and vehicle-to-vehicle. The de facto standard for the latter group is the WiFi-based IEEE 802.11p, while for the first group the cellular 4G/5G system can also be used depending on the application. For intersection safety, I2V systems are more promising in the near term than V2V, as long as the penetration of equipped vehicles is low.
- Satellite navigation and highly-precise digital maps are core parts for the route and motion planning of automated vehicles. For areas without satellite coverage or digital map data, the SLAM method can be applied.

2.3 Accident analysis and safety at road intersections

From the previous sections, it can be seen that a large amount of different technologies is applied to ADS. Evaluating these technologies is a main part of ADS development, reaching from the automotive industry to sensor manufacturers and software developers. Also, road and vehicle safety research is an important field to enable safe automated driving. As intersections are locations, where the paths of multiple traffic participants are crossed, they are considered high-risk spots for safety researchers. For automated vehicles, road intersections of whatever type constitute a major point of interest along their routes due to the increased likelihood of conflicts with other road users. This thesis addresses the problem of road intersection safety with regard to a mixed population of automated and non-automated vehicles. Therefore, the following section comprises an overview of the current situation about safety at intersections, by giving figures of macroscopic crash data as well as results from microscopic analyses.

2.3.1 Types of accident data analysis

Accident data coming from national accident databases, mostly collected from police records, can also be called *macroscopic* data. Macroscopic crash data is commonly used for descriptive statistical analysis aiming at describing general road safety problems, e.g. problems of each road user group. In contrast to microscopic data, this data type does not include detail on causation, but can mostly be considered statistically representative for the region where the data was collected. Examples for macroscopic databases are CARE (Community Road Accident Database, Sanz Villegas, 2011), where all EU member states except Germany are included, or STATS-19, the National accident database from the UK. CARE is a centralised database on road accidents in the European Union resulting in injury or death. It provides access to member states to identify and quantify road safety problems, to evaluate road safety measures and to facilitate the exchange of experience in the field of accident data analysis (Sanz Villegas, 2011). The full operation of CARE started in 1999 and it includes data from the respective National databases of all Member States.

A *microscopic* analysis aims at describing the accident mechanisms with the use of in-depth data. Contrary to macroscopic crash data, this type is rarely representative due to a limited number of cases available. This is because microscopic data is derived from in-depth accident investigations, which provide a much more comprehensive set of crash information. Accident investigation comprises the acquisition of factual information regarding an accident, which can include on-scene elements, elements recorded retrospectively, or both of these. Those investigations are typically done by on-scene inspections immediately after the occurrence of a crash, accident reconstruction and driver interviews. Prominent examples of in-depth databases are OTS (On the Spot, Cuerden et al., 2008) and RAIDS (Road Accident In-Depth Studies, Department for Transport, 2013) in the UK, or the German GIDAS (German In-depth Accident Studies). There is an ongoing project aiming to develop a harmonised European in-depth database, called iGLAD (Initiative for the global harmonisation of accident data, Hainig, n.d.), currently containing sample, mostly non-representative data from ten countries.

The project GIDAS was initiated in 1999 including two regions in Hannover and Dresden, with the aim to provide a comprehensive in-depth crash database for automotive manufacturers, suppliers, public bodies and research institutes. The locations were selected in a way to cover almost all relevant traffic situations and to include a representative sample of collision types. There are three requirements for an accident that must be fulfilled in order to collect accident data by the investigation teams: (1) it is an injury accident, (2) it happened within the two geographical regions and (3) it happened within the predefined daytimes, which alternate weekly between 00:00-06:00 and 12:00-18:00 or 06:00-12:00 and 18:00-24:00. The investigation teams consist of engineers, technicians and medical staff and attend the crash location as soon as it was reported. The collected crash information includes data about the road layout, environmental conditions, traffic rules, vehicle types and their deformations, crash causations, personal data, injury details etc. The on-site inspections are complemented by computer simulations, where the crashes are reconstructed, e.g. to obtain collision speeds. In this way, data relating to approximately 2,000 accidents are collected annually.

In the same year as GIDAS started, the OTS project was commissioned by the UK government to gather in-depth information at the scene of road accidents (Mansfield et al., 2008a). Between the year 1999 and 2010, data were collected in two geographical regions, namely Thames Valley and Nottinghamshire in the Midlands. For the Midlands, a research team from the Loughborough University collected data, while for the Thames Valley region, the Transport Research Laboratory (TRL) was responsible. Expert investigators typically attended accidents spots within 15 minutes after its occurrence and were thus able to collect vital information about the scene.

Data from both teams were collated into a single database that contains more than 2,000 variables.

In 2012, the UK Department for Transport brought together different types of investigation from earlier studies into a single accident data collection programme called RAIDS. It combines existing studies including OTS, CCIS (Cooperative Crash Injury Study) and HVCIS (Heavy Vehicle Crash Injury Study) with new data in a common database. OTS data was used for this thesis and is further described in Section 4.2.1.

2.3.2 Proportion of junction accidents and injury severities

A query from the CARE database was analysed to get a picture of the intersection accident situation in the European Union. In the following sections, accident figures for the years 2003 until 2013 are given for the EU-27. However, eight countries were excluded due to missing values in that decade. It is important to note that junction accidents in this analysis only include the types four-legged crossroad, multiple junction and T or staggered junction, both signalised and non-signalised, and exclude roundabouts and other types.

In Figure 6, the number of road injury accidents at junctions compared to all injury accidents is given per year. It shows the general decrease in accident numbers and that the proportion of junction accidents lies between 27 and 32 percent throughout the decade.

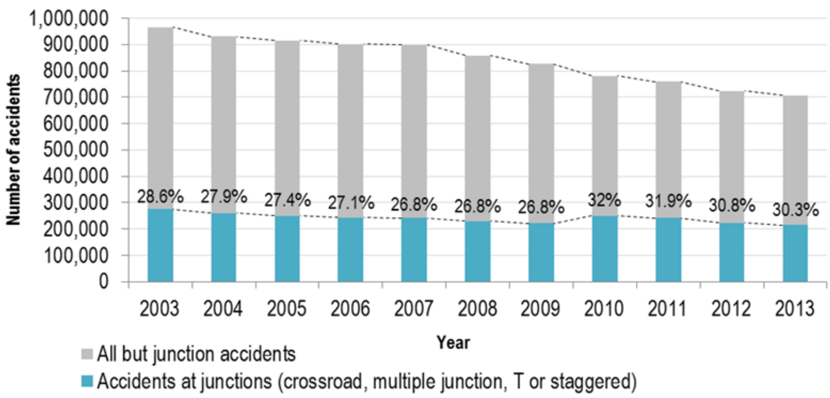


Figure 6: Proportion of road injury accidents at junctions compared to all accidents, by year in EU-27 (excluding BG, CY, DE, EE, LV, SE, SI and SK due to missing data), queried from CARE on 11 Aug 2015 (n=9,269,493)

A similar conclusion can be drawn from the analysis of fatalities, as depicted in Figure 7. The proportion of fatalities remains between 13 and 15 percent throughout the decade, and the total number of fatalities has been reduced to almost the half. In terms of serious injuries, an average proportion of 24.4 percent at junctions can be observed for the years from 2003 to 2013 (see Figure 8).

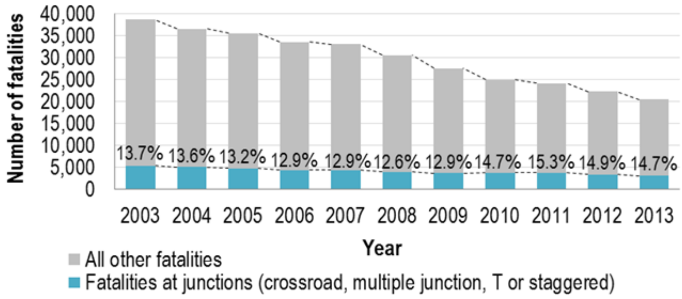


Figure 7: Proportion of fatalities at junctions compared to all fatalities, by year in EU-27 (excluding BG, CY, DE, EE, LV, SE, SI and SK due to missing data), queried from CARE on 11 Aug 2015 (n=327,039)

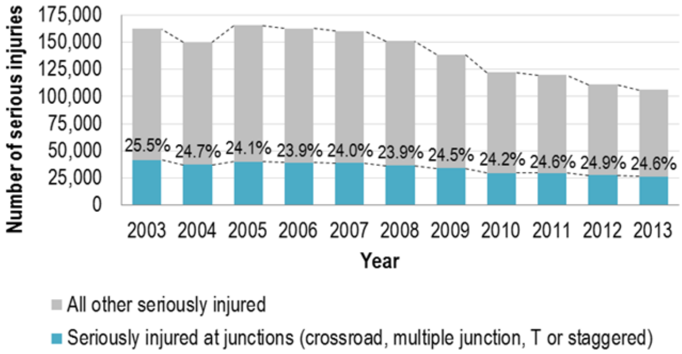


Figure 8: Proportion of seriously injured persons at junctions compared to all seriously injured, by year in EU-27 (excluding BG, CY, DE, EE, LV, SE, SI and SK due to missing data), queried from CARE on 11 Aug 2015 (n=1,551,039)

2.3.3 Differences in junction types

Due to the higher number of conflict points, four-legged junctions were found to be generally less safe than three-legged junctions (Bauer and Harwood, 1996; David and Norman, 1975; Hanna et al., 1976; Harwood, 1995), without distinguishing into signalised and non-signalised junctions. The pie charts in Figure 9 depict the proportion of fatally injured and seriously injured persons by junction type, which confirms this finding. For comparison, roundabouts have been included. Crossroads have the highest amount of both fatal and serious injuries with 43.9 and 43.2 percent, respectively. For both injury levels, roundabouts have the lowest proportion, which can be explained by the safer design principles such as limited points of conflict and lower speed (see Section 2.3.7). However, it must be noted that those percentages also depend on the exposure of different junction types, which has not been further analysed in this review. The category “Other or unknown” comprise the types “Multiple junctions”, “Not at grade”, “Other” and “Unknown”, with “Other” as the most frequent type.

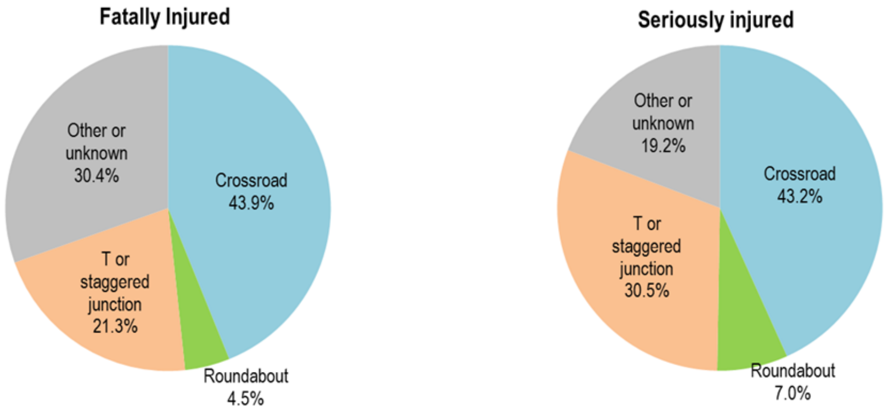


Figure 9: Proportion of fatalities (left) and seriously injured (right) by junction type, in EU-27 (excluding BG, CY, DE, EE, LV, SE, SI and SK due to missing data), years 2003-2013, queried from CARE on 11 Aug 2015

2.3.4 Proportions of different road user types

To get a complete picture of the situation of intersection safety, it is useful to know the modes of transport involved in fatal accidents at intersections. Figure 10 depicts the total proportion of junction fatalities by the different modes. Due to their higher exposure, passenger cars have the highest amount throughout the decade with more than a third of all fatalities at junctions, followed by pedestrians and motorcycles. However, all vulnerable road user types (pedestrian, motorcycle, moped and pedal cycle) combined have a proportion of almost 59.4 percent, which is much higher than for passenger cars.

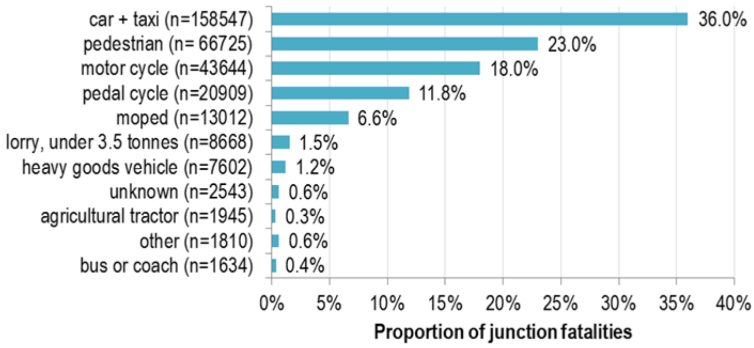


Figure 10: Proportion of fatalities at junctions, by traffic unit type, in EU-27 (excluding BG, CY, DE, EE, LV, SE, SI and SK due to missing data), years 2003-2013, queried from CARE on 13 Aug 2015 (n=327,039)

In contrast to that, Figure 11 gives the percentage of how many persons were fatally injured at junctions compared to all fatally injured in the respective modes, by stretching each bar to 100 percent. The chart reveals that persons on pedal cycles and motorcycles were more often fatally injured at junctions than persons using other modes of transport. Every fourth fatally injured person riding on a pedal cycle was killed at a junction, while only every tenth fatally injured car occupant died due to a junction crash. This means that junctions pose a particular risk for powered two-wheelers, cyclists and pedestrians, although their total number of fatalities at junctions is lower than e.g. for car occupants.

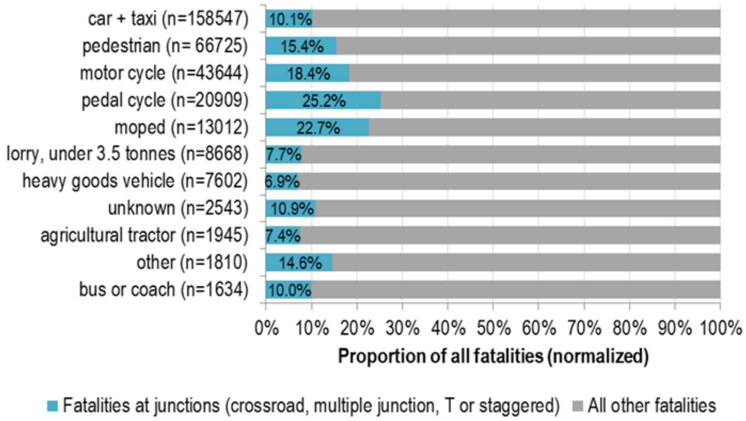


Figure 11: Normalized proportion of fatalities at junctions, by traffic unit type, in EU-27 (excluding BG, CY, DE, EE, LV, SE, SI and SK due to missing data), years 2003-2013, queried from CARE on 13 Aug 2015 (n=327,039)

2.3.5 Relevant collision scenarios

Van Maren (1980) reported that (multi-lane) non-signalised intersections have a lower number of crashes per million conflicts than signalised intersections. For signalised intersections, it was found that the dominant crash types are rear-end and head-on collisions (Obeng, 2007; Polders et al., 2015), however, Abdel-Aty et al. (2006) argues that this also depends on the number of lanes and traffic volumes. In comparison to that, the majority of non-signalised intersection accidents are angle collisions (Arndt, 2003; Layfield et al., 1996; Molinero Martinez et al., 2008; Pickering and Hall, 1985). Given the finding that signalised intersections have a higher number of crashes, it is difficult to say whether those areas are particularly risky for automated vehicles, too. As the other studies showed, the crash types differ between non-signalised and signalised junctions. Future studies are needed to analyse whether ADS are more effective in e.g. avoiding rear-end crashes than angle crashes.

The most important variables affecting the safety of non-signalised intersections were studied by Haleem et al. (2010). Accordingly, these include the traffic volume on the major road and the existence of stop signs, and among the geometric characteristics, the configuration of the intersection, number of right or left turn lanes, median type on the major road, and left and right shoulder widths. In particular for angle crashes at non-signalised intersections, the factors were found to be traffic volume on the major road, the upstream distance to the nearest signalized intersection, the distance between successive non-signalised intersections, median type on the major approach, percentage of trucks on the major approach, size of the intersection and the geographic location within the state (Abdel-Aty and Haleem, 2011).

A study by Molinero Martinez et al. (2008) for EU-27 in 2004 concluded that intersection accidents occur rather in urban regions (64 to 73%), during daylight and good visibility, and with passenger cars driven by male drivers (65 to 76%) having the highest distribution. They further identified relevant intersection scenarios that can build a basis for this thesis:

- Scenarios, where vehicles cross the trajectory of an opponent vehicle that turns left, right or not at all. (70% of all intersection accidents)
- Scenarios involving a rear-end crash, where the struck vehicle intended to turn right or left. (2% of all intersection accidents)
- Scenarios involving a rear-end crash with no special manoeuvre of the struck vehicle (e.g. going straight). (5% of all intersection accidents)
- Scenarios involving a head-on collision of two vehicles going different directions (2% of intersection accidents)
- Scenarios involving a crash on a roundabout (11% of intersections accidents)
- Other scenarios including pedestrian collisions (9% to 15% of intersection accidents)

2.3.6 Typical causation factors

An interesting aspect to understand intersection crashes is the critical event that led to an accident. According to a query from the SafetyNet Accident Causation database (Thomas et al., 2009) based on DREAM causation charts, the events leading to junction accidents differ from those in non-junction accidents (see Figure 12). In this in-depth study, the critical events “Late action”, “Premature action”, “Skipped action”, “Prolonged action” and ‘No action’ are recorded more often than for non-junction accidents. In contrast to that, surplus speed or surplus force (inappropriate for the conditions) as well as incorrect driving direction have much lower distributions compared to non-junction accidents. Similar findings were also published by Sandin (2009), where six pre-defined risk situations at intersections were analysed by using the DREAM method for 52 drivers. It was concluded that the most common causation patterns include missed observation due to distraction or sight obstructions, which then led to no, late or premature action. Furthermore, a common causation was found to be incorrect prediction or faulty diagnosis, e.g. they did not expect another vehicle to cross their path. One the one hand, assisted and automated driving systems such as automated emergency brake or forward collision avoidance are expected to mitigate the safety problems caused by human error (Cunningham and Regan, 2015; Fitch et al., 2014), i.e. through sensing and perception technologies. On the other hand, it is argued that automated driving modes reduce the overall driver vigilance and situational awareness, which makes it difficult to take back the driving control if needed (Merat et al., 2012; Neubauer et al., 2012; Saxby et al., 2013). Also, failures of the sensor systems cannot be excluded and sight obstructions and unexpected road user behaviour still pose problems for ADS.

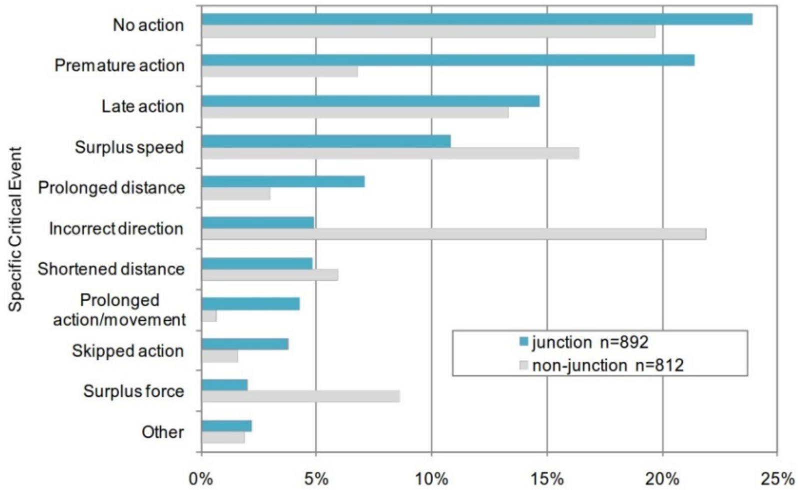


Figure 12: Distribution of specific critical events at junctions (SafetyNet 2005-2008) (Broughton et al., 2013)

Other studies (e.g. Lee et al., 2004; Molinero Martinez et al., 2008; Najm et al., 2001) achieved similar results concerning causation factors and have further shown that failure to yield right-of-way is the most dominant violation in crossing path scenarios. This is followed by running a traffic signal or sign as one of the most frequent violations. It was found that alcohol and drug violations are minor factors compared to the others.

2.3.7 Traffic conflicts as surrogate measures for safety

The previous sections have presented the safety situation at intersections by summarizing recent studies and analysing crash data records. For road safety research, the frequency and severity of accidents have traditionally been used for analysis, because they are direct measures of safety. However, there are well-known quality problems of accident data such as small sample sizes leading to inconclusive results, lack of details or underreporting (Tarko et al., 2009). Therefore, developing non-crash or so-called surrogate measures of road safety have been subject to research over the past five decades. The idea behind is to study near-misses, further denoted as conflicts, which are assumed to happen much more often than accidents and could be used as a complementary source of information for safety analyses.

Traffic conflicts are defined as critical incidents not necessarily involving collisions (Chin and Quek, 1997). The concept of the Traffic Conflict Technique was first published by Perkins and Harris (1967), who observed and counted instances in which cars took evasive actions to avoid a collision. The goal was to find out whether General Motors cars are less often involved in unsafe driving situations than other manufacturers' cars. While their study was industry-driven, this new safety technique gained immediate interest among other researchers, who recognised the potential of conflicts as surrogates for accidents (e.g. Allen et al., 1978; Baker, 1972; Hydèn, 1987; Williams, 1981; Zegeer and Deen, 1978).

So far, there is still a discussion about the understanding of what a traffic conflict is. One the one hand, traffic conflicts can be marked by evasive actions, i.e. that crashes and conflicts are of similar nature except for the conduction of a successful braking or steering manoeuvre to avoid the collision. For example, Parker and Zegeer (1989) stated that “a traffic conflict is an event involving two or more road users, in which the action of one user causes the other user to make an evasive maneuver to avoid a collision”.

On the other hand, a crash does not always involve an (unsuccessful) evasion action, e.g. due to inattention, which somehow weakens the relationship between conflicts and crashes according to the definition above. This led to the development of temporal and spatial proximity indicators, which should detect the closeness of two road users in time and space, even if there was no evasion action. In the first workshop on traffic conflicts in Oslo 1977, it was therefore agreed that a traffic conflict can be defined as “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 remained unchanged” (Amundsen and Hydèn, 1977). This adapted definition implies that there are quantitative indicators to assess temporal and spatial proximity, such as time-to-collision (Hayward, 1972), post encroachment time (Archer, 2005), time-to-accident (Hydèn, 1987), proportion of stopping distance (Gettman and Head, 2003) and others. In this thesis, both traffic conflict definitions are used for the safety performance evaluation and a combination of numerical indicators and thresholds is used, as described in Section 6.5.

In comparison to the collection of accident data, which normally requires several years of recording, traffic conflicts are traditionally collected over a shorter period of time. Earlier studies relied on on-site observations on particular risk spots, typically at intersections, either done by on-site observers or by reviewing videos afterwards (Hydèn, 1987; Parker and Zegeer, 1989). While this method might be more valid than objective measurements, it involves high costs and requires well-trained observers.

With the rise of image processing techniques, computer vision has become an alternative method to collect traffic conflict data. The automated detection and classification of events based on road-side cameras are relatively cost-effective and can be used on a larger scale. However, shortcomings are the fact that high-quality cameras are needed and that tracking algorithms are still not reliable under all conditions (Zheng et al., 2014). In the past years, image processing has become particularly interesting for analysing naturalistic driving data, when combined with kinematic triggers obtained from in-vehicle data (Bagdadi, 2013; Bagdadi and Várhelyi, 2013; Dingus et al., 2006; Valero-Mora et al., 2013; Wu and Jovanis, 2012). Using naturalistic driving data for traffic conflict analyses has the advantage that its longitudinal data helps to understand crash causation and driving behaviour better, because rare events can be studied in detail including information about driver vigilance and reactions.

There have been numerous studies investigating the predictive validity of conflicts, i.e. the statistical significance of the correlation between conflicts and collisions. In fact, this is still an open research question. In their survey paper, Zheng et al. (2014) compared different arguments for and against the validity of the traffic conflict technique. In summary, if there was a poor correlation found, this can either be explained by the inaccurate, underreported crash population itself or by the discrepancy between observation durations for conflicts and crashes. As mentioned, conflict data are usually collected over a much shorter period of time than crashes and hence may not cover sufficient variability in traffic. Researchers who found a strong correlation between crashes and conflicts recommend disaggregating both data sources into specific characteristics such as road type, manoeuvres or severity level.

Given the counts of conflicts and crashes, Hauer and Garder (1986) suggested the following relationship:

$$\lambda = \sum_i \pi_i c_i \tag{1}$$

with λ as the expected number of crashes, π_i the conflict-to-crash ratio and c_i as the number of observed conflicts of severity level i. Various regression techniques were proposed to calculate π_i (Lord, 1996; Lord and Mannering, 2010; Sayed and Zein, 1999), which differ in the accident types analysed.

Another promising approach to estimate the crash frequency from conflicts is the extreme value theory model, which was investigated by Songchitruksa and Tarko (2006) for signalised intersections. It considers interactions between vehicles as surrogate events that are analysed individually. Accordingly, the expected number of crashes $F(C)$ is a product of the number of observed conflicts $F(H)$ and the estimated conditional probability of a crash $P(C|H)$:

$$F(C) = F(H)P(C|H) \tag{2}$$

The probability of a crash is estimated with the fitted extreme value distributions given for the surrogate event H for the specific road location. Hence, crash counts are no longer taken into account, but instead, the method uses more frequent conflicts to predict less frequent crashes.

Simulations can help to automate conflict analysis and to increase the number of "virtual observation" of conflicts. Gettman and Head (2003) investigated the potential for deriving surrogate measures of safety from existing microscopic traffic simulation models for intersections and developed the Surrogate Safety Assessment Methodology (SSAM). To assess a specific road location with SSAM, the location is first modelled in a microscopic environment and then simulated with the desired traffic conditions. Trajectories obtained from each simulation run are post-processed to identify conflicts according to predefined proximity indicators. Gettman et al. (2008) carried out a field validation exercise to compare the output from SSAM with real-world crash records. By modelling and simulating 83 intersections, they found the following relationship:

$$\frac{Crashes}{year} = 0.119 \cdot \left(\frac{Conflicts}{hour}\right)^{1.419} \tag{3}$$

The authors conclude that this relationship provides a good estimate for future studies, because it is consistent with the range of correlations reported in several studies between ADT and crashes for urban, signalized intersections. They further found a conflict-to-crash ratio of 20,000 to 1, although this ratio varied by conflict type.

Despite the long-year research undertaken in the field of traffic conflicts, the relationship between conflicts and crashes is still not solved. It is recommended in literature to use longer conflict observation periods to improve the validity, or to combine on-site observations with simulations. Also, there is little research on the analysis of single-road-user conflicts (e.g. run-off-road) or multi-road-user and secondary conflicts.

2.3.8 General safety principles for intersections

The previous sections were devoted to safety analysis in relation to crashes and conflicts. This section briefly summarises the road design principles to avoid critical events at intersections. There are numerous design manuals and guidelines for intersection layout in different countries (AASHTO, 2012; Austroads, 2009; Fitzpatrick et al., 2004; Golembiewski and Chandler, 2011; “TD 42/95 - Geometric Design of Major/Minor Priority Junctions,” 1995), but most recommendations involve the following safety principles, adapted from the Safe System Intersection design principles published by Candappa et al. (2015):

Limit travel speeds through intersections to 50 km/h

It has been proven that collisions close to 90° impact angle with speeds over 50 km/h exceed the biomechanical tolerance threshold for humans (Bostrom et al., 2008). In order to keep this kinetic energy low, the approaching speed should be limited by safe intersection design instead of just setting new speed limits.

Minimize the points of possible conflicts

Every intersection has a certain amount of conflict points between vehicle paths (see Figure 13), but safer intersection design could avoid many of those. Limiting the points of conflict reduces the risk of a collision. A roundabout is an example, where the points of conflict have been successfully minimised to four to eight depending on the design. In comparison, a typical crossroad junction has 32 conflict points.

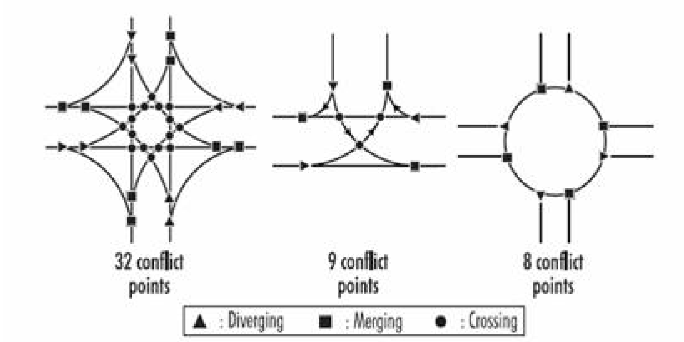


Figure 13: Number of conflict points for examples of intersections (European Commission, 2015)

Ensure sufficient sight distances

Providing appropriate stopping sight distance at intersections is a fundamental principle, especially with no traffic-control device. Each quadrant of a junction should contain a triangular unobstructed area, as depicted in Figure 14. The triangle dimensions depend on the design speed of the intersection as well as the type of traffic control.

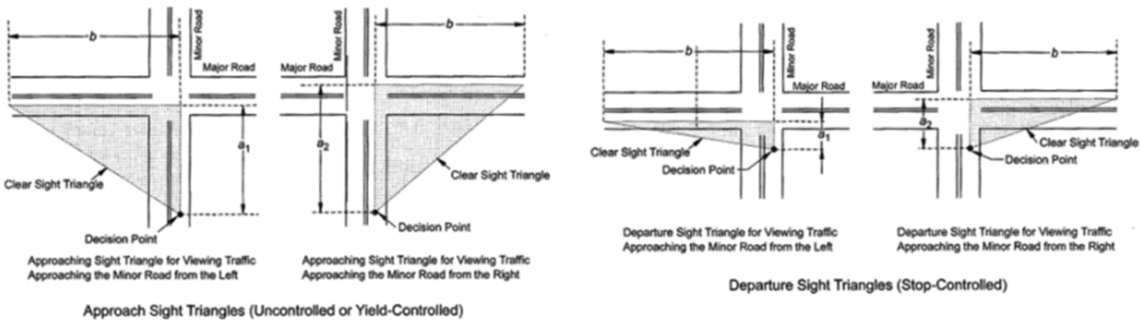


Figure 14: Principle of sight triangles at intersections (AASHTO, 2012)

Avoid 90° impact angle

Optimising impact angles to minimise the kinetic energy for vehicle passengers is crucial for a safe intersection design. At roundabouts, the impact angle is halved and therefore a lower injury severity can be expected.

Physically separate vulnerable road users or provide travel speeds <30km/h

Pedestrians and non-motorized two-wheeler users are often temporally separated from other vehicles, but only a physical separation ensures maximum intersection safety. Where physical separation is not possible, inducing travel speeds to below 30 km/h is recommended.

Raise awareness of intersections

Beck (2015) suggests that the layout, delineation and visibility, when approaching an intersection, must meet a driver's expectations. Warning signs such as "Intersection ahead", improvements to street lighting or treatments to the pavements, e.g. rumble strips, markings are examples to improve awareness.

Exclusive turn lanes

Turning movements constitute a particular risk at intersections. Li and Tarko (2011) found that introducing left-turn lanes can reduce rear-end crashes at junctions by 20 to 40 percent. It must further be ensured that the turn lane is of sufficient length to avoid queues (Beck, 2015).

Optimize signalling

There are several treatments to traffic signals that help to reduce the crash risk at junctions. For example, it has been proven that a separate phase for turning movement can reduce rear-end crash rates (Baldock et al., 2005). Furthermore, the coordination of signals along a route can decrease the frequency of stops and therefore the risk of rear-end crashes (Antonucci et al., 2004; Wang and Abdel-Aty, 2006).

Road designers have introduced various novel intersection types, but they are currently only implemented occasionally. Examples are turbo-roundabout, cut-through roundabouts or raised intersections, but they will not be further discussed in this thesis, because the focus is on evaluating safety on commonly existing road design and not on novel infrastructure measures such as new intersection types.

2.3.9 Findings

The following findings summarise the review of safety aspects of road intersections:

- The proportion of junction accidents among all accidents is approximately 30 percent, while the proportion of fatalities at junctions among all fatalities is around 14 percent.
- Crossroads have the highest amount of both fatal and serious injuries, and roundabouts have the lowest proportion. Signalised intersections have a higher number of crashes compared to non-signalised junctions. However, this does not lead to the conclusion that crossroads or signalised junctions are the major risk spots for ADS, because the effectiveness of ADS on certain crash types, e.g. rear-end or angle collisions, must be proven first. While signalised accidents have a higher proportion of rear-end and head-on collisions, angle collisions are the main crash type at non-signalised junctions.
- Due to their higher exposure, passenger cars have the highest amount with more than a third of all fatalities at junctions. However, relative proportions show that vulnerable road users have the highest risk of fatal injuries.
- The most common causation patterns for human drivers include missed observation due to distraction or sight obstructions, incorrect prediction or faulty diagnosis, e.g. they did not expect another vehicle to cross their path.

While human errors are likely to be reduced by ADS, sight obstructions and faulty sensor perception still pose current safety problems.

- Failure to yield right-of-way is the most dominant violation in crossing path scenarios. This is followed by running a traffic signal or sign as one of the most frequent violations. At the current state of research, there is no evidence to extrapolate this to the safety situations relevant for ADS. While the primary task of a vehicle should be to eliminate its own driving failures, a secondary task should be to avoid or mitigate collisions caused by the violation of others.
- The traffic conflict technique is a well-recognised supplement to traditional crash analysis and has been applied in various studies related to intersection safety. However, there is still no clear evidence on the relationship between crashes and conflicts.
- There are several basic safety principles for intersections, among which are minimizing possible conflict points, ensuring proper sight distance, avoid right angles or separate vulnerable road users from other vehicles. It is subject to research, which additional safety principles should be applied for automated driving at intersections.

2.4 Automated intersection collision avoidance and mitigation systems (ICAMS)

Due to their high number of conflict points and possible critical scenarios, intersections are one of the most challenging spots for ADS. There have been numerous studies and developments of systems, which aimed at avoiding or mitigating collisions at intersections. In principle, ICAMS can be categorised into the following groups (Mages, 2008):

1. *Vehicle-based systems*, solely based on the perception of the in-vehicle sensors
2. *Infrastructure-only systems*, based on roadside sensors and roadside warnings
3. *Cooperative infrastructure-to-vehicle (I2V/V2I) systems*, based on roadside sensors and data communication to the vehicles
4. *Cooperative vehicle-to-vehicle (V2V) systems*, solely based on inter-vehicular communication messages

While this categorisation distinguishes between different technologies, Mages et al. (2015) presented another way of grouping intersection assistance systems based on the assistance function:

a. *Crossing Traffic Assist*, which assists drivers in merging into or crossing a major road.

b. *Turning Assist*, which assists drivers in avoiding collisions with oncoming traffic while turning at a junction.

c. *Stop Sign Assist*, which informs or warns drivers to avoid failure to stop.

d. *Traffic Light Assist*, which assists drivers in approaching a traffic light, primarily to avoid a red light violation.

In context with the scope of this thesis, which is to evaluate the safety of automated driving systems at road junctions, this list of functions is extended with Forward Collision Avoidance (FCA) systems, as they are a core function to avoid frontal collisions. Considering the two ways of categorisation, ICAMS can be structured as given in Table 3. Past developments of stop sign and traffic light assistants were realised by infrastructure-only or V2I/I2V systems. Vehicle-based or V2V systems can also realize the other three categories.

In the following sections, the first three groups of technology categories are further explained by summarising relevant literature. For the reasons argued in Section 2.2.2, V2V is out of the scope of this thesis and will not be further reviewed.

Table 3: Categories of ICAMS

Technologies	FCA	Crossing Assist	Turning Assist	Stop Sign Assist	Traffic Light Assist
Vehicle-based	x	x	x	x	
Infrastructure-only	x	x	x	x	x
V2I/I2V	x	x	x	x	x
V2V	x	x	x		

2.4.1 Vehicle-based systems

Modern vehicles are equipped with collision avoidance and mitigation systems, mostly as an optional feature. Those systems belong to the group of ADAS, which intervene when a potential collision is sensed, either by warning the driver, by supporting emergency manoeuvres or by acting autonomously. This section surveys recent literature of in-vehicle systems relevant for intersection collisions, as they are a core function of ADS.

2.4.1.1 Forward Collision Avoidance

Many accidents are caused by late or insufficient braking, e.g. due to driver distraction, poor visibility or unexpected behaviour of other road users. FCA systems, also known as Automated Emergency Braking (AEB), are designed to detect an imminent frontal collision with other road users or obstacles and to intervene by applying the brakes. Usually, drivers are warned beforehand (Forward Collision Warning, FCW) and some systems deactivate when they detect avoidance action being taken by the drivers. Of course, this step would not be necessary for future fully automated vehicles. Although FCW and FCA systems are similar in terms of sensor technology, they present different challenges. For example, FCW relies on a quick driver reaction and must thus deal with appropriate HMI design. Faulty interventions from FCW systems may be tolerated to a certain extent, while FCA decisions are much more critical and are usually taken at the very last second before a collision (Winner, 2015).

FCA systems are made to avoid frontal collisions, including rear-end collisions as well as collisions with crossing traffic pedestrian and cyclist accidents. Some examples for real-world applications are

- FIAT City Brake Control,
- BMW Pedestrian Warning with City Brake Activation,
- Mitsubishi Forward Collision Mitigation,
- Skoda Front Assistant,
- Audi Pre Sense Front Plus,
- Volkswagen Front Assist or City Emergency Brake,
- Mercedes-Benz Collision Prevention Assist,
- Volvo City Safety,
- Tesla Model S,
- Google Self-Driving Car

FCA (as well as FCW) systems use data collected by the in-vehicle radar, LIDAR, camera sensors or a fusion of those. Radar is the most common sensor type used in FCA systems thanks to its high detection accuracy, robustness, and wide sensing capability (Bloecher et al., 2009). Low-cost cameras can be applied, because they can detect lanes, vehicles, and pedestrians in short range (Stein et al., 2010). To detect the right of way, the vehicles use digital maps and GNSS positioning, camera-based traffic sign recognition or a combination of both (Chen et al., 2011; Lindner et al., 2004; Ruta et al., 2011).

In their survey paper about intersection safety systems, Shirazi and Morris (2017) distinguished between “mild”, “moderate” and “intense” ADAS for intersection collision avoidance. While “mild” ADAS comprise information systems to enhance the driver’s perception, e.g. through in-vehicle displays and maps, “moderate” ADAS goes one step further towards warning systems by visual, audio or haptic means to suggest appropriate driving actions. “Intense” ADAS would be the relevant group for ADS, because they actively intervene in risky situations. As a conclusion, the authors suggest that vehicle-based systems should cooperate with infrastructure-based systems.

In literature, many research papers can be found that deal with rear-end collision avoidance and mitigation methods, algorithms or strategies. One of the crucial tasks for FCA is the assessment of critical states or threats in dynamic road situations. Typical measures to characterise the threats are time-to-collision (TTC, Leonard et al., 2009), headway time (Polychronopoulos et al., 2004) or required deceleration (Karlsson et al., 2004). However, these measures are tailored to avoidance systems for frontal collisions. Intersection situations require more complex measures, since vehicles may approach from different angles. For example, time-to-intersection and distance-to-intersection are useful measures to evaluate the threat of traffic movements at intersections (Chan et al., 2004).

Focussing on road intersections, Hillenbrand and Kroschel (2006) used additional metrics to assess the situation, namely the time-to-react (TTR), time-to-brake (TTB), time-to-kickdown (TTK) and time-to-steer (TTS), which all refer to the time the driver has left to begin with a manoeuvre. This could be translated to fully automated vehicles, e.g. by lowering the thresholds for reaction times. However, this issue was not covered in the paper.

Developing a robust FCA system for intersections requires a prediction model of the future motion of the vehicles (Aoude et al., 2010). De Campos et al. (2014) presented a probabilistic collision detection and decision-making algorithm for safety interventions at intersections. The method comprises 1) a path prediction model based on a Kalman filter, 2) a threat assessment model to detect the risk of a future collision based on vectors defined by reference points on the vehicle's structure and 3) a minimally invasive intervention protocol that triggers an emergency brake as late as possible. Although experimental results showed promising reaction times and efficiency, the research does not take into account road geometry and lane topology. Moreover, only braking and no steering manoeuvres are considered.

Aoude et al. (2010) combined an intention predictor based on support vector machines with a threat assessor using rapidly-exploring random trees. This threat data was used to evaluate the safety of several possible escape paths at busy intersections, by maximising TTC and thus minimising the threat level. Through simulation and experimental results with a small autonomous vehicle prototype, Aoude et al. (2010) demonstrated that their threat assessment algorithm can be used for real-time applications. A limitation of this research is that it was only conducted on a single type of intersection.

Not necessarily targeted at road intersections, Li et al. (2014) presented a fuzzy-based active control strategy to avoid or mitigate rear-end crashes in low-speed urban traffic. A crisis index was introduced to evaluate the risk of a crash with the following vehicle. A fuzzy logic based controller was designed to identify the crisis state based on the velocity of the vehicles and the distance between the vehicles. Ultimately, the research focussed on finding the right appropriate deceleration in certain driving conditions based on the evaluated crisis state, which may be useful information for vehicle actuators.

An et al. (2014) proposed a warning system for rear-end collisions by using Linear Discriminant Analysis (LDA). Features for the LDA are the state of the vehicles and the remaining TTC, which were projected into linear space to indicate alarm thresholds for collisions. Simulation studies resulted in an average alarm time of 1.86 seconds before the collision, which would be sufficient to activate the brakes in most cases. The paper only focusses on the algorithm development, apparently assuming ideal sensor signals, because they do not mention a specific set of in-vehicle sensors.

Kim and Jeong (2014) used Monte Carlo simulations, i.e. a stochastic sampling technique for a given probability distribution, to assess the crash probability for FCA systems in three different scenarios, namely rear-end, cut-in and T-bone collision. To detect an imminent collision, varying driver behaviour models characterised by steering and braking as well as vehicle dynamics are taken as inputs for the Monte Carlo simulations. A threat assessment algorithm combined with a vehicle tracking algorithm compute the point of no return and warning times in different scenarios. The authors discriminated between collision and near-miss cases, which are also considered critical for drivers or passengers. Related to the objectives of this thesis, this approach is highly relevant, as it includes various safety-critical situations and driver models. However, it is limited to three scenarios, which would not be sufficient to evaluate the safety performance on intersections in detail. Further evaluations are necessary to include more scenarios.

An FCA system based on emergency steering instead of braking has been studied by Keller et al. (2014). A path-planning algorithm for steering torque overlay in critical rear-end situations was developed by using a 5^{th} order polynomial. The polynomial's coefficients are determined by the bounding conditions that are implied by the traffic situation and the course of the road. A drawback of the paper is that it does not include automated braking. Future ADS might use a combination of braking and steering actions to avoid a collision. However, this depends on many parameters of the situation. The work does not target intersection in particular, so the FCA algorithms would have to be validated in more complex situations.

Research has been carried out to develop robust methods for detecting vulnerable road users, which is particularly important for the safety at road intersections. The challenges are to distinguish pedestrians and cyclists from each other or from other objects as well as to predict their future path of motion. Many approaches to tackle these challenges are based on image processing, such as the method proposed by Chien et al. (2013). Their method detects pedestrians with a single camera mounted in the vehicle and includes the vehicle‘s motion, lane types, boundary detection as well as the driver's visual focus (i.e. line of sight) to assess the situation. The driver's line-of-sight is analysed by a fuzzy-rule-based system using an interior camera. Pedestrians were detected successfully in more than 93 percent of the test cases. It is not explicitly mentioned in this paper, but there may be limitations in severe weather conditions or poor light conditions.

Fernandez Llorca et al. (2011) presented a pedestrian collision avoidance system for automated vehicles, based on a stereo-vision based system that detects pedestrians and computes the TTC. The proposed method implements a fuzzy-control based automated steering manoeuvre to avoid the collision. Lateral displacement and the actual speed of the vehicle are used as fuzzy inputs, and the steering wheel position is the controller's output. The setup with an experimental vehicle showed promising results at speeds of up to 30 km/h. For higher speeds, the collision avoidance performance was not acceptable. Furthermore, the study is restricted to the case where the vehicle drives in the right lane, the pedestrian has to be in the same lane and the left lane has to be free and long enough.

A prototype active pedestrian protection system was introduced by Eckert et al. (2013), which executes both steering and braking to avoid collisions. A stereo-video camera combined with a radar sensor recognises pedestrians as well as available manoeuvring space and can predict pedestrian movements. The system achieved promising results in 135 vehicle-pedestrian situations. However, the authors suggest future research on detecting pedestrian of a more generic appearance, e.g. sitting in wheelchairs or pushing a trolley.

A similar system was presented by Hayashi et al. (2013), combining radar and stereo-video cameras including infrared vision at night time. The system detects vehicles, pedestrians and other objects, estimates the collision probability and activates the brakes automatically. Parameters such as the position, speed and predictive courses of the vehicles and pedestrians are taken into account.

Coelingh et al. (2010) demonstrated a robust AEB system for both rear-end and pedestrian collision scenarios. The authors raise an interesting point, namely that the risk of automated emergency braking has to be minimised in non-collision situations, such as pedestrians on the kerb, which do not enter the road. The issue of false alarms in complex scenarios is still one of the main research gaps.

In the European project PROTECTIVE, a method to assess the risk of collision with a pedestrian was developed with the aim to minimise false alarms (De Nicolao et al., 2007). Pedestrian detection was done by onboard LIDAR, radar and video fusion. The risk assessment is based on extensive offline Monte Carlo simulations conducted to determine collision probability functions in certain situations. The system has been successfully assessed with real drivers in demonstrator vehicles. However, it was not designed for ADS and may need to be adapted.

2.4.1.2 Crossing and turning assistance

The previous section summarised methods to avoid or mitigate forward collisions, either rear-end or vulnerable road users collisions. Another significant aspect for intersection collision avoidance systems is assistance to avoid turning or angle collisions. In literature, most approaches to avoid those types of collisions are based on roadside sensors, which are explained in Section 2.4.2 and 2.4.3. Nevertheless, there have been several developments of vehicle-based systems that become active before an imminent collision. The following paragraphs give an overview of recent literature in that field.

One of the main factors that lead to collisions at non-signalized intersections is misjudgement of the speed and acceleration of the vehicles in the cross traffic stream (Pierowicz et al., 2000). Dabbour and Easa (2010) developed a collision warning algorithm for non-signalised intersections, which can detect imminent angle and turning collisions at semi-controlled intersections by utilising a pair of sensors, either radar sensors or laser scanners. Approaching vehicles are detected and their speed and acceleration characteristics are determined. Ultimately, the driver can be aided in finding the right gap in crossing the junction. Further research is necessary to evaluate the functionality of the radar/laser detectors under adverse weather conditions. For vehicles crossing non-signalised intersections autonomously, the appropriate judgement of approaching vehicles and gaps must be a core function.

Another approach for mitigating unavoidable collisions for crossing traffic was proposed by Heck et al. (2013). The authors developed a system based on a mono-camera, which performs an automated emergency brake and/or steering manoeuvre when a collision is detected. This study, along with others (Klappstein et al., 2007; Sato et al., 2011) has demonstrated that low-cost camera-based systems perform reliably in detecting crossing objects. However, a robust mitigation system may intervene differently according to the type of the crossing vehicles and thus has to classify those types and set appropriate intervention actions. In practice, a fusion of video cameras and LIDAR/radar is preferable.

In the European project INTERSAFE-2, an onboard stereo-video system was combined with a LIDAR sensor to perceive and assess collision risks with road users at an intersection (Aycard et al., 2011). A risk assessment module takes as inputs the position and speed of the host vehicle along with the list of detected objects and delivers an estimation of the collision risk between the host vehicle and objects around based on the TTC. This is done by a trajectory prediction algorithm for detected vehicles. Pedestrians are modelled as circles with a predefined radius, because detecting their movements were reported as difficult.

Brännström et al. (2011) presented a real-time implementation of collision avoidance for both rear-end and intersection angle collisions. A sensor fusion system is used to estimate the motion and properties of surrounding road users and objects. Then, a decision-making algorithm based on a model predictive control approach uses the estimates to assess the traffic situation and may brake or steer autonomously. However, one major drawback of this approach is that it did not use real in-vehicle sensors, but positions of surrounding vehicles communicated by WiFi. The argument was that they focus on the development of the risk assessment method first, before suitable onboard sensors are available.

Chakraborty et al. (2011) investigated this idea by developing a control mechanism for T-bone collisions, involving a rapid yaw rotation of one of the vehicles. The manoeuvre is posed as an optimal control problem, whose numerical solution yields the optimal control strategy. The study utilises differential braking and active differential, which means that the torque of the wheels can be controlled individually by an automated system. A follow-up paper by Chakraborty et al. (2013) extends the previous work by modelling additional vehicle dynamics and by utilising conventionally available control commands such as conventional braking and steering. The experimental results confirm the existence of an "option zone" for some situations, within which a sudden yaw rotation manoeuvre may be possible and perhaps even preferable to straight braking.

An intersection collision avoidance and mitigation algorithm that also takes into account near-misses in addition to imminent collisions was proposed by Kim and Jeong (2014), whose paper is explained in the preceding section.

2.4.1.3 Stop sign and traffic signal assistance

There has been little research regarding vehicle-based systems to avoid violations at stop signs or traffic signals, since most of the recent approaches utilise V2I/I2V communication. However, an important study to mention was published by Lee et al. (2004). They investigated an intersection crash avoidance and violation warning system that consists of four functional subsystems: Positioning, in-vehicle sensors, a dynamic warning algorithm and a driver-vehicle interface. The traffic signal warning function included an additional subsystem, namely a communication link to the traffic signal interface to determine signal phase and timing. This relatively early work was further developed in the American research programme CICAS (Cooperative intersection collision avoidance systems) carried out between 2006 and 2009, which will be explained in the later sections.

2.4.2 Infrastructure-only systems

In addition to vehicle-based systems, roadside units can be used to detect and warn traffic participants at intersections. Infrastructure-only systems are independent from vehicle technologies as they do not exchange data with vehicles. Information to road users is given by Variable Message Signs (VMS) or simple warning lights. For example, if the roadside sensors detect an approaching road user or a potential misbehaviour such as failure to stop at a stop sign, warning signs are displayed to the other road users. Radar, video or induction loops may be used as infrastructure sensors (Chan et al., 2004; Frye, 2001, pp. 41–46).

Decision support for rural, non-signalised intersections was investigated by Alexander et al. (2007). In their system to support drivers in turning into a crossing road, roadside radar sensors are used to detect position and speed of surrounding vehicles. Cameras recognise vehicles waiting to turn and an appropriate gap is determined. Drivers waiting at the junction are informed by additional signs, e.g. as a supplement to a stop sign.

Chan et al. (2004) focussed on threat assessment for signalised intersection to give warnings to the drivers through an infrastructure-based or vehicle-based display. For example, a driver intended to make a permissive turn will have to judge how fast a vehicle coming from the opposite direction is approaching. They used two measures for evaluating potential threats of traffic movements towards a signalised intersection, namely time-to-intersection (TTI) and distance-to-intersection (DTI). Their paper concluded that the simple TTI is a good indicator for threat assessment, but with its limitations. Pulsing traffic streams and wave behaviour, such as at the beginning of green, TTI may be fluctuating or changing quickly. Therefore, they suggest combining it with DTI to take the distance gap into account. Furthermore, those indicators should be evaluated by considering traffic signals. Since this paper focussed on the threat assessment algorithm, implementation of the system in real-world scenarios was foreseen as future work.

Infrastructure-only systems can be used for all of the five assistance functions listed in Table 3. Automated vehicles may use such infrastructure-only systems to support their environment perception. However, most recent approaches tend to utilise V2I/I2V communication, as the next section explains.

2.4.3 Cooperative infrastructure-to-vehicle systems

From the previous sections, it can be concluded that both vehicle-only and infrastructure-only systems have their limitations when avoiding collisions at intersections. It is commonly agreed that vehicular communication, either between vehicles or between vehicle and infrastructure, is crucial for future automated road transport. I2V-based intersection systems can be used for all of the five assistance functions listed in Table 3. In the following, the most relevant research studies are summarised.

The American cooperative research programme CICAS aimed to reduce the number of crashes and fatalities at intersections by equipping dangerous intersections with sensors, warning systems and communication units. The program was divided into three functional segments based on the crash type, namely 1) Stop Sign Assist, 2) Signalized Left Turn Assist and 3) Violation of stop signs or traffic signals. The functions correspond to four of the categories given in Table 3, because the FCA system was not included in the programme. The CICAS functions were validated in a field operational test, where the main functionality as well as the drivers' acceptance and reactions were assessed. After all, the CICAS programme focussed on driver warnings and appropriate driver interfaces, either displayed on the roadside or on-board. Hence, the projects were not carried out to assist automated vehicles, which would mean that only active intervention in the driving control must be considered.

One of the largest recent European project dealing with cooperative intersection assistance was INTERSAFE-2, which was completed in 2011. The project combined warning and intervention functions that were demonstrated on three vehicles. The system architecture includes both vehicle and infrastructure systems, linked by V2I/I2V communication. Four vehicle support functions were investigated: 1) Right turn Assistance to prevent VRU collisions, 2) left turn assistance, 3) assistance when crossing a priority road and 4) traffic light assistance to prevent red light violations (Roessler, 2010)

2.4.3.1 Forward Collision Avoidance

As reviewed in Section 2.4.1.1, the application of FCA systems is not restricted to intersections, because e.g. rear-end collisions can happen at any part of a road.

However, rear-end collisions pose a major problem before entering a junction, and therefore, cooperative intersection assistance can help to warn or intervene in critical situations. For example, V2I/I2V messages can be distributed within an intersection's surrounding area to warn other drivers. The warning message can be triggered by a vehicle's braking system or roadside sensors (Le et al., 2009).

Milanes et al. (2012) presented an approach to avoid rear-end collisions by using two fuzzy controllers, namely an FCA and an FCW controller. The former was developed to generate an output control signal for the steering wheel in order to avoid the collision. The inputs provided to the system are the vehicle speed and the displacement required to perform the manoeuvre safely. Based on I2V communication, the system detects a potential collision and performs an evasion manoeuvre without leaving the road. Although the proposed system might be useful to avoid rear-end collisions at intersections, there is more research needed in more complex situations, e.g. when the leading vehicle performs a steering manoeuvre or when driving on a curvy road stretch.

A cooperative FCA system that focusses on collision with crossing pedestrians is proposed by Köhler et al. (2013). The active pedestrian protection system was developed for urban traffic scenarios to perform an autonomous evasive manoeuvre if the intention of a pedestrian to cross the road is detected. This is realised by 1) a mobile roadside unit with a camera and an image processing engine, which can be placed at hazardous spots to observe the sidewalk or the parking lane, and 2) an onboard unit to receive the message of the warning algorithm from the roadside unit. For future automated road traffic, this approach seems promising, although the system has only been tested with one pedestrian at a time. Future work may be to enhance the method for detecting multiple pedestrians or other road users, such as bicyclists.

2.4.3.2 Crossing and turning assistance

There have been numerous developments regarding a warning system based on wireless communication, particularly for intersections. Usually, roadside sensors monitor oncoming traffic and a warning system provides information to the drivers when there is a high collision probability. Some approaches determine the risk for multiple collision types (Basma et al., 2011; Basma and Refai, 2009), while others focus on particular collision types.

The CICAS signalised left turn assist aims to address collisions caused by vehicles making left turns at signalised intersections where there is no protected left-turn signal (Misener et al., 2010). The system helps to judge the gaps in oncoming traffic and also informs drivers when other users, such as pedestrian and cyclists, pose hazards. The left turn assist combines roadside sensors such as radar and cameras, roadside messaging signs, communication and positioning units, dynamic maps and traffic signal interfaces.

If there are no traffic lights present, other approaches need to be taken. Lee and Park (2012) developed an agent-based algorithm that could be called cooperative virtual traffic lights. The algorithm was designed to manipulate individual vehicles' manoeuvres so that all vehicles can safely pass a non-signalised intersection without collision. The assumptions are that all vehicles are automated and thus controllable, that all vehicles are equipped with communication devices and that there are no communication latencies. The agent projects the vehicles' trajectories and computes the intersection point of two vehicles. In case of a possible collision, the trajectories are adjusted. The approach assumes a centralised intersection control agent, which could be realised by infrastructure-based roadside units. However, considering the scope of this thesis with a mixed vehicle population, it is unlikely in the near future that all vehicles approaching an intersection are automated.

An intersection agent that takes vehicular status information from vehicle agents and learns, detects and warns collisions at a road intersection was proposed by Salim et al. (2008). In their paper, they tackle intersection collision problems entirely instead of focussing on a particular collision type. They introduced a learning component that enables the collision warning system to adapt to different types of intersections taking into account varying collision patterns (Salim et al., 2007).

Dabbour and Easa (2014) proposed an infrastructure-based collision warning system to aid right-turning drivers at stop-controlled rural intersections. The system utilises a roadside radar that measures the location, speed, and acceleration of the approaching vehicles on the major road. The algorithm evaluates if there will be any potential collision between the approaching and the turning vehicles and warns the driver of the turning vehicle if such a conflict is found. Computer simulations were conducted to assess the performance of the system, so there was no real-world test included. However, the turning vehicle's acceleration profile was derived from real-world measurements.

A system developed in the European project INTERSAFE was called "Intersection Assist", which aimed to inform drivers about potentially dangerous situations when turning into an intersection (Chen et al., 2007). This was achieved by using path prediction of road users based on laser scanner data and I2V communication. The follow-up project INTERSAFE-2 further developed this assistant. Fusion techniques for in-vehicle and roadside sensor data and V2I as well as V2V communication were investigated to interpret the scene, to assess the collision risk and to perform warning or intervention actions (Roessler, 2010).

One of the research objectives in the project of Donath et al. (2007) was the design and implementation of a rural intersection decision support system that determines the safe gaps in traffic and communicates this information to the drivers intending to enter a major road. One of the conclusions was that the thresholds for a safe gap should be individualised to the driver, or in the case of automated vehicles, the passengers. In their work, they used a single threshold based on behaviour data from older drivers, which is clearly a limitation. Overall, it can be expected that passengers of automated vehicles will have different acceptance levels for safe gaps when crossing or turning into an intersection. This must be taken into account, especially when actively performing an intervention action by accelerating and steering.

2.4.3.3 Stop sign and traffic signal assistance

Stop sign assistants and traffic sign assistants are designed to avoid a red light or stop sign violations and subsequent collisions. In cooperative intersection systems, the technologies are similar and mainly differ in the information given to the driver. Therefore, those systems are reviewed together in this section.

Le et al. (2009) suggested that V2I/I2V communication can be used to inform approaching vehicles about the traffic light status and the remaining time until the status changes. They outline that when traffic lights become dynamic, e.g. due to inductive loops or push buttons, a real-time communication system is necessary.

The CICAS function to avoid stop sign and traffic signal violation assists drivers in avoiding crashes in the intersection by warning the driver of an impending violation (Maile et al., 2008). Equipped vehicles approaching an equipped intersection receive messages about the intersection geometry, GPS differential corrections and status of the traffic signal. A warning algorithm in the on-board processing unit determines whether the driver is predicted to violate the signal and issues a warning by a visual icon, a brake pulse or an audio signal.

The CICAS stop sign assistant for rural intersections aids drivers in deciding when to proceed onto or across a major road after stopping at a rural road stop sign. This is realised either via animated display signs or wireless communication (Le et al., 2009). A field trial study compared different warning systems, both in-vehicle and roadside-based warnings. The use of an in-vehicle stop sign assistant resulted in improved intersection crossing performance, measured by an increased likelihood of making a complete stop at the stop sign and a decreased probability of accepting critical gaps (Becic et al., 2012).

The project INTERSAFE-2 also focussed on traffic light and stop sign assistance. The systems warn or intervene with a brake jerk, when a violation is predicted or when the driver is approaching the traffic light or the stop sign with a speed that is too high (Schirokoff et al., 2012).

Jang et al. (2012) presented a cooperative intersection collision warning system for non-signalised intersection without a stop or yield sign. Multiple sensors located at intersection approaches monitor the vehicles' location and speed, taking into account sight distance relationships. In contrast to most of the other approaches, they utilised the traffic conflict technique, i.e. near-miss detections that are included in the dangerous situation forecast. Microscopic simulation models were conducted to assess the system performance, which resulted in promising prediction rates for potential conflicts. However, the study lacks to incorporate detailed varying driver behaviour such as acceleration patterns or reaction times. Nonetheless, the use of near-miss indicators can be seen as useful decision support for automated vehicles, since not only collisions must be avoided, but also almost-collisions (e.g. close contact to opposing traffic) that certainly have a negative impact on the passenger's comfort.

2.4.4 Findings

The previous sections have surveyed recent literature on intersection assistance and collision avoidance, distinguished into vehicle-based, infrastructure-only and cooperative I2V/V2I systems. All the studies reviewed so far, however, show certain limitations when considering the safety aspects required for robust operation of automated vehicles at intersections. Despite the many projects and research studies in the field, there is still no commercially available intersection assistance system for automated vehicles on the market. The following remarks summarise the review:

- Forward Collision Avoidance systems are the most advanced driving assistance systems available to avoid or mitigate collisions. However, current solutions are primarily tailored to rear-end and pedestrian collisions. For intersections, detecting and avoiding angle or turning collisions must be a key feature of ADS.
- Instead of relying on a single sensor technology, a fusion of data from different sensors is preferable. Vehicle-based as well as roadside-based collision avoidance systems mainly use a combination of LIDAR, radar and stereo-video. Traditional technologies such as induction loops may be a useful supplement.
- Motion prediction of vehicles and other road users is a crucial function of intersection collision avoidance and mitigation systems. Numerous prediction models exist, but further research is needed to make them reliable in complex environments or under adverse weather conditions.
- An intersection collision detection system must be able to adapt to different intersection types. It has been found that most studies do not validate their models on a large variety of intersection layouts and characteristics.
- Various collision detection algorithms have been proposed, but the coverage of these algorithms is limited to only a few scenarios. Validation and verification of the systems must take into account the most critical combinations of collision parameters. Detailed accident analyses may be necessary.
- Infrastructure-only systems are mainly designed to inform and warn human drivers. For automated vehicles, vehicular communication technologies are preferable, since active steering, braking and acceleration intervention is required.
- Cooperative roadside-based I2V/V2I systems help to enhance the vision of automated vehicles and make it possible to inform the vehicles about approaching road users that are obstructed by parked cars, buildings or other obstacles.

- Instead of solely focussing on collisions, it is also recommended to consider near-misses as threat measure. Near-misses do not lead to a collision, but may result in discomfort for vehicle passengers and should therefore be avoided, too.

2.5 Current testing procedures for automated vehicles

This thesis aims to provide critical scenarios for testing automated vehicles. Hence, current testing activities and virtual testing procedures are reviewed in the following sections.

2.5.1 Testing on public roads

Many countries are signatories to the Vienna Convention (UN, 1968) on Road Traffic and have also ratified it. The Vienna Convention originally states that "every moving vehicle or combination of vehicles shall have a driver" as well as that "every driver shall at all times, be able to control his vehicle". An amendment to the Vienna Convention entered into force on the 23rd of March 2016. Accordingly, "the driver, still in control of the vehicle, can be helped by a system under some conditions" as long as the system can be overridden or switched off by the driver (UK Department for Transport, 2015). However, it has been argued that a further change is needed to allow automated vehicles on the roads in various countries. The UNECE regulation No. 79 states that whenever automated steering becomes operational, this shall be automatically disabled, if the vehicle speed exceeds 10 km/h by more than 20 percent. This is certainly a hurdle, which would require another amendment. At the time when this thesis was compiled, a proposal for the amendment of regulation 79 was under review.

Several European countries have already started in-depth research and have taken action to address the legal challenges of automated vehicles, some of which are introduced below. Mostly, special permits granted by the government are currently necessary to allow public road testing of ADS.

As the first European country to allow large-scale testing of automated vehicles on public roads, the Netherlands commenced the Dutch Automated Vehicle Initiative (DAVI, Happee, 2015) in 2015. This public-private partnership aims to investigate, improve and demonstrate ADS at SAE level 3 to 5 on public roads. The initiative involves several research projects covering safety, vehicular connectivity, human factors and the exploration of legalisation including type-approval procedures for automated vehicles aligned with the applicable EU and ECE forums.

In Germany, vehicles have been tested with varying levels of automation. Test drives require a special permit granted by the individual Federal states. However, the German minister for transport announced at the beginning of 2015 that the A9 motorway between Munich and Berlin will be equipped with technology to allow automated vehicles to drive on the road and communicate with other vehicles and the road infrastructure (Eckardt, 2015). This has been realized by the Digital Motorway Test Bed project running until 2022 (BMVI, n.d.).

The current Swedish vehicle legislation, driver's license and liability rules may need amendments to permit the testing of highly automated vehicles. However, the launch of the joint initiative "Drive Me – Self-driving cars for sustainable mobility" (Volvo Car Group, 2013), endorsed by the Swedish Government allowed testing of 100 self-driving Volvo cars on selected roads in and around the area of Gothenburg.

The UK government has invested significantly in the development of ADS, which led to numerous R&D projects (CCAV, 2017). In 2015, four test trials were started: in Milton Keynes and Coventry as part of the LUTZ pathfinder project (Burn-Callander, 2015) in Bristol by the Venturer consortium and in Greenwich, where the Gateway project (TRL, 2014) tested automated electric shuttles buses, plus robotic valet parking for automated cars. The three-year project MOVE_UK started in August 2016 and aims to accelerate the development and market readiness of ADS. This will be achieved by trialling a small fleet of self-driving cars in real-world conditions on the roads of Greenwich, London.

Other European countries such as Austria, Finland, France, Italy and Spain are also currently conducting or preparing experiments and field trials on public roads. For example, the Italian PROUD (Public Road Urban Driverless) car test 2013 demonstrated a fully autonomous vehicle around the city of Parma. This was carried out in a mix of rural, motorway and urban traffic, but requiring a police escort at all times and a passenger ready to take over in case of an emergency. With projects such as SARTRE (Robinson et al., 2010), Citymobil (van Dijke, 2011) and Citymobil 2 (Roberts, 2015), the Spanish Government has invested in an outdoor test track for testing the most advanced technologies. Austria has recently initiated test regions for ADS for different forms of automated road traffic such as passenger light vehicles, public buses and freight transport.

Currently, consortia from different European countries are collaborating in various research projects as given in Table 1. The projects HAVE-iT (Hoeger et al., 2008), InteractIVe (Etemad, 2013), and AdaptIVe (Amditis and Ghosh, 2014), CityMobil (van Dijke, 2011), CityMobil2 (Alessandrini et al., 2015) or PEGASUS (Hallerbach et al., 2017) are examples of large European projects, where various levels of automated driving have been and are currently being tested. AdaptIVe started in January 2014 and aimed to demonstrate automated driving in complex scenarios, as well as define and validate specific evaluation methodologies.

In InteractIVe, several automated driving functions such as Curve Speed Control, Lane Change Collision Avoidance, Emergency Steer Assist and others have been tested for different use cases: rear-end, head-on or blind-spot collisions, collisions with VRUs, lane departure accidents, traffic rule violations. Seven demonstrator vehicles were used for real-world testing, as well as various simulators for virtual testing and evaluation. Similarly, they employed their own-developed testing procedures.

In HAVEit, use cases included normal driving in a lane and activation of different automation levels possible (from Driver assisted to Highly Automated), driver unresponsive and transition to Minimum Risk Maneuver, driving and speed limit change, etc.

As one of the recent ongoing projects, the German research project PEGASUS focuses on quality criteria, tools and methods for testing the highway chauffeur application. This includes the determination of safety level through assessment of the probability of occurrence and mechanical manageability in critical situations, determination of critical traffic situations and quality measures (Lemmer, 2017).

In the United States of America, Nevada was the first state to authorise the operation of automated vehicles on public roads in 2011. Currently, more than twenty states have passed legislation related to automated vehicles and others already issued executive orders (status from 27 May 2017). However, the NHTSA does not recommend permission to operate self-driving vehicles for anything other than testing purposes at the moment. Nevertheless, in the absence of a clear European legal framework, in September 2014, vehicle manufacturers such as Daimler and Audi obtained two permits each for testing self-driving vehicles in California. Google won 25.

In Japan, Toyota stated that “an infrastructure-cooperative type of automated driving” is a priority and manufacturers such as Honda and Nissan have also planned to release various ADS, such as motorway assistant and fully automated parking. The Japanese government announced that ADS will be tested on public roads and highways between 2017 and 2019, inviting foreign car manufacturers to participate. The test area comprises a road network of 300 km (“Automated driving systems to be tested on Tokyo roads from 2017,” 2016). A “Strategic Innovation Program” is underway in Japan, aimed at introducing next-generation ADS by the target year of 2020, when Tokyo is scheduled to host the Olympic and Paralympic Games.

2.5.2 Virtual testing methods

For testing assisted and automated driving functions, virtual testing is being used by car and sensor manufacturers, as it can decrease costs in the development cycle (see Figure 15). In contrast to real-world testing, where only a limited number of testing scenarios can be evaluated involving high costs, virtual testing using computer simulation models can comprise high numbers of scenarios with combinations of varying factors leading to more flexibility and repeatability. However, the fidelity of real-world tests is higher. In practice, automotive developers use a combination of simulation, laboratory experiments and real-world testing.

V ... Virtual R ... Real world	MIL	SIL	ECU HIL	ECU+Sensor HIL	Driving VIL	Dynamic VIL	Real World
ECU-Code / Function	V	R	R	R	R	R	R
ECU	V	V	R	R	R	R	R
Sensor	V	V	V	R	R	R	R
Vehicle	V	V	V	V	R	R	R
Vehicle Dynamics	V	V	V	V	V	R	R
Driver	V	V	V	V	V	R	R
Static road environment	V	V	V	V	V	V R	R
Surrounding traffic	V	V	V	V	V	V	R

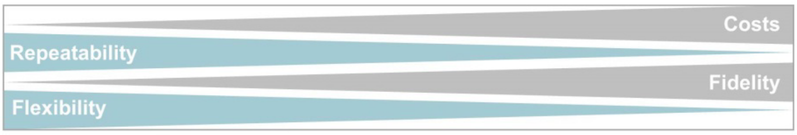


Figure 15: Different levels of vehicle testing (source: Valeo Germany)

Apart from driving simulator studies, there are four categories to simulate traffic, namely sub-microscopic, microscopic, mesoscopic and macroscopic, according to the level of detail from high to low, respectively (Aria, 2016).

Sub-microscopic refers to the most detailed scale of investigation, because single vehicles are simulated using physical models, e.g. for tires, suspension, engine or sensors, and their interaction with the surrounding road area can be studied. The next lower scale of detail is *microscopic*, where individual vehicles are simulated in a traffic stream to investigate the effects on traffic performance due to changes in infrastructure or driving behaviour (Olstam, 2009). In comparison to sub-microscopic simulation, vehicle models are simplified, but a larger road area can be studied. Prominent software tools for microscopic simulation are VISSIM, AIMSUN or the open source software SUMO. *Mesoscopic* and *macroscopic* models have the lowest level of detail and are used to investigate travel demand between origin and destination. For the scope of this thesis, sub-microscopic simulation is the most useful approach, since automated vehicles and their interaction in specific situations are investigated. Moreover, simulated scenarios are entirely quantifiable, controllable and reproducible. There are multiple simulation tools currently on the market that allow testing of various advanced driver assistance systems, as well as ADS. In the following, some of the most common sub-microscopic simulation tools are introduced.

The software CarMaker distributed by the company IPG is a vehicle simulation software with a variety of additional functionality and tools that can be used to implement complex simulation cases. All models are real-time capable and support the model-based development from MIL (Model-in-the-loop), SIL (Software-in-the-loop) through HIL (Hardware-in-the-loop). Several modules are available to construct a simulation model: IPG Driver, IPG Road, IPG Traffic, etc. Representations of the road can also be developed by using measured road data as a three-dimensional track model. The software tool also includes an upgraded module called TestWare for Advanced Driver Assistance Systems that contains test cases, test automation and post-processing routines. A test and evaluation catalogue developed by TÜV SÜD Automotive can be used for testing ADAS within a series of test scenarios build on the basis of real driving tests. With the latest version 6, CarMaker supports realistic in-vehicle sensor models to simulate environment recognition and object detection. Since CarMaker was used for this thesis, the following paragraphs give an overview on relevant literature in the field of virtual testing for safety purposes.

Riegger et al. (2016) used CarMaker to developed and validate a centralised system, which takes control over autonomous vehicles within a certain surrounding of an intersection to optimise their trajectories for safe crossing. However, the study assumes that all vehicles approaching an intersection are automated, and is limited to straight crossing because turning manoeuvres were not included. Interaction with driver-operated vehicles, more complex manoeuvres as well as sensor noise were left for future studies.

Überbacher et al. (2017) applied CarMaker's vehicle-in-the-loop (VIL) function to evaluate a lane change warning and side collision warning. Information about the environment is modelled via virtual sensors and can be displayed to a driver in a real vehicle by means of a monitor or augmented reality glasses using the function IPGMovie. The advantage of VIL is that there is no need for additional vehicles or test drivers, as the environment is virtually created. However, it normally requires a closed test track with enough space to perform the defined manoeuvres.

Another VIL test using CarMaker was conducted by Pfeffer and Haselhoff (2016), where artificial pedestrian models were injected into the video stream of a camera-based driving assistance system. This was done to validate the performance of the camera-based ADAS in complex scenarios involving pedestrians. The implementation required a transfer of the vehicle's position in the real world into the simulated world.

Ni et al. (2016) developed a lane-departure prevention system, which was validated as MIL in CarMaker. In particular, the driver behaviour models of CarMaker were applied to take different driving steering inputs into comparison. Improved stability and velocity-keeping performance across a range of different drivers could be achieved.

A trajectory planning algorithm for lane change manoeuvres on motorways has been developed by Hansen et al. (2016). They used CarMaker to test different lane change scenarios and to analyse the algorithm's performance. During the simulation run, the planning adapts the trajectory to the changing surrounding situation.

In a study by Erbsmehl and Wagner (2012), CarMaker was used to reconstruct road accidents stored in the German in-depth database GIDAS. The study demonstrated the applicability of CarMaker in accident research by simulating a junction crash and by changing the configuration of the crash to make comparisons. However, the experiment was conducted for a simple test case (sight obstruction yes/no) and was limited to one reference accident.

Holzmann (2006) applied the CarMaker software to a simulation-based development of a chassis control systems. HIL test runs using electronic vehicle units were conducted to validate the system.

Being an alternative to CarMaker, PreScan is a physics-based simulation platform developed by Tass International and TNO used by the automotive industry for developing ADAS and ADS based on sensor technologies such as radar, LIDAR, camera and GNSS. PreScan is also used for designing and evaluating V2V and V2I communication applications. The software can be used from model-based controller design (MIL) to real-time tests with software-in-the-loop (SIL) and hardware-in-the-loop (HIL) systems. A dedicated pre-processor (GUI) allows users to build and modify traffic scenarios using a database of road sections, infrastructure components, actors, weather and light sources. Applications include autonomous and connected driving, emergency braking, lane assistance systems, pedestrian detection and avoidance or parking assistant systems. Bours et al. (2014) used PreScan to develop virtual driving scenarios from real-world tests, which are then varied to evaluate different situations. Another study was carried out by Schubert et al. (2014), who used PreScan to validate their probabilistic ADAS sensor models with real-world data. Autonomous emergency braking as well as a pre-safe seat belt system were evaluated by Seo et al. (2014), where PreScan provides a realistic radar sensor model, which is then coupled with control logics in Matlab/Simulink.

dSPACE is a producer of engineering tools for developing and testing mechatronic control systems, with applications in the automotive, aerospace and industrial automation. In the context of driver assistance systems, dSPACE offers a toolchain for function development, automatic production code generation and simulation that include Automotive Simulation Models. The models allow users to construct road networks, traffic, add objects and sensors and predefine manoeuvres. Representations of real roads can also be imported: map data (Open Street Map, Google Maps), ADAS RP or Open Driver format. Moreover, dSPACE provides an extensive test environment for executing Euro NCAP tests and validating the relevant systems by means of simulation. Standardized testing procedures for Automated Emergency Brake procedures for pedestrian detection have been incorporated into the latest versions. Virtual validation and hardware-in-the-loop simulation can be performed for adaptive cruise control, lane-keeping assistant, pedestrian detection, intersection traffic assistant etc.

The German company Vires developed the software VTD (Virtual Test Drive), which can be used to test ADAS and ADS. Similar to the abovementioned tools, it covers the whole range from MIL, SIL, HIL and VIL as well as driver-in-the-loop (DIL) (von Neumann-Cosel et al., 2009).

The simulation software Pro-SiVIC, originally developed by IFSTTAR and industrialised by a company called CIVITEC, was designed to reproduce various sensor functions of a vehicle. This allows simulating a vehicle's automated behaviour in varying road environments (Grapinet et al., 2013). The sensor models in the software package include vision sensors and individual test environments (such as road, roadside, weather and other road users) can be created.

2.5.3 Generation of test scenarios

The previous section has introduced different methods and tools to virtually validate ADS. Demonstrating the reliability, safety and robustness of ADS in all possible traffic situations under all road and environmental conditions is today's main challenge towards the approval and certification of ADS (ERTRAC, 2017). Kalra and Paddock (2016) investigated the miles needed to demonstrate that the failure rate of automated vehicles is below those of humans. Their statistical approach is based on per-mile injury rates from historical accident data and equivalent assumptions for autonomous vehicles. Figure 16 depicts the results by showing the miles needed in contrast to the assumed percentage of improvement in failure rates. For instance, a 20 percent improvement can be demonstrated by driving 11 billion miles, which would take a few hundred years. The number of miles decreases with a falling assumed reduction of failures.

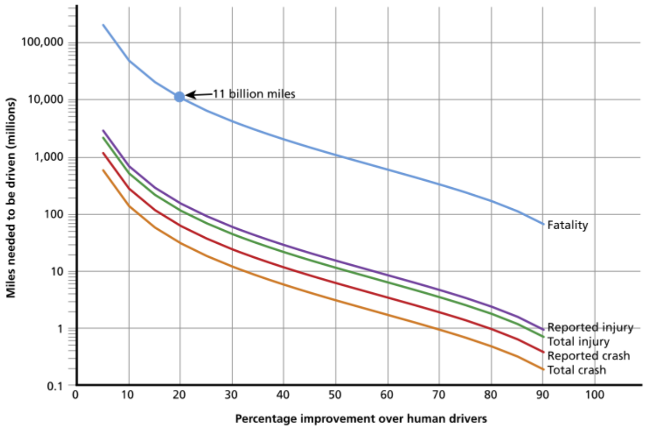


Figure 16: Miles needed to demonstrate with 95% confidence that the autonomous vehicle failure rate is lower than the human driver failure rate (Kalra and Paddock, 2016)

Wachenfeld and Winner (2015) studied the approval requirements for automated cars and based their calculations on German accident data. In average, there are 210 million kilometres between two fatal accidents with human-operated cars. They found that if the automated car were twice as safe as the conventional car, the distance would increase to 2.1 billion kilometres assuming a statistical success rate of 50 percent. Although their calculations are simple and theoretical, it could be shown that the safer a vehicle is expected to operate, the higher is the required test amount.

Those findings support the need for alternative testing methods to supplement real-world tests, in particular by virtual tests and simulations, which is recommended in a large number of publications (e.g. Beglerovic et al., 2017; Chen and Chen, 2010; Masuda, 2017; Olivares et al., 2016; Rolfsmeier, 2013). Simulations allow to run considerably more tests for a large number of situations. The most crucial aspect for defining virtual tests is the selection of critical scenarios, because they can significantly reduce the distances needed to be covered (Bock, Julian, 2017; Junietz et al., 2017).

In literature, many different definitions for scenarios can be found. According to the European AdaptIVe project, a driving scenario is an abstraction and general description of a driving situation without any specifications of the parameters of the driving situation. In contrast to that, a situation describes a specific scenario in more detail (Rodarius et al., 2015). A synthesis of various definitions of scenarios was proposed by Ulbrich et al. (2015). Accordingly, a scenario spans a certain amount of time, which consists of one scene or a sequence of scenes. A scene would be a snapshot of a scenario, where a certain driving action is performed.

In the PEGASUS project (Lemmer, 2017), the definition of a scenario has been categorised into

1. *functional scenarios* that contain natural language with a high level of abstraction (e.g. "ego car approaching with high speed on a three-lane motorway"),
2. *logical scenarios* that describe parameter ranges to reduce the level of abstraction (e.g. approaching speed: [60-90] mph, lane width: [2.7-3.9] m) and
3. *concrete scenarios* that have the lowest level of abstraction and specify each parameter precisely (e.g. approaching speed: 81.4 mph, lane width: 3.5 m).

Naturally, the number of scenarios increases from functional to concrete scenarios. The reason why these three categories were introduced is that different phases of the development cycle in the V-model require different levels of abstraction. For example, the concept phase necessitates a vocabulary that is readable and easy to understand for human experts, while in the validation phase the scenarios must be interpretable by computers for running the simulations.

This thesis has its own definition of a scenario, as explained in Section 6.2. Accordingly, a scenario defines dynamic traffic objects, how road users move, while a scenery includes the static elements such as the road environment. Translating the thesis' approach to the PEGASUS definition above, this thesis first derives functional scenarios from the accident data in study 2, and second, specifies concrete scenarios for the simulation models in study 3.

Several data sources can be used to generate and collect appropriate testing scenarios, which are:

- Real-world driving data from in-vehicle recordings, e.g. from previous field operational tests, naturalistic driving studies such as UDRIVE or SHRP2

- Real-world traffic observation data, e.g. from roadside traffic cameras
- Historical accident data, preferably in-depth statistics to get enough information about the scenario parameters
- Reused results from previous projects and pilot tests
- Results obtained from risk analysis methods such as FMEA (Failure Mode and Error Analysis) or FTA (Fault Tree Analysis)

The European research project ENABLE-S3 uses a combination of those sources to derive validation scenarios, as Figure 17 illustrates. It can be seen that real-world data is linked to scenarios obtained from other projects and scenarios derived from risk analysis methods. The parameters within those scenarios, e.g. weather conditions, vehicle types, are then varied to generate test runs for the validation platform. Following the V-model, the scenarios are tested on several levels from MIL/SIL to driving tests on the proving ground. The system was applied to evaluate testing methods for several use cases such as the highway pilot.

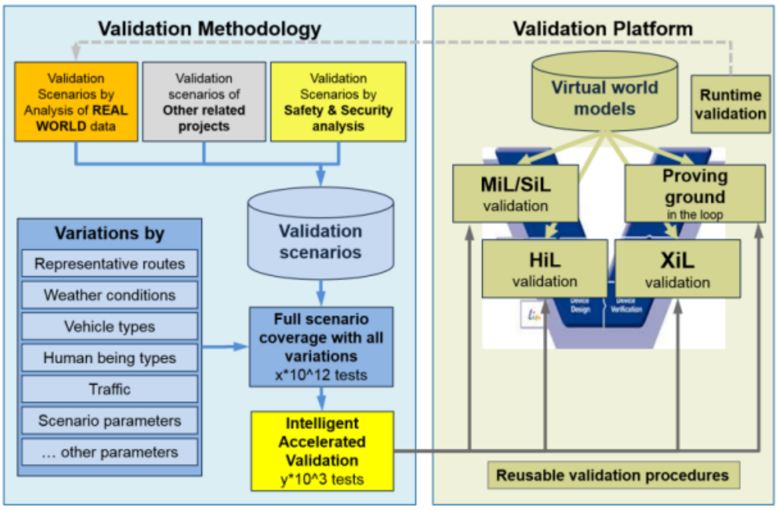


Figure 17: Validation architecture of ENABLE-S3 (Beglerovic et al., 2017)

In the project PEGASUS, testing procedures were developed for the highway pilot, including a novel scenario generation method (Bock, Julian, 2017; Pütz et al., 2017; Zlocki et al., 2017). A database was implemented to process data from different sources in a uniform toolchain, which comprises the analysis of FOT data, accident data, driving simulators, traffic simulations and experts knowledge. The concept in Figure 18 shows that the database is used to extract information for different levels of the V-model to define the tests, but is also used to store information from those tests. They call it a circuit of relevant scenarios, which is constantly updated with new information. A main element of the database is the storage of parameter distributions obtained from driving data, which are important to parametrise simulation models.

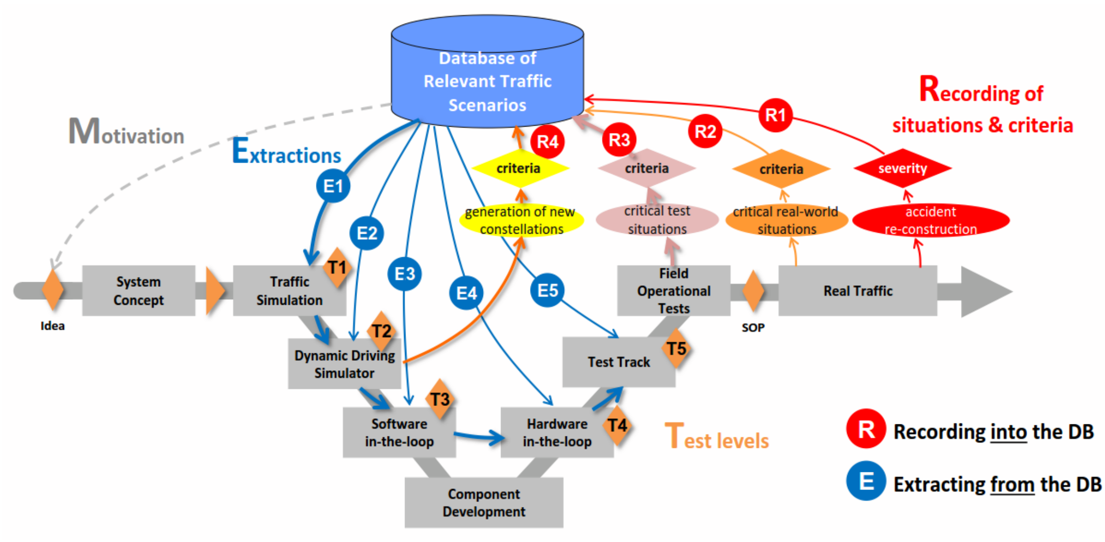


Figure 18: Database-driven scenario generation in PEGASUS (Zlocki et al., 2017)

Kim et al. (2017) evaluated collision warning systems (camera-based and radar-based) for intersections based on scenarios derived from naturalistic driving data and accident records. They identified sixteen vehicle-to-vehicle accident scenarios and studied the accident prevention capabilities in those scenarios. However, their study considers only one safety indicator called safety-remaining distance and no conflicts, and the interaction with vulnerable road users is not included.

An interesting approach to reduce the approval effort and parameter variations is the principle of functional decomposition (Amersbach and Winner, 2017). The approach is commonly used in robotics and informations, but also in the analysis of the human driving task. Graab et al. (2008) suggest five levels of failure classification, distinguishing into information access, information reception, information processing, behavioural decision and the final driving action. Each of those levels can have different causes of failures. For instance, was the information perception influenced by distraction, was the information correctly interpreted or did the drive take the right decision for an action? Amersbach and Winner (2017) translated this approach into a systematic decomposition of automated vehicle operations. The different levels were used to set specific critical scenarios based on an FTA. Their approach certainly has the potential to reduce the approval effort, but this reduction has to be quantified in future studies.

A conclusion that can be made after reviewing current projects regarding scenario generation and testing of ADS is that they focus on near-term applications of automated vehicles, such as the highway pilot. There is little research on testing scenarios for other applications of ADS, in particular for urban or suburban sceneries. This underlines the need for a validation method for junction sceneries, which pose a particular risk for ADS.

2.5.4 Testing standardisation

Currently, standardised and harmonised testing procedures or safety guidelines for the on-road testing of vehicles equipped with prototype automated vehicles do not exist. However, in March 2015, the SAE On-Road Automated Vehicle Standards Committee has released J3018, which provides general guidelines for performing tests of prototype ADS equipped on test vehicles operated in mixed traffic environments on public roads (On-road Automated Vehicle Standards Committee, 2014). Additionally, SAE is currently preparing the standard J3092, which will contain dynamic test procedures for verification and validation of ADS, such as those performed on a test track.

In 2013, Euro NCAP released the test protocol for Automated Emergency Braking (AEB) systems for low-speed and urban type car crash scenarios (Fildes et al., 2015), in which they describe the standardised testing procedures (e.g. measuring equipment, test conditions etc.) for this type of system. A similar protocol for more complex scenarios, such as pedestrian detection followed in 2016, but there are still much more scenarios that must be considered.

Testing of ADS has been and is being performed in various European and national projects, by vehicle manufacturers and automotive suppliers. However, the literature review revealed that there are no common and harmonised procedures for testing ADS. For the automotive industry, the international standard ISO 26262 (ISO, 2011) is of main importance, when it comes to the functional safety of vehicle systems. The ISO 26262 is named "Road Vehicles – Functional Safety" and provides regulations and recommendations throughout the product development process for automotive-specific electrical and electronic (E/E) systems. Although ISO 26262 compliance is not mandatory, it is considered state-of-the-art concerning functional safety for road vehicles up to 3.5 tonnes (Weigl, 2014). The standard introduced a risk-based approach for determining risk classes, namely the Automotive Safety Integrity Levels (ASILs) from A to D, with D requiring the highest integrity requirements on the product. The ASILs ask the question "If a failure arises, what will happen to the driver and associated road users?" (National Instruments, n.d.). Hence, level D rated systems have a likely potential for severe or fatal injury when they fail and therefore require the highest QM standards. The risk level is determined by three variables: severity, probability of exposure and controllability. And this poses a problem when defining ASILs for fully automated vehicles, because as the standard reads now, the absence of a human driver means zero controllability (Lee and Hobbs, 2013).

2.5.5 Findings

In summary, the review of testing and validation procedures for ADS can be concluded as follows:

- Regulations such as the Vienna Convention on Road Traffic and UN regulation 79 have been amended to allow testing of automated vehicles on public roads. However, there are still legal hurdles.

- Worldwide, ADS testing is being initiated. In Europe, several test sites have been opened to validate automated driving functions. Twenty states of the USA have allowed public testing and Japan aims to introduce automated road traffic by 2020.
- Before testing on the road, the automotive industry validates their systems in virtual environments to cover numerous scenarios and different testing conditions. Commercially available simulation software packages are CarMaker, dSPACE, PreScan, VTD or Pro-SiVIC.
- Virtual or real-world tests require a clear definition of relevant test scenarios. A combination of different data sources is recommended to derive such scenarios, including accident data, real-world driving data or risk analysis methods. However, there is a need to derive scenarios for ADS applications on non-motorway areas, since most current projects focus on motorway scenarios.
- SAE is currently preparing the standard J3092, which will contain dynamic test procedures for ADS on a test track. However, there are currently no standardised procedures for testing ADS, especially not for intersections.

2.6 Research gaps addressed by the thesis

Among the many challenges that ADS face, the functional safety and reliability of environmental perception and motion planning were found to be the most crucial one for market introduction. Therefore, the automotive industry keeps on testing their vehicles under different circumstances. However, the literature review showed that there are currently no common validation procedures and benchmark scenarios. As a summary of this chapter, three main research gaps are defined in the following, which are addressed by the thesis.

Research Gap #1: Ensuring high road safety at road intersections with mixed vehicle population of automated and non-automated vehicles:

One of the high-level research gaps in the world of automated driving is how to achieve full reliability of highly automated vehicles on public roads, taking into consideration other traffic participants and non-automated vehicles. Especially in urban areas, automated vehicles still struggle with complex environments and unexpected behaviour of other drivers or vulnerable road users. The purpose of this thesis is to provide novel methods for improving testing procedures, which lead to avoidance or mitigation of intersection crashes involving automated vehicles.

Research Gap #2: Identifying key scenarios and testing procedures for ADS at intersections:

Another main research gap addressed by this work is that there are no standardised procedures for evaluating automated driving systems in junction environments. The literature review has revealed that it is too comprehensive to cover all relevant driving situations when testing ADS, either in real-world or in virtual test environments. Therefore, there is a need for identifying key driving situations used as "benchmark scenarios", which constitute the core population of driving situations to be passed by automated vehicles. This thesis contributes to the problem of finding such scenarios for the areas of road intersections, e.g. including relevant crash types, collision scenarios or junction characteristics. This research gap is linked to the research questions 3 and 4 given in Section 1.2.

Research Gap #3: Providing a method to evaluate the safety performance of ICAMS under a representative variation of real-world conditions:

An intersection assistance system for automated cars must be able to adapt to different intersection types. The literature review has shown that most studies do not validate their models on a large variety of intersection layouts and characteristics. Furthermore, various collision detection algorithms have been proposed, but the coverage of these algorithms is limited to only a few scenarios. Validation and verification of the systems must take into account the most critical combinations of collision parameters. However, those parameters may not be available from real-world observations or accident data. This gap is particularly addressed by the research question 4 and 5.

3 Research methodology

This purpose of this chapter is to explain the methods used to achieve the research objectives. First, the overall research design is presented, followed by a detailed description of each research study's methodology.

3.1 Overall research design

The thesis is structured into three studies, as depicted in Figure 19. The initial study 1 was conducted to set the research scope according to relevant research gaps as well as challenges and problems for automated traffic at intersections. The research gaps and challenges were identified by reviewing literature in various areas and by surveying experts with an online questionnaire. After that, a preliminary analysis of junction accident data helped to understand the main safety problems and accident circumstances at junctions. Thus, study 1 addressed the research question 1, 2 and 3 (see Chapter 1.2).

Study 2 built upon the findings of study 1 to investigate junction crashes in detail. By clustering historical in-depth accident data from the UK, the key crash scenarios at junctions were identified and described. The underlying data comprised accidents involving human drivers only and was used to draw a picture about critical intersection situations that drivers usually encounter. The results of study 2 deliver answers to the research question 3.

In the third and final study, a novel, modular simulation and evaluation framework is presented to reconstruct the crash scenarios in a virtual testing environment, where automated driving functions replace the driver. The framework is demonstrated on a particular junction scenery, where automated collision avoidance systems are evaluated. The simulation output comprises various safety indicators to quantify the relative safety performance. Thus, the third study addresses research question 4 and provides the basis for recommendations on safety performance indicators as part of research question 5.

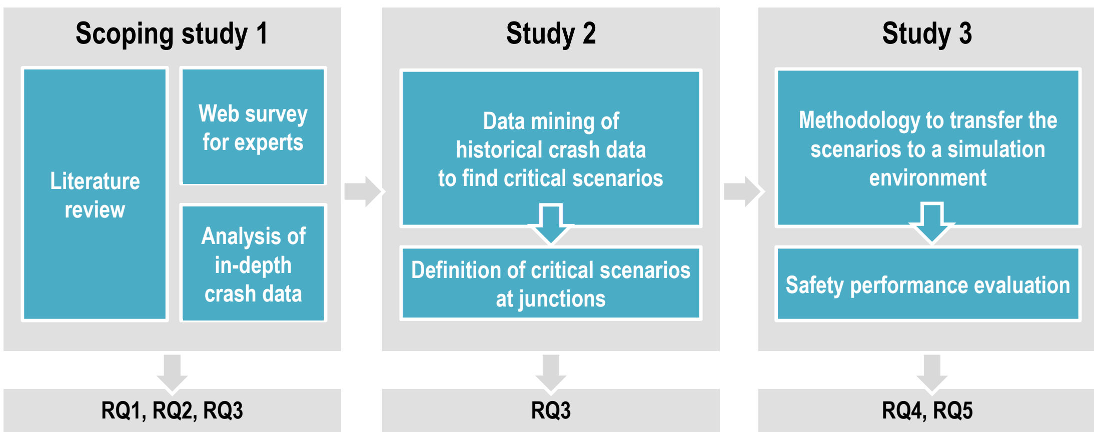


Figure 19: Overall research design

3.2 Study 1

As the first step of the scoping study, literature was reviewed to explore the context and identify research gaps. The review comprised the following topics (see Chapter 2): Technical hurdles for automated driving systems including their limitations of vehicle technologies, accident data types and investigation methods, safety at road intersections including accident statistics, automated intersection collision avoidance and mitigation systems as well as current testing and validation procedures for ADS.

A problem that occurred during the literature review is the novelty of the topic, which limits the number of available research papers and reports. Furthermore, given the fact that this topic is mainly industry-driven, recent developments are rarely made public. Additional information was necessary to get a complete picture of current developments and challenges in road transport automation. Therefore, a web questionnaire was set up to address experts in the field of ADS, working in the industry, academia, research institutes and public authorities. For the survey study, the following steps, loosely based on the guide by Burgess (2001), were undertaken:

1. *Decide on the information required according to the research aims:* The literature review showed that more information on the technical readiness of automated driving functions is needed. Furthermore, factors that influence the performance of automated vehicles on public roads needed to be surveyed. This information helped to select appropriate accident variables and indicated, which automated driving subsystems needed to be investigated.

2. *Define the target respondents*: The idea was to reach as many representatives from industry as possible. However, due to limited response, the target population was widened to research institutes and academia, as well as public authorities and consultants. In total, 54 complete responses were received.

3. *Design the questionnaire*: Given the purpose of the survey, an exploratory questionnaire design was chosen to gather preliminary information to define the problem. For each topic of the questionnaire, namely technologies for ADS and influencing factors, closed questions were stated with predefined categories for the answers, followed by a text box for comments. The questionnaire was designed using Google Forms and did not take longer than 10 minutes on average. All the questions from the survey can be found in Appendix A.

4. *Run a pilot survey:* A pilot survey was conducted among colleagues to check for mistakes or usability issues.

5. *Carry out the main survey:* After making some amendments, the main survey was carried out by distributing the questionnaire link among colleagues and professional contacts, networking organisations and the professional social network LinkedIn.

6. *Analyse the data:* After the deadline given for survey responses, data were analysed using the exported Google summaries in MS Excel. See Section 4.1 for further information and results of the survey.

The expert survey was followed up by an initial data analysis of junction accidents. It was decided to focus on accident data, as this can draw a picture of the current safety aspects. First, it was investigated and selected which accident database to be used for the thesis. To obtain in-depth information, the following microscopic UK databases were investigated: CCIS, OTS and RAIDS (see Section 4.2.1). RAIDS was found to be too sparse for the desired data query with only 75 cases to investigate. So, CCIS and OTS were among the candidates, but difficult to merge data-wise. Instead of conducting two separate analyses, and because CCIS only contains car accidents, OTS was chosen as dataset, having a proper sample size and a sufficient number of attributes included.

From the OTS database, data elements to be queried were defined and exported from an online interface. After data post-processing, i.e. reducing variables and removing missing values, a descriptive statistical analysis of the main accident parameters was conducted. A significance test was done to confirm the generalizability of the OTS data sample in comparison to the overall accident population taken from the CARE database.

3.3 Study 2

The second study is related to the research question 3 (see Chapter 1.2). The goal of the second study was to identify distinct groups of junction accidents by applying an appropriate clustering technique. Due to the different principles of the algorithms, one method might produce different clusters to another method. Hence, one has to choose the most appropriate method for the underlying dataset, taking into account the sample size, the number of attributes, the attribute types as well as the desired output of the study.

An important prerequisite of the clustering exercise is that it needs to deliver results for defining the simulation parameters in study 3. Modelling a virtual road environment and collision scenarios necessitates detailed information on the circumstances and location of the accidents, which assists the definition of static and variable simulation parameters. For example, a static attribute can be the junction type and the type of vehicles involved, while varying parameters can be the driving trajectories, traffic conditions or driving velocity. Therefore, the overall methodology of study 2 is strongly linked to the requirements of the subsequent simulation study. One major requirement is that the number of varying parameters is kept low in order to reduce the number of possible combinations. Also, the range of variations should be narrowed down as much as possible. Consequently, study 2 aims at providing enough information on possible junction scenarios and parameter combinations.

The methodology flowchart is depicted in Figure 20. Inspired by a study from Kumar and Toshniwal (2015), the idea was to initially partition the data by a clustering technique for categorical data and then apply the association rule method on the data subsets. Note that the elements of study 3 are greyed out.

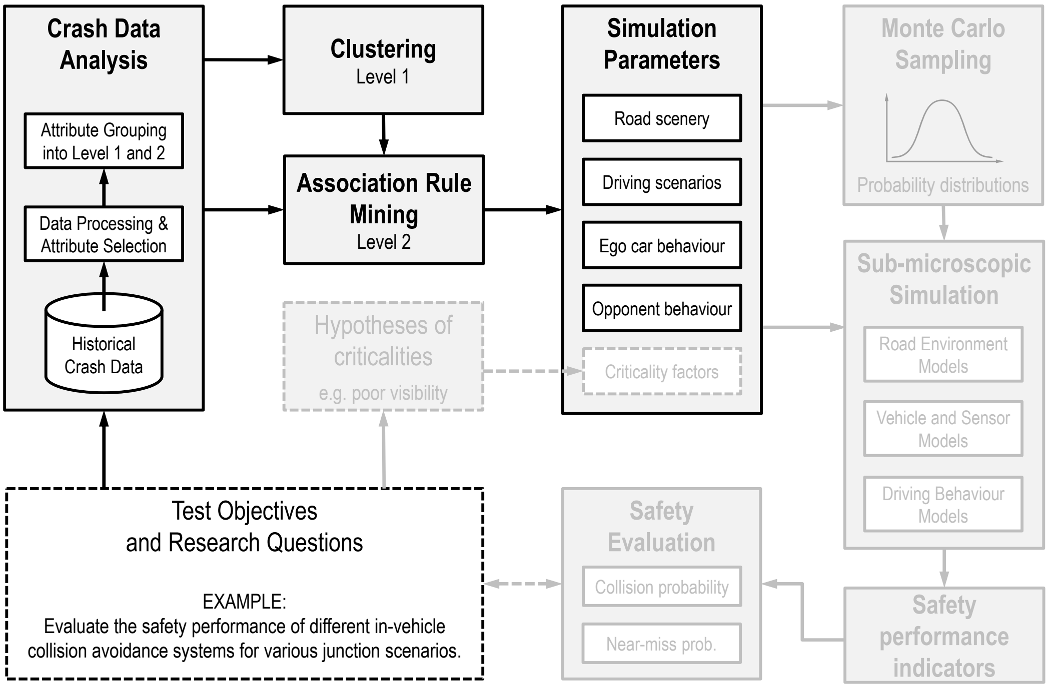


Figure 20: Methodology of study 2

The historical crash data used and its processing steps are explained in Section 4.2, including the procedure of attribute selection. Section 5.2 describes the attribute coding and grouping into two levels. Level 1 is a reduced set of attributes describing the collision parameters, for better partitioning and easier interpretation of the results, while Level 2 adds additional attributes describing the environment and causation factors. Level-1 data is used as input for the clustering algorithm and level-2 data for finding association rules. The main reasons why this two-level approach has been chosen are the following:

1. Most clustering methods achieve superior clustering results on a smaller set of attributes. No clear partitioning would be achieved when using all available attributes.
2. The results from applying the association rules on the whole dataset (without prior clustering) would be hard to interpret due to the high number of obtained rules. It must be noted that depending on the sample size and attribute dimensionality, millions of rules might be computed. This requires post-processing by applying dedicated algorithms or pruning techniques.

The accident data used in this study is of categorical nature, i.e. described by qualitative attributes (also called nominal attributes) of mainly arbitrary order.

Although the categories can be coded as numbers, e.g. 1: female, 2: male, those numbers would not have mathematical meaning (Han et al., 2011; Lourenco et al., 2004). Therefore, dedicated statistical methods are necessary to analyse categorical data. Data mining techniques are based on a dissimilarity or distance measure between samples. This is especially true for clustering, which can be described as a collection of data objects such that the objects within a cluster are similar to each other and dissimilar to the objects in other clusters. Conventional clustering techniques often use the Euclidean distance measure, which would not make sense for the given OTS dataset.

In literature, a range of different definitions are given for clustering, but one of the most apposite ones for this thesis is: "Finding of natural groups from a data set, when little or nothing is known about the category structure" (Anderberg, 1973). In fact, natural groups of accident circumstances at UK junctions are unknown, and the goal of this thesis is to discover such groups.

Clustering methods can be classified into the following groups (Berkhin, 2006; Han et al., 2011):

1. *Partitioning methods* decompose a data set into a set of disjoint clusters. Popular methods are e.g. k-means and *k*-medoids that usually start with a random partitioning and refine it iteratively to find a (not necessarily global) optimum. See Section 5.3.1 for a more detailed explanation.

2. *Hierarchical methods* build a tree of clusters, also known as dendrogram, where each cluster node contains child clusters. The number of clusters depends on where the tree is cut. This approach has the advantage that data can be explored at different levels (hierarchies).

3. *Density-based methods* can cope with clusters of arbitrary shape and data outliers, since they are based on a density metric instead of a distance metric. The most popular algorithm, DBSCAN (Ester et al., 1996), clusters points that are closely packed together and takes note of single outlier points in low-density areas.

4. *Grid-based methods* quantise the object space into a finite number of cells that form a grid structure (Schikuta, 1993). The idea is to organise the value space surrounding the patterns and not the patterns, as the other three cluster method groups do. For large datasets, this approach tremendously increases execution times.

Elavarasi and Akilandeswari (2014) and Rezanková (2009) published surveys on clustering algorithms for categorical data and presented possible methods, among which are SQEEZER (He et al., 2002), ROCK (Guha et al., 1999), LIMBO (Andritsos et al., 2004), STIRR (Gibson et al., 1998), Link Clustering (LC, Zengyou et al., 2004) or CACTUS (Ganti et al., 1999). Also, conventional clustering algorithm were modified to deal with categorical data, such as *k*-modes (Huang, 1997; Huang and Ng, 1999), *k*-histograms (Zengyou et al., 2003), *k*-medoids (Kaufman and

Rousseeuw, 1990) or Generalized Self-Organizing Maps (SOM, Hsu, 2006), all of which having their advantages for different applications. Basically not a clustering method, but a popular classification algorithm for categorical data is Latent Class Analysis (LCA, Goodman, 1974), which is a model-based approach, assuming that a mixture of underlying propability distributions generates the data. In contrast to conventional clustering techniques, LCA assigns probability of each sample to be in different classes, instead of partitioning the data into fixed clusters. In the following, it is explained how clustering has been set up in this study and which technique and parameters were chosen.

According to Anderberg (1973), the procedure of clustering should contain at least nine elements. Äyrämö and Kärkkäinen (2006) extended this list by adding a missing data strategy:

1. Choice of objects
2. Choice of variables
3. What to cluster: data units or variables
4. Normalization of variables
5. Choice of (dis)similarity measures
6. Choice of clustering criterion (objective function)
7. Choice of missing data strategy
8. Algorithms and computer implementation (and their reliability, e.g., convergence)
9. Choice of appropriate number of clusters
10. Interpretation of results

In this study, this process was undergone as follows:

(1) The choice of objects, i.e. data samples, was done by filtering and exporting OTS data from the online database (see Section 4.2.2).

(2) The variables were selected according to the scope of this study. For example, detailed occupant injury data, or vehicle damage information is not relevant for this study. Hence the corresponding variables were removed. As Section 4.2.4 will describe in detail, many of the remaining variables were removed as well, among which are low-semantic variables (e.g. Case ID, weekday), variables with mainly high-occurrence values, highly correlated variables or those with many missing values. For clustering, the number of variables was again narrowed to achieve interpretable results, as explained in Section 5.2.

(3) According to the research questions, it is necessary to cluster the data units and not the variables, although the latter might reveal similar variables that could be grouped in order to reduce the dimensionality.

(4) "Normalisation is the process of scaling individual samples to have unit norm" ("Normalization - PHP-ML - Machine Learning library for PHP," n.d.). It is a common pre-processing step for many machine learning methods. However, since this

study uses categorical attributes and not numeric ones, normalisation is not needed. It must be noted that all attributes were coded in a binary format (see Section 4.2.4).

(5) The choice of dissimilarity measures was narrowed down by the fact that the underlying OTS data is of categorical nature. Categorical variables require a non-numerical distance metric. Commonly used measures for binary-coded categorical data, as used in this study, are the Hamming distance or Jaccard distance. While the Hamming metric simply gives the number of elements that are different between two binary vectors, the Jaccard distance gives the percentage of non-zero elements that differ. See Section 5.3.1.

(6) The clustering criterion is predefined by the algorithm used. For partition-based clustering, the criterion is the minimisation of the within-cluster distances of the points in a cluster and likewise, the maximisation of the intra-cluster distances, i.e. the distances between the cluster centres.

(7) For the purpose of clustering, samples with missing values, i.e. samples with unknown attributes, were removed (see Section 4.2.2).

(8) After investigating different clustering techniques, the k-medoids method was chosen, in particular an algorithm called PAM (Partitioning Around Medians, Kaufman and Rousseeuw, 1990) was applied to compute the clusters. k-medoids was found to be most appropriate for the given dataset, since the method is robust against outliers and can cope with categorical data. See Section 5.3.1 for further information.

(9) As for k-means, k-medoids requires a predefined number of clusters (k's) to partition the data accordingly. In an iterative process, the number of k was incremented to a maximum k and a validity measure called silhouette value was used to identify the best k. See Section 5.3.2.

(10) One of the trickiest part of cluster analysis is the interpretation of the results. The aim is to describe each cluster by words so that a critical scenario can be derived. A χ^2-goodness-of-fit test was carried out on the sample population within each cluster compared to the population of all clusters to identify significant differences and over-represented attributes. See Section 5.3.3.

Association rules are a popular method in data mining to discover hidden associations within variables in large datasets. A popular application of association rules is the so-called market basket analysis, where associations between supermarket products from a large record of transactions are computed (Agrawal et al., 1993). For example, a resulting simple rule could be "beer → crisps", which would indicate that customers who buy beer also buy crisps. The information derived from such rules can be used as a basis for decisions about marketing activities and product placement. In the case of this thesis, association rules are used to get insight into the collision circumstances within each cluster. For example, a rule "SpdLimit=20mph → MaxInj=Slight" would indicate that low-speed accidents are associated with slight injury. It is important to note that association rules do not indicate causality, which the arrow might suggest.

Section 5.4 explains the principles of association rules in more detail and presents and discusses the results obtained.

Ultimately, the association rules result in various crash scenarios that can be described by four groups of simulation parameters: Road configuration, driving scenarios, ego car behaviour and opponent behaviour (see definitions in Chapters 6.8.2 and 6.8.3). Those parameters are further modelled in the virtual environment.

3.4 Study 3

The third and final study is related to the research questions 4 and 5 (see Section 1.2). The goal was to develop a methodology to transfer the derived collision scenarios (see Section 5.4.3) to a sub-microscopic simulation environment for examining the safety performance of ADS at junctions. The methodology for study 3 is given in Figure 21 and shows that the simulation module is strongly related to the clustering analysis. Note that the elements of study 2 are greyed out.

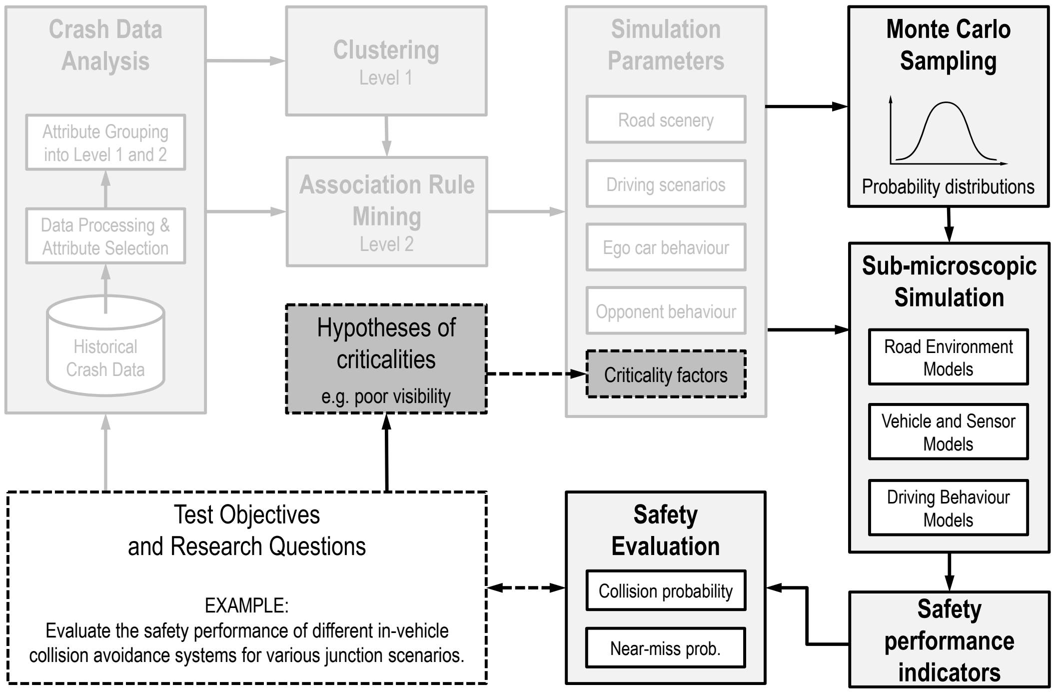


Figure 21: Methodology of study 3

In general, the proposed simulation framework is based on one or more experiments to test pre-defined hypotheses. Those hypotheses are related to the objectives of the virtual test, as described in Section 6.1. While study 2 analysed junction accidents involving driver-operated vehicles only, the simulation framework replaces the driver with an automated vehicle, further denoted as ego car. On the one hand, this affects the parameters for the ego car model, such as reaction times or driving behaviour. On the other hand, factors that are assumed to be problematic for automated vehicles must be taken into account. Risks that could negatively influence the safety performance of ADS are specified as “criticalities” (see Section 6.2.5). Hence, criticality factors are added to the parameters derived from the association rule analysis. For the demonstration of the framework, the simulations were realised by using the computer simulation tool CarMaker, which allows to virtually evaluate a large number of variants in comparison to real-world tests. The reasons for using CarMaker are given in Section 6.3. However, any other simulation tools such as dSPACE or PreScan might be used instead, as long as they can provide the same output parameters used for the safety performance evaluation.

As Section 6.2 defines, a simulation experiment contains various sceneries, i.e. road and junction environments to be analysed. Each scenery itself contains various scenarios, selected from those presented in Section 5.4.3. It is significant to mention that not all simulation parameters can be derived from the clustering study. For example, approaching speeds, lane width or vehicle parameters have to be assumed, because they were not given in the underlying accident dataset. The original crash scenario is modified and simulated until the point of impact, by implementing automated driving functions and varying collision and road infrastructure parameters. The method can therefore also be called a pre-crash simulation, because it does not replicate post-collision damage, injury and vehicle positions.

For each scenario, some of the simulation parameters remain static and some are varied within a certain range, such as the lateral position in the lane and the speed of the opponent, the friction coefficient of the road surface or the level of collision avoidance systems to study their impact. The Monte Carlo sampling method, in particular the Latin Hypercube Sampling (LHS, McKay et al., 1979), was applied to randomly select samples among the value range of the varying parameters. There is no single Monte Carlo method definition, but most approaches follow this pattern: 1) model a system as a (series of) probability density functions (PDFs), 2) repeatedly sample from the PDFs and 3) compute the statistics of interest (Harrison et al., 2010). The three steps were also taken in the underlying study, by defining PDFs for each varying parameter, sampling from the PDFs by taking into account the previously generated sample points and computing a collision and near-miss probability from the simulations. A detailed explanation of the sampling method is given in Section 6.7.

The simulation block in Figure 21 is divided into the model groups

- road environment (see Section 6.6.1),
- vehicle and its sensors (see Sections 6.6.3 and 6.6.4) as well as
- driving behaviour models (see Section 6.6.5).

CarMaker allows to accurately designing road environments including road layout, surface parameters and roadside elements. A large number of car models can be chosen and parametrised for the ego car and surrounding traffic such as the opponent can be modelled as all road user types (e.g. cars, trucks, pedestrians, cyclists). The sensor models are used to replicate an automated car’s ability to sense its environments and also reflect their limitations due to poor visibility or sensor failure.

The vehicle control module that perceives the sensor data as well as the opponent behaviour is specified by the driving behaviour model. This includes the modelling of longitudinal and lateral control, or in the case of ADS, automated lane assistance (lane keeping, lane change), automated speed and/or distance control system (when approaching and crossing an intersection) and automated emergency braking to avoid a collision.

An important task to assess the safety performance of ADS is to use appropriate safety indicators. In this study, not only crash indicators are investigated, such as collision yes/no, impact speed or angle, but also near-miss indicators are taken into account, e.g. post encroachment time or time-to-collision. By varying parameters, many different scenario configurations are simulated and the actual collision is then unlikely. Hence, by evaluating near-misses one can obtain additional information on safety risks. See Section 6.4 for further information. The resulted safety performance indicator values are then used to compute a collision and near-miss probability for each scenario variation to derive findings. The findings are discussed against the hypotheses stated for each experiment.

The developed simulation framework is demonstrated for a specific junction scenery to show its level of usefulness and versatility. Instead of modelling an artificial, virtual road scenery, a real-world junction in the East Midlands, UK, was chosen to showcase the method for a practical example. This had the advantage that the environment parameters such as lane and shoulder widths, position of roadside elements, gradients, i.e. parameters that may not be directly derived from the crash clusters, were taken from an on-site inspection. Furthermore, taking a real-world scenery for the simulation experiment demonstrates the use of the framework for road operators and authorities to evaluate particular junctions in their network.

This approach is different from traditional crash simulation studies, where real-world road locations are reconstructed case-by-case from a crash sample of the underlying accident database and the effectiveness or crash avoidance rates of safety interventions is studied (Brunner et al., 2003; Canu et al., 2016; Cliff and Moser, 2001; Helmer, 2014; Sander, 2017; Sander and Lubbe, 2018). Commonly, those studies use a crash reconstruction tool such as the software PC-Crash. In this thesis, the virtual junction is not included in the OTS samples, because the crash samples are not evaluated case by case. Instead, generalised abstract scenarios are obtained from clustering and further specified by association rules. More concrete, less abstract scenarios are then produced by parametric variation using the LHS method (see Table 4). In other words, artificial critical scenarios are evaluated instead of accident scenarios that really happened. In this way, more parametric variation can be implied than by using a limited crash population.

In the demonstration experiment, the safety performance of two different automated collision avoidance systems is studied and the capabilities as well as the limitations of the framework are discussed later on.

Table 4: Types of scenarios and how they are derived

Type:	**Functional scenario**	**Logical scenario**	**Concrete scenario**
Level of abstraction:	High	→	Low
Source:	Clustering results	Association rule results	Sampling of parameter distributions (LHS)
Described by:	Natural language	Parameter ranges and attributes	Precise parameter values
Examples:	A car turns left on an urban T junction and hits another PTW.	Surface: wet (friction coefficient: 0.3-0.9), speed limit (30-40mph), traffic control: light, road type: dual carriageway	Position in lane: 0.3m, velocity: 27mph, Friction coefficient: 0.43

4 Study 1: Scoping and initial analysis

The following sections describe the work carried out during the first study of this thesis, to set the research scope according to relevant research gaps as well as challenges and problems for automated traffic at intersections. Besides a comprehensive literature review, research activities included a web expert survey as well as an initial analysis of in-depth accident data. The work of this initial study addresses research question 1 and was published by Nitsche et al. (2014).

4.1 Web expert survey

The literature review described in Chapter 2 was complemented by a web survey to get insights, which have not been published by the industry or are not yet available due to the novelty of the topic. At the time the survey was conducted, the scope of the thesis was not completely defined. It therefore helped to get ideas on the challenges that ADS currently face and the readiness of ADS technologies. The survey was designed as exploratory questionnaire targeting experts in the field of automotive development and testing. For each topic of the questionnaire, namely technologies for ADS and influencing factors, closed questions were stated with predefined categories for the answers, followed by a textbox for individual comments. The questionnaire was designed using Google Forms and did not take longer than 10 minutes on average.

To ensure acquiring expert opinion, the questionnaire invitations were sent to contacts in well-known institutions active in the fields of automotive research and development, with the request to forward it to relevant persons. For example, the link was distributed among the CLEPA (European Association of Automotive Suppliers) partner network. In addition, project teams of the DARPA challenges were approached, as well as other European networks such as ERTICO (European Road Transport Telematics Implementation Coordination) and FEHRL (Forum of European Highway Research Laboratories). Furthermore, the questionnaire link was distributed through dedicated LinkedIn groups with members in the field of automated driving and ADAS. Due to this approach, a response rate is difficult to calculate, because the overall number of recipients, who have actually read the survey invitation, is not clear.

In total, 54 persons completed the questionnaire, which was considered a satisfactory sample size. The questions included the role of road infrastructure, market readiness as well as to which extent certain factors influence the performance of selected ADS groups. The survey was kept anonymously, with the option for the respondents to state their institution and email address. As shown in Figure 22, the majority of responses (36 of 54) came from research and development experts, which could either be from research institutes or industry companies. Multiple answers were possible, if the expert worked in several domains. 13 out of 54 respondents worked in academia, followed by 8 experts from the industry domain, 3 from public authorities and 5 from other domains such as consultants or civil engineers. It must be noted that the survey was conducted in 2014 and their opinions and developments might have changed since then. However, for the further work in the thesis, the survey provided a relevant complement to the literature review.

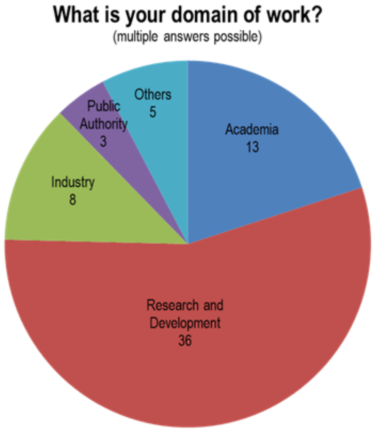


Figure 22: Number of responses per domain of work (n=54, multiple answers possible)

As a first question, the role of the road infrastructure was rated as "very important" by 76 percent of the respondents (see Figure 23). Only one participant believed that the infrastructure is not important at all. This question served as a simple introductory question and was asked to get a general idea if ADS require additional road infrastructure and if the infrastructure itself might have an influence on the performance of ADS.

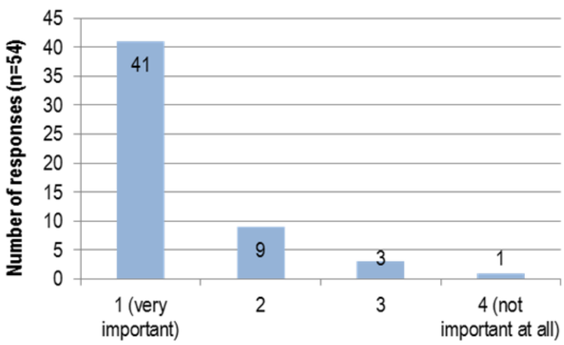


Figure 23: Number of online survey responses to the question "How important would you rate the role of road infrastructure in a world of self-driving cars?" (n=54)

In terms of technological readiness for market (see Figure 24), selected automated driving systems were rated as "still under research" (green bars), "prototyping" (red bars) and "ready for market" (blue bars). The selected automated driving subsystems are explained in Table 5.

Table 5: Explanation of automated driving subsystems listed in the survey

System	Description
Automated lane keeping	Systems which take over steering, keep the car centred in the lane, and ask the driver to take over when needed.
Dynamic speed control and adaptation	Systems which take over longitudinal control by selecting, keeping and adapting the driving speed according to speed limits and environment conditions.
Fully automated parking	Systems which take over steering and accelerating/braking to bring the vehicle into a parking position.
Fully automated braking	Systems which activate the brakes in an emergency situation, when a collision is imminent.
Obstacle crash avoidance	Systems which take over steering and braking to avoid a collision with an obstacle on the road or roadside.
Collision avoidance with other road users	Systems which take over steering and braking to avoid a collision with other road users such as vehicles or pedestrians or cyclists.
Fully automated steering	System which take over steering in all driving situations even if there is no lane marking.
Automated safety pull-over in emergeny cases	Systems which perform a safe exit manoeuvre with minimal risk, e.g. drive to the shoulder and stop when the human passenger does not respond to take over.
Automated lane changing	Systems which change from one lane to another by observing the surrounding traffic.
Automated lane merging	Systems which change to another lane, because the current lane is ending, e.g. on motorway on-ramps. The surrounding traffic must be observed and a safe gap must be found.
Automated overtaking	Systems which change from one lane to another, overtake and change back the lane, by observing surrounding and oncoming traffic.

It is important to note that the rating is based on technological readiness in the year 2014, disregarding legal, liability and privacy issues. The coloured areas in the background were added to visualise the decline of responses for systems that are market-ready, while the responses for “still under research” are increasing for those systems. This is logical and shows that e.g. automated lane keeping only needs little research, because it is already ready for the market, while more complex systems such as automated overtaking are still under research and not ready for the market yet. In general, responses indicate that lane keeping, speed adaptation and parking systems are in an advanced stage of deployment, while assistance systems that must handle more complex manoeuvres, e.g. overtaking, are still under research. However, there is still some discrepancy, such as with automated braking, where more than twenty responses were given for either “prototyping” or “still under research”. A reason for this could be that the definition of this ADS category was not fully clear to the respondents.

The respondents were asked to mention additional systems, which they consider relevant for automated transport, among which are traffic sign recognition, detection of markings other than lane markings, trajectory planning, negotiation and compromise with other road users, platooning, intersection communication assistant or traffic signal communication.

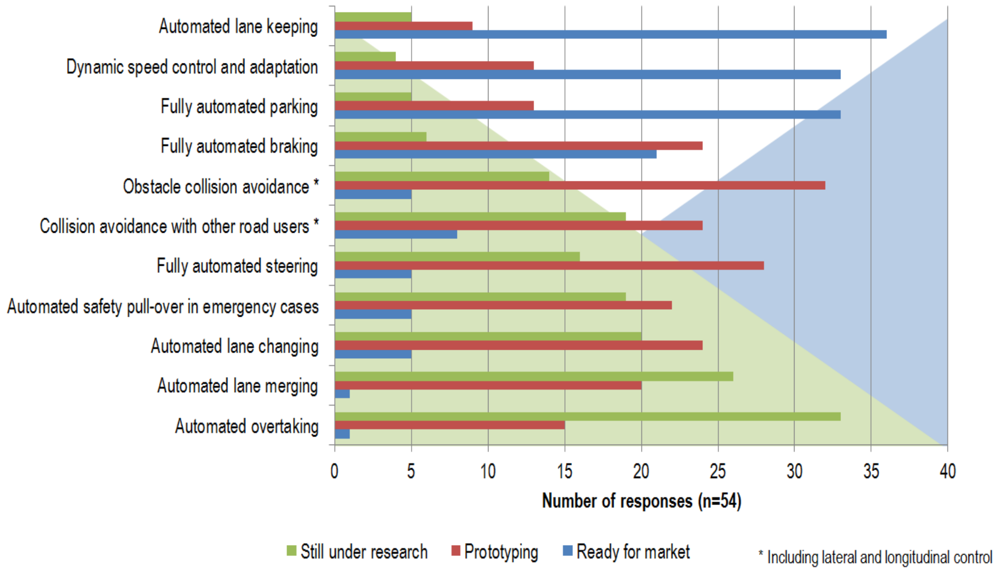


Figure 24: Number of online survey responses regarding the technical readiness of ADS

The survey further focused on three distinct groups of ADS, which differ in their purpose of use, technologies applied and market readiness:

1. *Lane assistance*, aiding drivers by performing actions such as keeping or changing the lane automatically, comprising of the subsystems "Automated lane keeping", "Automated lane changing", "Automated lane merging" and "Automated overtaking" given in Figure 24.
2. *Collision avoidance*, aiding drivers in avoiding collisions with other traffic participants (e.g. vehicles, pedestrians, cyclists) or other obstacles by means of automated steering or braking, including the systems "Obstacle collision avoidance", "Collision avoidance with other road users"
3. *Speed control systems*, aiding drivers by adapting the vehicle's speed automatically, based on legal speed limits or other external factors, such as road alignment, weather conditions, congestion, road works etc. (Fancher et al., 2004; Yamamura et al., 2001). The group consists of Intelligent Speed Adaptation (ISA), dynamic speed adaptation and curve speed control.

It is significant to mention that the three groups analysed in this study do not cover the whole range of subsystems in automated transport. For example, automated parking systems, truck platooning, re-routing or automated docking systems for public transport are also available or under development.

To the question regarding to which extent certain factors influence the performance of specific ADS groups, the respondents gave a rating from very high to very low regarding the influence of a factor, e.g. poor visibility. To present the rating on a linear scale, the following weighting function has been applied:

$$Weighted\ influence\ of\ a\ factor = \frac{I - N}{3N} \tag{4}$$

with N as the total number of responses for each factor, and $I = \sum_{i=1}^{4}(n_i \cdot i)$, with n as the number of responses for each rating category i from 1 (very low) to 4 (very high influence). This 4-point scale was preferred over a 5-point or Likert scale to avoid a neutral response, which keeps the number of analysable answers higher. Having a mid point on the scale would mean that the respondent had the possibility to opt out of the question and that a decrease of responses is likely. Resulting in a scale from 0 to 1, the influence weights were simply divided into three equal ranges indicating 'low', 'medium' and 'high influence' for each automated driving subsystem (see Figure 25 to Figure 27).

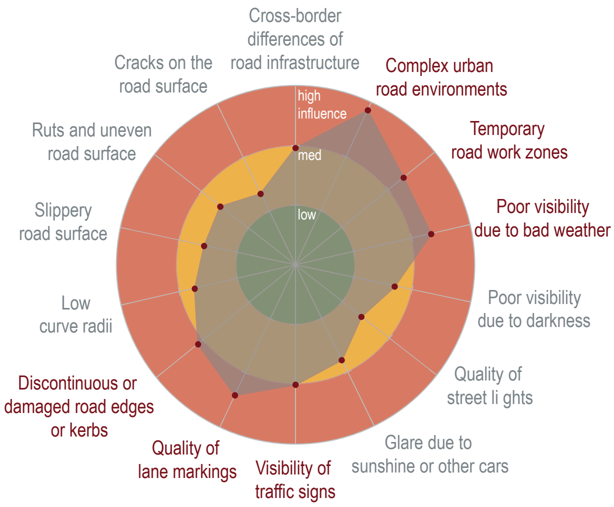


Figure 25: Factors influencing the performance of lane assistance systems

Considering the fact that lane assistance systems (see Figure 25) are mainly based on vision sensors, it is not surprising that poor visibility and low quality of lane markings and road edges are among the factors with high influence. Complex urban environments were rated with the highest influence, meaning that lane assistance systems might struggle with obstructed intersections, parked vehicles or crossing pedestrians and cyclists. Temporary work zones may constitute a problem for lane assistance of ADS, since temporary markings often replace lane markings.

As seen in Figure 26, the ratings for collision avoidance systems indicate that visibility due to bad weather is still an issue, but not the visibility and quality of road infrastructure and equipment. For this subsystem of ADS, the interaction between road users comes into play, which could obviously be a challenge in complex urban environments.

Speed control systems are mainly based on traffic sign detection and digital maps, which are influenced by changing environments, such as it happens on temporary work zones or urban roads and intersections (see Figure 27).

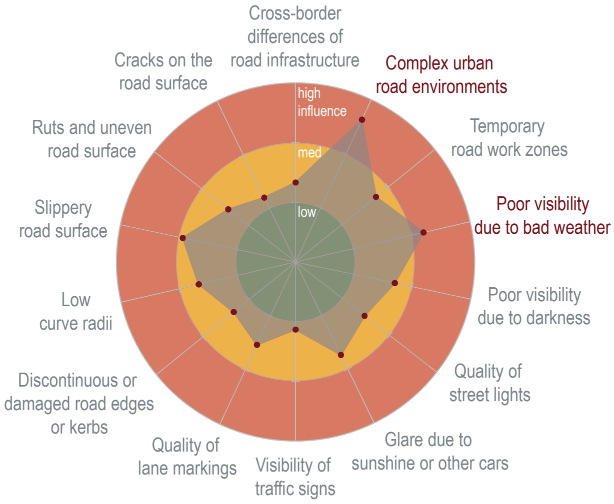


Figure 26: Factors influencing the performance of collision avoidance systems

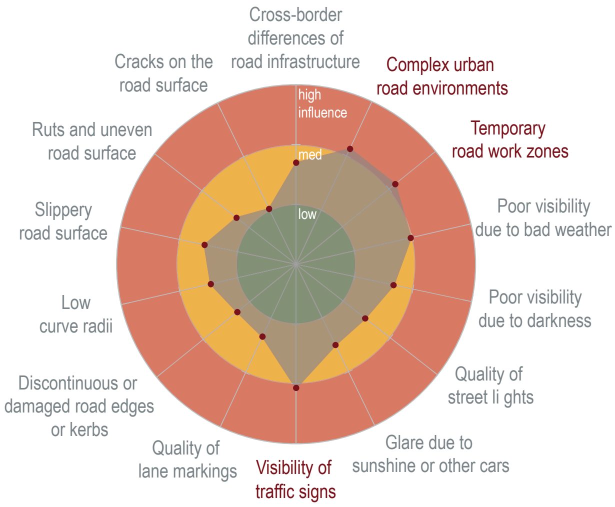


Figure 27: Factors influencing the performance of speed control systems

Summarizing the web survey results, the main challenges for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather conditions. Road surface characteristics, road alignment and lighting were rated as minor influencing factors. Intersections are assumed to be complex road environments for ADS due to the likely occurrence of other road users and crossing traffic. Given the survey results, it can be concluded that intersections with poor visibility due to obstructions or bad weather, or unclear lane markings and traffic signs pose particular problems. This helped to select the variables from the in-depth crash database, as explained in the following section, as well as to define the parametric variation in the simulation study.

4.2 Initial analysis of in-depth accident data on junctions

As an essential task of the scoping study, in-depth crash data from OTS was collected to enhance the findings given in Section 2.3, where the general safety aspects at intersections have been discussed. While the macroscopic analysis of European CARE data revealed the proportion of accidents and injured persons at junctions compared to all cases, this task is carried out to explore the data in detail and to set the scope for the upcoming analyses.

4.2.1 The OTS study

The data used for this thesis stems from a project called OTS (On-The-Spot), which was commissioned by the UK Department for Transport and the Highways Agency (HA). It aimed to establish an in-depth research database of a representative sample of road accidents in the UK, to better understand the cause of accidents and injuries (Hill et al., 2001). Two crash investigation teams collected data from the years 1999 to 2010. One team was located at Loughborough University covering the South Nottinghamshire area in the East Midlands, and the other at the Transport Research Laboratories (TRL) covering the Thames Valley region. See the locations of team in Figure 28. The teams were responsible for collecting information at the scene of the accidents or, when the accidents already occurred, by liaison with emergency services, hospitals and local authorities. To arrive at the accident scene as quick as possible, the teams had a direct link with the local police, and response vehicles driven by an OTS police officer were used (Cuerden et al., 2008).

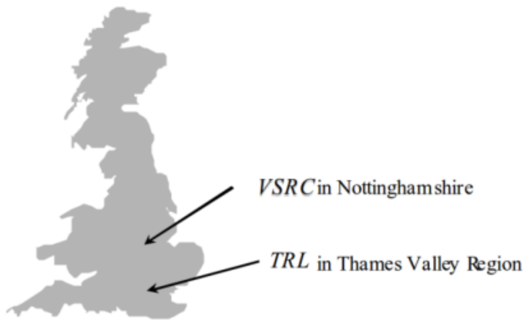


Figure 28: Locations of the OTS investigation teams (Mansfield et al., 2008)

The collected data was structured into eleven hierarchical sections as depicted in Figure 29. The "scene" level contains approximately 200 variables, e.g. relating to the date, time and location of the accident, general environmental conditions and possible contributory factors. For each scene, there is at least one "path" record, defined by the direction of travel. If two vehicles collided while travelling in the same direction (e.g. rear-end collisions), they share one path, while typical head-on collisions would have two paths. The path section includes data about road layout, surface condition, traffic control or potential distractions or obstructions.

For each vehicle on each path, there is a "vehicle" record, comprising information on the vehicle type, condition and safety features installed. Pedestrians and bicycles are also counted as vehicles. There are several subsections for each vehicle, such as accident position or information on involved persons ("human"), which is further subdivided into "medical", "injury", "interaction" and "questionnaire".

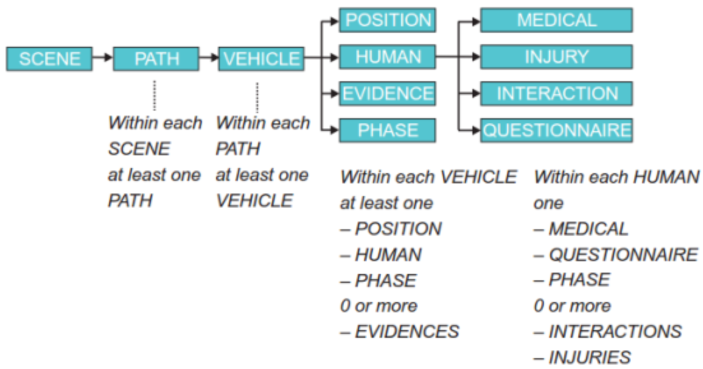


Figure 29: Hierarchical structure of the OTS database (Mansfield et al., 2008)

In addition to OTS, several other crash data collection programmes for different purposes have been completed in the UK. For example, CCIS (Cooperative Crash Injury Study) is a study for car accidents only with focus on injury causation and crashworthiness of vehicles (Department for Transport, 2006). For causes and personal injury consequences of crashes involving trucks and bus crashes, the Heavy Vehicle Crash Injury Study (HVCIS, Department for Transport, 2003) was carried out using a similar methodology to CCIS. As a follow-up programme addressing each of the areas of the former work, the project RAIDS (Road Accident In-Depth Studies) was initiated in 2012. Similar to OTS, TRL attends the scene of crashes and examine vehicles that have been involved in police-reported collisions in the Thames Valley and Hampshire regions, while the Loughborough University team focused on Nottinghamshire. The data structure is also similar to OTS. However, RAIDS was found to be too sparse for the desired data query with only 75 cases to investigate. So, CCIS and OTS were among the candidates, but difficult to merge data-wise. Instead of conducting two separate analyses, and because CCIS only contains car accidents, OTS was chosen as dataset, having a proper sample size and a sufficient number of attributes included.

4.2.2 Data collection and processing

OTS is part of the RAIDS project, whose data query and export tool was used to download all necessary data elements including collisions with the following prerequisites:

- Junction type = "T or staggered junction", "Crossroads", "Multiple junction", "Other junction" or "Using private drive or entrance".
- Police Accident Severity = "Fatal", "Serious" or "Slight"

It is important to mention that although the police reported a certain injury level, this might have been adapted by the crash investigation team based on more precise evidence. Therefore, there can be uninjured cases within the dataset according to the OTS investigations, although the police reported an injury accident.

As mentioned before, roundabouts were excluded from this study. The junction types included comprise signalised and non-signalised junctions of different shapes. This query resulted in 1,056 crash cases (see Table 6), including more than 400 variables. The entire export was securely stored in an MS Access database for further queries and easier handling. All tables are related via key fields corresponding to their respective ID fields (see Appendix C).

Table 6: Results from the initial OTS query

Number of accidents:	1,056
Number of involved persons:	3,112
Number of injured or killed*:	1,278
Number of fatalities*:	36
Number of seriously injured*:	198

(* Police accident severity)

It was decided to analyse the data on the driver level, i.e. every sample corresponds to one driver involved in a crash, regardless if he/she was injured or not. This also means that every sample contains one vehicle. Consequently, if two or more vehicles are involved in the same crash, the underlying crash and environment data is simply duplicated. Furthermore, there should be at least one car (including VANs, minibuses and SUVs) involved. This principle is illustrated in Figure 30 and required a second query from the MS Access database as follows:

- Seating position of occupant = "Driver / Rider"
- At least 1 vehicle = "Car"
- Total number of vehicles > 1

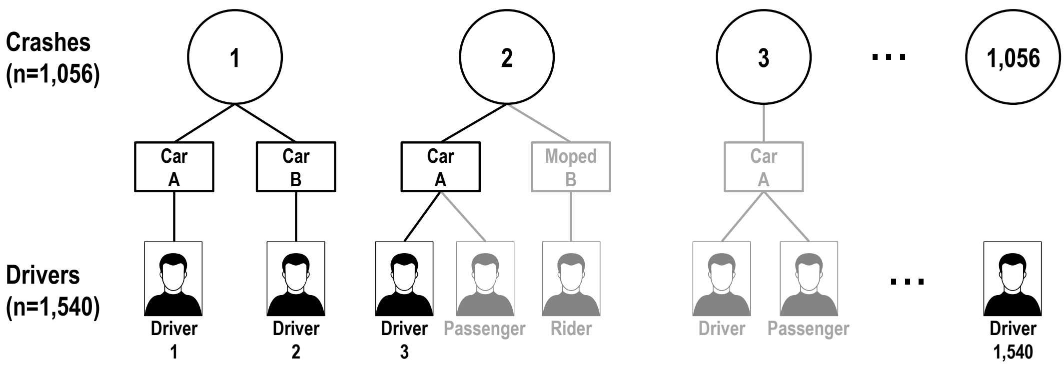


Figure 30: Principle of crash data filtering to obtain one driver per sample. Single vehicle accidents are excluded and only car drivers are selected.

The requirement of more than one vehicle means that single-car accidents were intended to be excluded, because speeding, fatigue or other human causation for single vehicle accidents are likely to be minimised by ADS and are therefore not relevant to the thesis. Also, single-vehicle accidents do not belong to the common crash types at intersections (Abdel-Aty et al., 2006; Polders et al., 2015). This additional query resulted in a final sample size of 1,540, i.e. 1,540 drivers, regardless if they are injured or uninjured. Compared to an analysis on the crash level, with 1,056 cases, the sample size has been increased that way. The background of the analysis gives another reason for scaling it to the driver level: In the further steps, the thesis will investigate safety risks involving automated vehicles instead of drivers. To this end, it is necessary to know the critical situations to be handled by drivers nowadays, as they are likely to happen to automated vehicles as well. Each sample is thus associated with an ego car, later denoted as car A, which collides with a secondary vehicle or road user, later denoted as B.

To filter out the drivers, the field "Seating position" was relevant, as shown in Figure 31. It further gives the distribution of the gender showing that approximately two-thirds of all drivers were male.

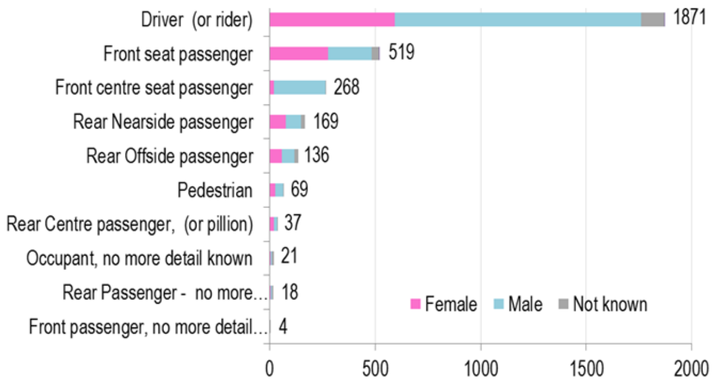


Figure 31: Number of all involved occupants by seating position and gender (OTS, n=3112)

4.2.3 Generalizability of the OTS data sample

The following charts were produced to explore the queried data and most important accident figures. This exercise was not done to prioritise or to identify the highest risks, but to compare OTS data to the National UK data from CARE in terms of generalisability of analysis results as well as to get a picture about the proportions of accidents and injury severities for different junction types.

When comparing data from OTS with CARE, it must be noted that some definitions differ, e.g. the junction types. Therefore, only the two most frequent junction types "T or staggered" and "Crossroad" have been compared (see Figure 32 and Figure 33), for all injury accidents, fatal accidents and serious injury accidents, respectively). It can be seen that the proportions highly correspond between the two datasets. The largest difference can be observed for the fatal accidents (see Figure 33), where in the CARE data, the T or staggered junction has a higher proportion and the crossroad a lower one than the OTS dataset. For better comparison, the police injury severity has been used instead of the severity assessed by the OTS investigators.

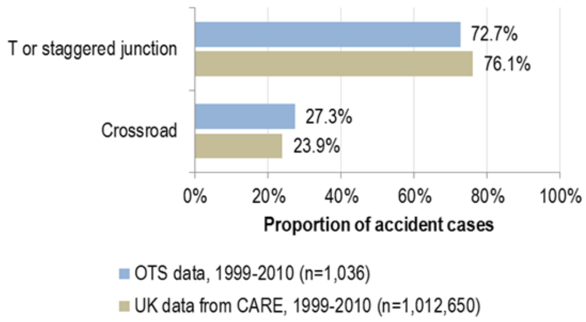


Figure 32: Comparison of the proportion of all accidents at two different junction types between CARE and OTS (other junction types not included due to different definitions)

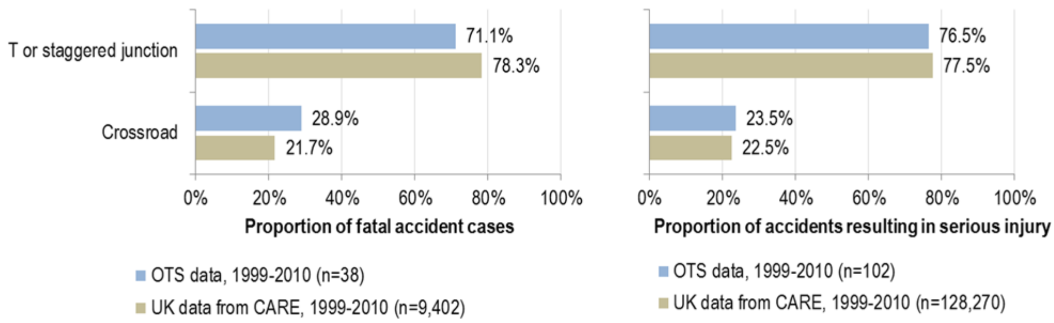


Figure 33: Comparison of the proportion of accidents with fatalities (left) and serious injuries (right) at two different junction types between CARE and OTS (other junction types not included due to different definitions)

A χ^2-test (using an α value of 0.05) was made to see whether the proportions of injury levels in the two datasets are the same. In Table 7, it can be seen that the null hypothesis, i.e. the hypothesis that there is no difference between OTS and CARE, has to be rejected when including all three injury levels separately. The p-value is lower than $\alpha = 0.05$. Since it is assumed that this is due to the difference in fatalities and their low sample size, severe and fatal injuries were combined to determine the effect on the chi-square value. Apparently, the null hypothesis becomes now valid with a p-value much greater than 0.05. This means that the proportions between CARE and OTS are similar in a statistically significant manner. Since it is not the aim of this study to investigate injury levels in detail separately, the difference in the fatality proportions can be considered negligible.

Table 7: Results from the χ^2 homogeneity test (alpha=0.05) on OTS and CARE

Junction type	Variables	Chi-Square Value	Degrees of freedom	p-value
Crossroads	Fatal/severe/slight	33.80	2	0.00
	Fatal+severe/slight	0.04	1	83.97
T or staggered	Fatal/severe/slight	58.08	2	0.00
	Fatal+severe/slight	0.01	1	94.33

Figure 34 and Figure 35 give the percentage of the injury severities for the respective junction types, by stretching each bar to 100 percent. For example, the bar charts show that 3.9 percent of all accidents at crossroads ended fatally in the OTS dataset, while the CARE dataset gives only 0.8 percent. This could be explained by the low sample size of fatalities in the OTS data and reflects the results from the χ^2-test. However, the proportions of accidents resulting in slight injury have almost identical values for T or staggered junction and crossroads. It must be mentioned that only the upper two bars can be compared directly, since "Other junction" have a different definition and "Private drive or entrances" are not included in the CARE data.

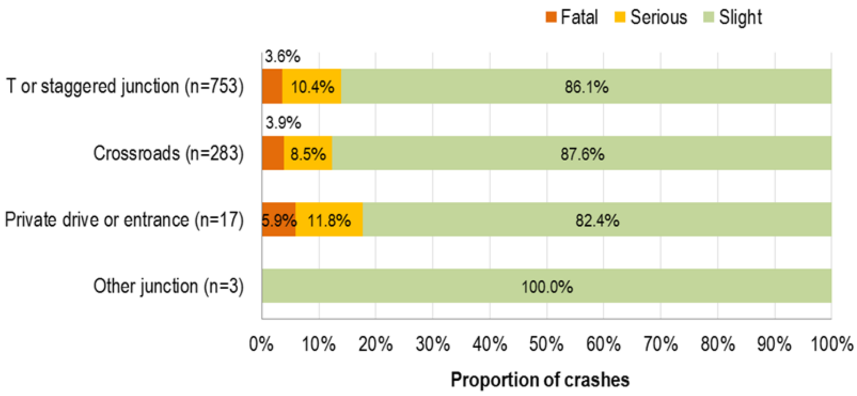


Figure 34: Proportion of queried OTS junction accidents by max. injury severity and junction type, 1999-2010 (n=1,056)

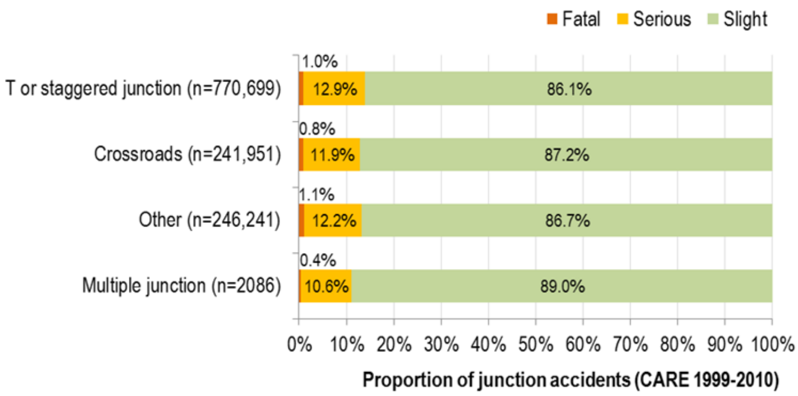


Figure 35: Proportion of queried CARE junction accidents by max. injury severity and junction type, 1999-2010 (n=1,260,977)

When exploring the data, it was interesting to see that the proportions of the injury levels "Fatal" and "Serious" are similar for the different junction types "T or staggered junction" and "Crossroads". While the private drive crashes ended fatally in 5.9 percent, crossroad and T-junction crashes have a fatality proportion below 4 percent in the OTS dataset.

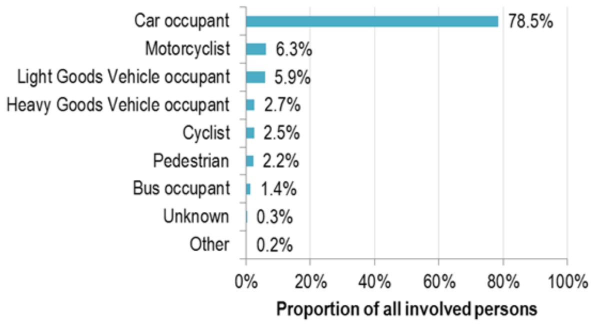


Figure 36: Proportion of involved persons in junction accidents, by road user type (OTS, 1999-2010, n=3,112)

In Figure 36, the proportion of all involved persons in the exported OTS accidents is given by road user type. Not surprisingly, car occupants show the highest proportion with more than 78 percent, also due to their higher exposure, followed by motorcyclists and lorry occupants.

Although car occupants are most frequently involved in junction accidents, they are among the safest when looking at the injury levels, as depicted in Figure 37. Only 0.5 percent of all involved car occupants were fatally injured, and more than the half remained uninjured. In comparison, motorcyclists died in 6.7 percent, and pedestrians in 16.1 percent of all cases. Interestingly, the proportion of fatalities among cyclists is relatively low with 1.4 percent, but they were injured in approximately 96 percent of all cases, having the highest amount of slight injuries.

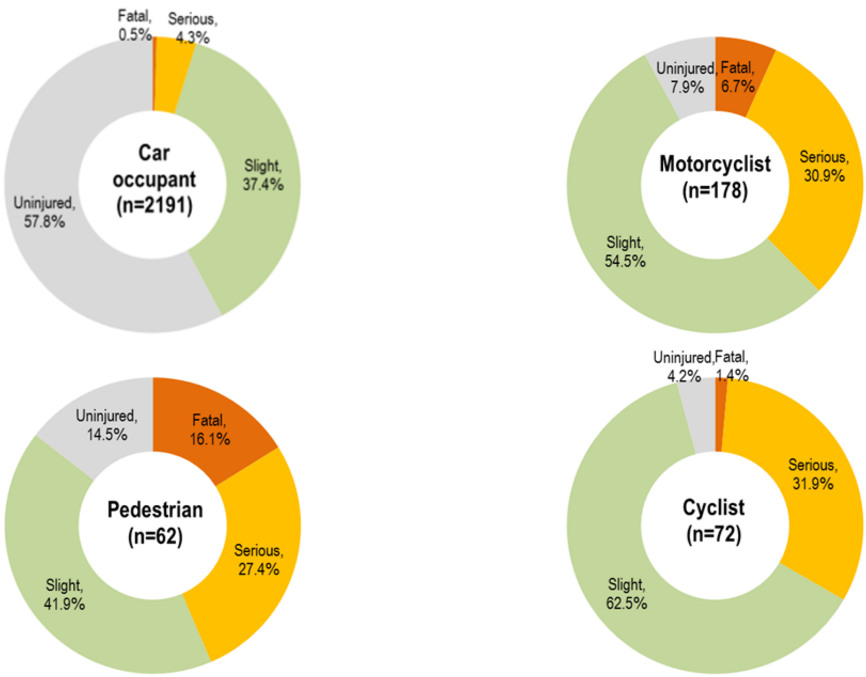


Figure 37: Proportion of involved persons in junction accidents, by OTS injury severity and road user type, excluding unknowns (OTS, 1999-2010)

The proportion of fatally and seriously injured persons by road user type was compared to CARE, as given in Figure 38. In the OTS dataset, car occupants have the highest proportion of fatal and severe injuries, followed by motorcyclists and pedestrians. CARE data show a different picture. It seems like pedestrians are underrepresented in the OTS dataset.

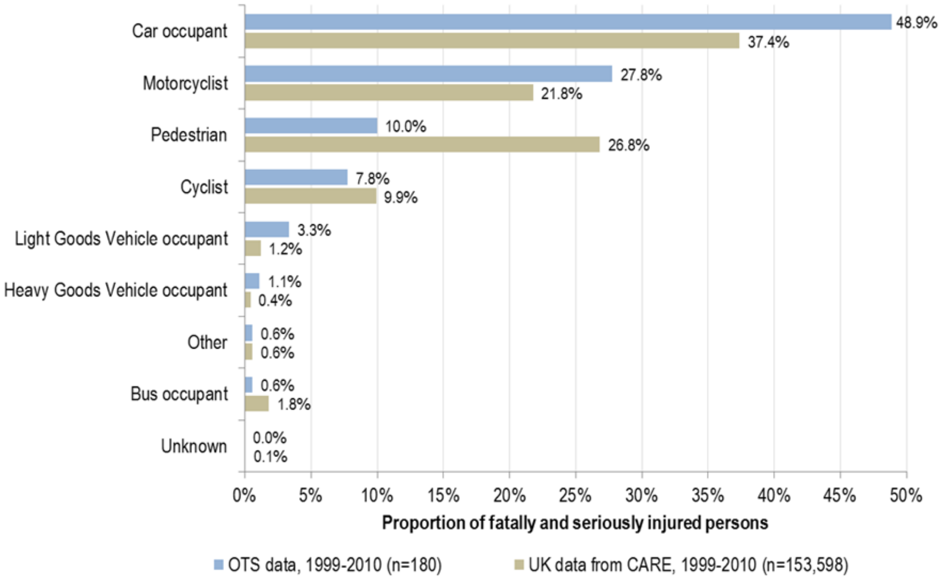


Figure 38: Comparison of the proportion of fatally and seriously injured persons (police severity) at junctions (crossroads and T/staggered) between CARE and OTS, by road user type

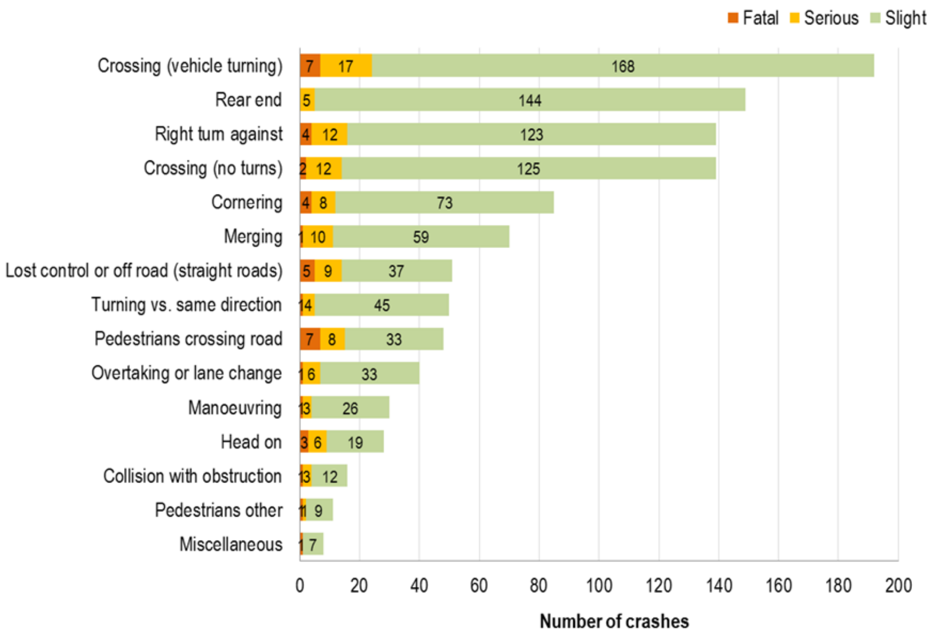


Figure 39: Number of OTS junction crashes, by collision type and max. injury (police severity) (n=1,056)

The collision types in the queried OTS dataset are given in Figure 39. For better readability, only the collision type labels are plotted, and not the subtypes. It shows that crossing collisions while one of the involved vehicles was turning are the most frequent types, followed by rear-end collisions, collisions at right turns and right-angle collisions (i.e. Crossing with no turns). Not surprisingly, pedestrian crossing accidents do not happen that often, but when they happen, the injury severity is among the highest. It can further be observed that head-on collisions or overtaking and lane change collisions occurred less frequently.

4.2.4 Attribute selection and reduction

In the previous sections, the underlying database queries and data processing steps have been explained. What follows is a description of the variable selection process. The second query, resulting in 1540 samples, originally consisted of more than 400 variables including categorical, continuously numeric as well as binary types. The number of variables was further reduced according to the following steps, similar to a study by Uno et al. (2013):

1. Include only variables that fit the scope of the study, e.g. not relevant were weekday or time of the crash, occupant data such as age or gender, vehicle damage or detailed injury data of different body parts.
2. Exclude variables with low variance, because they would fail to make a positive impact on model performance. In this study, all observations with more than 95 percent same values were excluded.
3. Group or combine highly correlated variables, e.g. OTS injury severity and police injury severity.
4. Exclude variables having unknown values in more than 30~percent of all samples.

Following this reduction process, the number of variables was reduced to 41, which were grouped according to the original OTS data hierarchies „scene“, „vehicle“ and „path“. The „scene“ variables include general attributes about the crash, such as collision type and maximum injury of all involved persons. The „vehicle“ variables are related to the pre-crash and collision circumstances from the perspective of the individual vehicle, i.e. driver, and includes for example the precipitating factor attributed to the vehicle, driver injury level or the pre-impact manoeuvre. The „path“ variables describe the road environment, e.g. junction type, weather, traffic density or speed limit.

4.3 Analysis of processed OTS data

After post-processing the data, the most important variable distributions were analysed. In total, the dataset contains 619 injured car drivers (OTS injury level, excluding unknowns), with 91 percent slight injuries, 8 percent serious injuries and around 1 percent fatal injuries in total. For further analyses, the OTS injury level is preferred to the injury level recorded by the police in the first place, because it is based on more accurate evidence. Figure 40 shows a comparison of both records, where the numbers of unknown injury level differ the most. Cases, where the injury level could not be estimated in the OTS project, were set to “unknown”, which resulted in less “uninjured” and “slightly injured”.

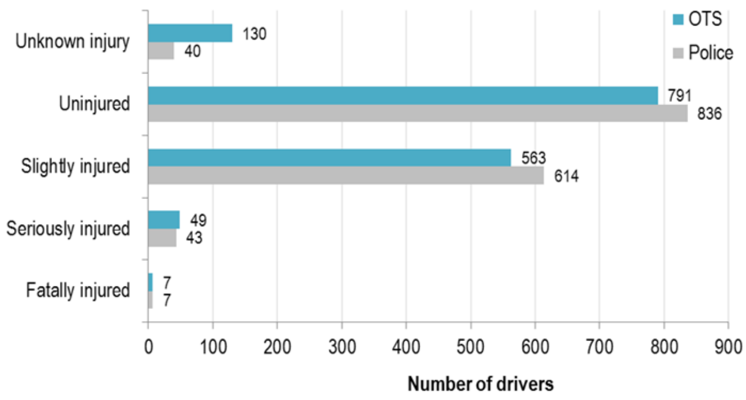


Figure 40: Comparison of injury levels from OTS and police records

It was found that there is a high number of uninjured drivers (791). This means that in the underlying accident case of these samples, the other driver(s) or road users (e.g. pedestrians) involved were injured, but not the driver of this sample. The proportion of the car drivers’ injury levels by junction type are depicted in Figure 41. T-junctions or staggered junctions have the highest share of injured car drivers, followed by crossroads. Regarding the percental distribution of injury levels, it can be seen that T-junctions and crossroads have similar values, with the crossroads showing a slightly lower proportion of serious injuries. There were no serious or fatal injuries recorded for private drives or entrances and other types of junctions, but the sample size is low.

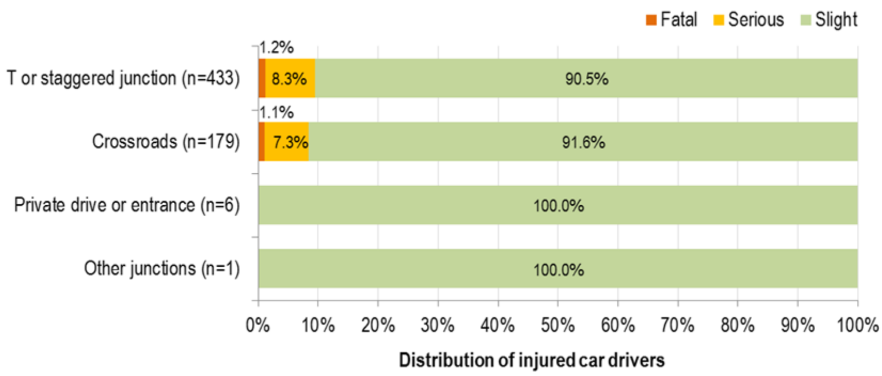


Figure 41: Proportion of injured car drivers by OTS injury level and junction shape (OTS, n=619)

Figure 42 shows the number of car drivers for each of the collision code letters. The letter represents the supercategory of a collision, whereas the collision code number expresses the subcategory. Interestingly, rear-end collisions (letter F) are the most frequent collision types with 327 drivers involved, followed by turning collisions (letters J and L). This is a different finding to what Molinero Martinez et al. (2008) reported. Accordingly, collisions, where vehicles cross the trajectory of an opponent vehicle, are the most frequent. However, the study included EU-27, while this analysis only focuses on two regions in the UK, with roundabouts excluded and at least one car involved.

Pedestrian accidents occurred to 3 percent of all drivers (letters N and P). Conditioned by the OTS query, where single-vehicle accidents were excluded, the dataset comprises low numbers for the collision letters C (Lost control or off road) and E (Collision with obstruction). An overview of all collision code letters and numbers is given in Appendix B.

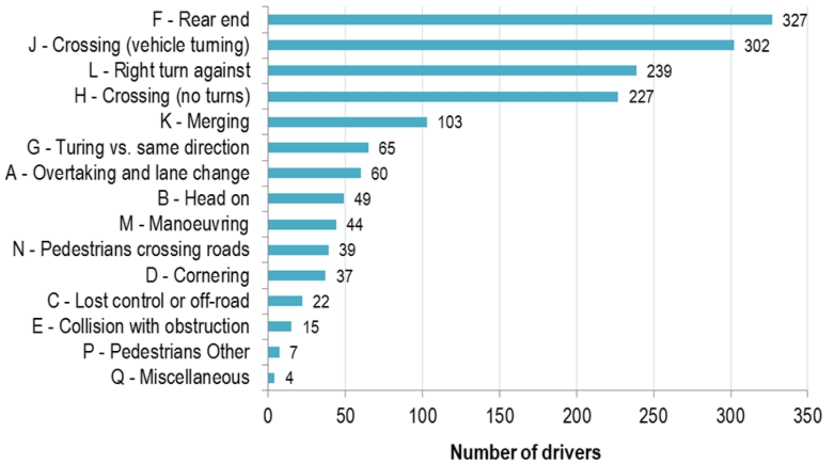


Figure 42: Number of car drivers by collision letter (OTS, n=1,540)

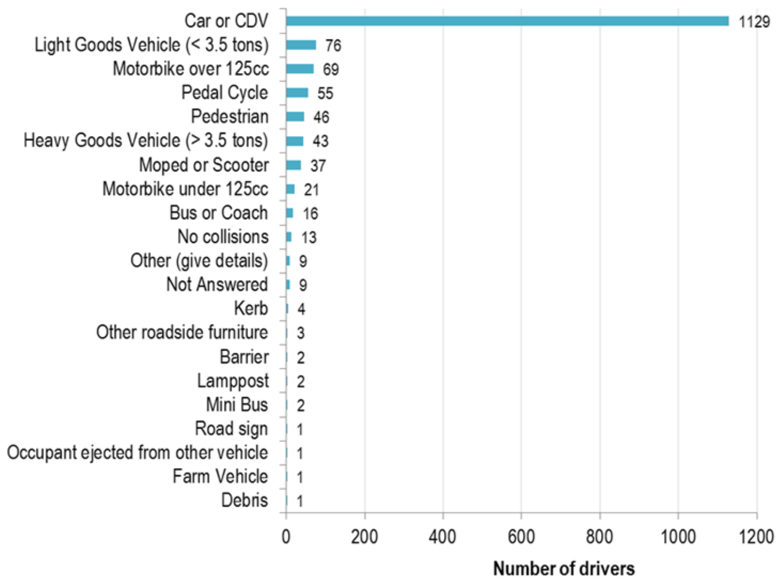


Figure 43: Number of car drivers by first interaction (OTS, n=1,540)

The variable "First interaction" reveals the potential collision partners for the car drivers (see Figure 43). An interaction could also mean a near-miss or no impact at all. However, only 13 of the 1,540 car drivers had no collisions as their first interaction, which equals 0.8 percent of all samples. By far, other cars or car-derived vans (CDV) are the most frequent first interaction partners for car drivers with 73 percent (1,129 out of 1,540). Bicyclists (55) had slightly more interactions with car drivers than pedestrians (46) and interactions with objects are negligible.

In the OTS study, precipitating factors indicate the trigger for the crash, meaning that it is most likely that the crash could be avoided if those factors did not occur. In Figure 44, the ten most frequent precipitating factors are plotted. There is also a variable for a second precipitating factor included in the OTS data, but it was discarded due to a high number of unknown or "not answered" values. Around 40 percent of all queried intersection crashes were caused by a failure to give way, which corresponds to the two most frequent collision types involving crossings with no turns. "Failed to stop" and "failed to avoid object or vehicle on carriageway" can be associated with rear-end collisions.

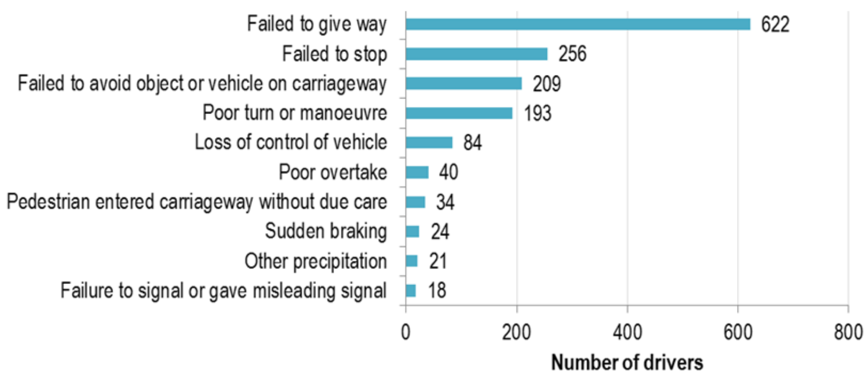


Figure 44: Number of car drivers by the Top 10 precipitating factors (OTS, n=1,501)

4.4 Conclusions

To set the scope of the thesis, an expert survey was conducted including questions on the role of road infrastructure, market readiness as well as to which extent certain factors influence the performance of selected automated driving functions. In summary, the main challenges found for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather conditions. Road surface characteristics, road alignment and lighting were rated as minor influencing factors.

In addition to the survey, a comprehensive set of junction crash data was collected to identify common crash figures and circumstances. 1,056 junction accident cases were exported from the OTS (On-The-Spot) database, including 3,112 involved persons, of which 1,278 were injured or killed. A comparison of OTS data with National UK data queried from the CARE database showed high similarities in terms of the proportions of accident cases by junction type as well as by injury level. Although there are differences in the distribution of fatalities due to the low number of OTS samples (n=36), the queried OTS data can be considered generalizable and reliable for further investigations in the UK. However, a direct extrapolation of the data to other countries is not recommended due to the differences in road layout, traffic regulations or driving styles. For example, the data is based on left-hand driving, which cannot be directly used for analyses in countries with right-hand driving, because other scenarios may arise. Also, the design of junctions in the UK differs from the layout in other central-European countries, e.g. pedestrian fences and traffic islands are more frequently used, road markings and pedestrian crossings are different as well as speed limits.

The OTS data was prepared to categorise and reduce variables, which resulted in a number of 41 categorical variables. The data was further analysed on the driver level so that each sample corresponds to one car driver, and consequently to one vehicle. This was done, because this thesis focuses on critical scenarios for automated vehicles depending on their environment, and because the sample size could be increased that way.

More than three-quarters of the involved persons were car driver occupants, of which 38 percent were injured. The road user group showing the highest injury rate are pedestrians. In general, car occupants are among the safest road users, together with HGV occupants.

An initial descriptive analysis of the causation factors showed that a failure to give way, to stop and to avoid are the top three precipitating factors for junction accidents.

The data was further analysed on the driver level so that each sample corresponds to one car driver, and consequently to one vehicle. This was done, because this thesis focuses on critical scenarios for automated vehicles depending on their environment, and because the sample size could be increased that way.

While a larger sample size is desirable in most cases, a lower number of variables is recommended to avoid high dimensionality in data. To this end, data has been reduced to 41 categorical variables by following a pre-defined reduction process, e.g. by excluding variables with many missing values or by grouping highly correlated ones. This resulted in a set of mainly uncorrelated variables describing relevant information.

Preparing and analysing the queried in-depth data built the basis for clustering on a multivariate level. Sufficient data was prepared to carry on studying safety-critical scenarios at junctions. To get a deeper insight into the high-dimensional data, data mining techniques are applied in the second study to reveal undiscovered structures and thus safety-critical situations at junctions.

5 Study 2: Identifying critical scenarios from crash data

The previous chapter explained how the in-depth accident data was processed and analysed. The descriptive analysis is now complemented by a data mining technique to derive safety-critical scenarios. This chapter describes how the accident data was clustered and how the scenarios were identified. The study addresses the research question 3 (see Section 1.2) and was published in the journal Accident Analysis and Prevention (Nitsche et al., 2017).

5.1 Problem definition

Virtual vehicle testing has many advantages, but especially regarding automated vehicles, which have to fulfil functional safety requirements, there are still some open questions: When is a system "good enough" for public roads? Which scenarios must be passed in the MIL, SIL and HIL tests? How many scenarios must then be physically re-tested on real roads?

Concerning road safety, it is still not clear what impact automated vehicles will have on crash risk, and what kinds of (new) risks they might cause. In particular, the safety risks coming with a mixed vehicle population, namely traffic with both automated and driver-operated vehicles are still subject to research. Although automated cars use sophisticated onboard sensors to recognise their environment, they have limitations, e.g. in challenging urban traffic situations, inclement weather conditions or when facing unexpected behaviour of traffic participants. In study 1, an expert survey was conducted including questions on the role of road infrastructure, market readiness as well as to which extent certain factors influence the performance of selected automated driving functions on public roads. In summary, the main challenges found for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather conditions. Road surface characteristics, road alignment and lighting were rated as minor influencing factors.

Three-legged and four-legged junctions are high-risk areas, which future automated cars should be capable to pass safely. Therefore, intersections play a particularly important role in testing assisted and automated driving. Automated vehicles should be capable of safely manoeuvring through an intersection and of avoiding or mitigating a collision. One of the main research gaps addressed by this work is that there are no standardised procedures for evaluating automated driving systems in junction environments and that the key scenarios need to be defined. Hence, this study addresses research question 3.

As explained in Section 3.3, the method for study 2 clusters historical accident data to understand the critical situations and factors at road junctions. Similar research has been conducted (INTERSAFE, 2005; Molinero Martinez et al., 2008; Plavsic, 2010; Polders et al., 2015; Wiltschko, 2004). However, the usage of k-medoids clustering and association rules in this context is novel. Due to a lack of accident data involving automated vehicles, it is a reasonable starting point to analyse historical accidents with human drivers, assuming there is a certain overlap to the crash risk for ADS. It is not the goal of this study to identify new, still unknown scenarios that will arise with a higher penetration of automated vehicles on the road. Instead, the study is preparatory research to develop and demonstrate a simulation and evaluation framework (see Chapter 6). The scenarios obtained in the underlying study will help to reduce the possible number of model parameter variations, such as vehicle trajectories, velocities or road and junction parameters.

5.2 Preparation of OTS data for clustering

In Section 4.2, it is explained how the accident data was exported, filtered, reduced and grouped. However, one important step is missing: The variable coding. The original data contains variables in the following format: "Maximum injury level = Serious" from the four possible values uninjured, slight, serious and fatal. For the further calculations, all variables were converted to the binary-coded format as follows:

	Uninjured	Slight	Serious	Fatal
Maximum injury	0	0	1	0

Consequently, this resulted in much more attributes, as each possible value was assigned to its own column, but it is a necessary step for applying most clustering algorithms.

The high number of attributes of the pre-processed OTS dataset made it necessary to further prepare the data for clustering. Usually, the fewer attributes, the easier the cluster results can be interpreted. Initial experiments with a varying number of attributes as input showed that the performance of the k-medoids method suffers from a higher dimensionality. Therefore, all attributes were divided into two levels as follows:

1. **First level (5 variables, 25 attributes, see Table 9):** This level of attributes was used as input for the k-medoids clustering. The idea is to derive clusters based on a set of main collision attributes first, before association rule mining is applied to each cluster with the second level attributes.
2. **Second level (10 variables, 61 attributes, see Table 11)**: This level adds more detailed attributes on road infrastructure and accident causation to the level-1 attributes. They are intended to help telling a "story" describing each cluster by association rule mining.

The first-level attribute groups are given in Table 8, including a description of the group. For this study, these five attribute groups were used to separate the data space into easily interpretable clusters. Apart from that, this particular set of attributes resulted in the highest validity in comparison to other combinations of attributes. Hence, the original number of 41 accident variables was further reduced to 20 (5+15). Furthermore, all attributes describing unknowns (e.g. “Area=Unknown”) and all samples that are true for these attributes were removed.

Table 8: Collision attribute groups (first level)

Category	Hierarchy	Description
Maximum injury	Scene	The maximum injury of all persons involved in the accident
Junction shape	Path	The detailed shape of the junction, attributed to the vehicles' path
First interaction	Vehicle	Road user type or object, which the vehicle first interacted with
Manoeuvre	Vehicle	Action of the vehicle immediately before the accident
First point of impact	Vehicle	First point to come into contact with another vehicle, pedestrian or other objects

The maximum injury gives information on the severity of the clustered accidents. The junction shape does not only differentiate between T-junction, crossroads and other junction, it also gives the shape from the perspective of the driver's path. For different vehicle paths of an accident case, there are different ways to describe the junction shape, although all involved vehicles approached at the same junction. All attributes of the first-level groups are listed in Table 9, including the short name used for presentation, a description of the attribute as well as count and relative frequency.

Table 9: Collision attributes (first level)

Category	Short name	Description	Count*	Rel. frequ.
Max. injury (of all persons involved in the crash)	MaxInj=Uninjured	No person injured (OTS injury level)	196	14.8%
	MaxInj=Slight	At least one person slightly injured (OTS injury level)	919	69.4%
	MaxInj=SeriousFatal	At least one person seriously or fatally injured (OTS injury level)	210	15.8%
Junction shape (attributed to the vehicle's path)	JctShp=X-minJoin	Road continues straight on with (minor) road joining from the left and right (crossroad)	224	16.9%
	JctShp=X-brkMaj	Road is temporarily broken by a (major) road passing across the path of the vehicle (Crossroad)	144	10.9%
	JctShp=NoJct	No junction present	20	1.5%
	JctShp=Other	Private drive, entrance or other junction type	7	0.5%
	JctShp=T-minLeft	Road continues straight on with (minor) road joining from the left	350	26.4%
	JctShp=T-minRight	Road continues straight on with an additional (minor) road joining from the right (T-Junction)	309	23.3%
	JctShp=T-termMaj	Road terminates with a (major) road passing across the vehicles path (T-Junction or acceleration lane)	271	20.5%

First interaction (Road user type or object which the vehicle first interacted with)	1stIntAct=Car	Driver interacted with another car	987	74.5%
	1stIntAct=LGV-HGV	Driver interacted with a large or heavy goods vehicle	97	7.3%
	1stIntAct=PTW	Driver interacted with a powered two-wheeler (motorcycle or moped)	115	8.7%
	1stIntAct=Other	Driver interacted with another type of vehicle or object	37	2.8%
	1stIntAct=Cycle	Driver interacted with a bicyclist	50	3.8%
	1stIntAct=Pedestrian	Driver interacted with a pedestrian	39	2.9%
Manoeuvre (Action of the vehicle immediately before crash)	Manvr=GoingAheadOther	Driver was going straight ahead	781	58.9%
	Manvr=TurnL	Driver was turning left	59	4.5%
	Manvr=TurnR	Driver was turning right	79	6.0%
	Manvr=WaitTurnR	Driver was waiting to turn right	353	26.6%
	Manvr=Other	Driver was reversing, doing a u-turn, overtaking, undertaking, held up or waiting to turn left	53	4.0%
First point of impact (First point to come into contact with another vehicle, pedestrian or other object)	1stImpact=Back	First point of the impact was the car's back	126	9.5%
	1stImpact=Front	First point of the impact was the car's front	674	50.9%
	1stImpact=Nearside	First point of the impact was the car's nearside	218	16.5%
	1stImpact=Offside	First point of the impact was the car's offside	307	23.2%

**The count represents the number of samples on a driver level, i.e. the number of drivers related to the respective attributes.*

As described above, the second-level attributes deliver more information on the accident environment and causation. Most of the additional attribute groups in Table 10 are related to the vehicle's path describing the road layout, e.g. road type, speed limit or curvature. The attribute groups "collision code", "precipitating factor" and "driver injury" were added to the list to understand the accident circumstances better.

Table 10: Additional attribute groups (second level)

Category	Hierarchy	Description
Collision code	Scene	The category letter of the STATS-19* collision code
Precipitating factor	Vehicle	The main cause of the accident, attributed to the respective occupant
Driver injury	Occupant	OTS injury level of the respective driver
Area	Path	Local area around the location of the accident
Horizontal geometry	Path	Qualitative assessment of curvature of road at locus
Lighting	Path	Light conditions at locus, at the time of the accident
Road type	Path	Type of the road on which the accident occurred
Speed limit	Path	Speed limit posted at the location of the accident
Surface	Path	Road surface condition due to weather at the accident location
Traffic control	Path	Type of traffic control at the location of the accident

** STATS19 is the National UK protocol for information to be collected whenever an injury crash is reported to the Police (see section 2.3.1)*

Table 11 gives all additional second-level attributes including the short name and a description of the attribute. Originally, it was planned to use the collision code (letter

+ number) as the attribute to distinguish between the different sub-categories of collisions (see Appendix B). However, the collision number was found to be erroneous due to a coding problem in the OTS database error. At the time when the study results were compiled, the data error was not eliminated yet, which made it necessary to use the collision letter (supercategory) only. Nevertheless, enough information can be derived from other attributes such as “manoeuvre” and “first point of impact”.

Table 11: Additional collision attributes (second level)

Category	Short name	Description	Count*	Rel. frequ.
Collision code (The category letter of the UK STATS-19 collision code)	Coll=D-Cornering	Cornering (D)	16	1.5%
	Coll=H-CrossingNoTurns	Crossing (no turns) (H)	202	18.9%
	Coll=J-CrossingVehTurning	Crossing (vehicle turning) (J)	236	22.1%
	Coll=M-Manoeuvring	Manoeuvring (M)	104	9.7%
	Coll=Other	Other collision code	11	1.0%
	Coll=A-OvertakingLaneChange	Overtaking and lane change (A)	30	2.8%
	Coll=P-PedestrOther	Pedestrians Other (P)	25	2.3%
	Coll=F-RearEnd	Rear end (F)	188	17.6%
	Coll=L-RightTurnAgainst	Right turn against (L)	204	19.1%
	Coll=G-TurningVsSameDir	Turning versus same direction (G)	54	5.0%
Precipitating factor (The main cause of the crash, attributed to the respective occupant)	Prec=FailAvoidDriver	Driver failed to avoid object or vehicle on carriageway	64	6.0%
	Prec=FailAvoidOther	Other road user failed to avoid object or vehicle on carriageway	58	5.4%
	Prec=FailGiveWayDriver	Driver failed to give way	266	24.9%
	Prec=FailGiveWayOther	Other road user failed to give way	217	20.3%
	Prec=FailStopDriver	Driver failed to stop	84	7.9%
	Prec=FailStopOther	Other road user failed to stop	95	8.9%
	Prec=LossCntrDriver	Driver lost control of vehicle	23	2.1%
	Prec=LossCntrOther	Other road user lost control of vehicle	17	1.6%
	Prec=OtherDriver	Other precipitation by driver	27	2.5%
	Prec=OtherOther	Other precipitation by another road user	29	2.7%
	Prec=PedEnter	Pedestrian entered road without due care (driver not to blame)	17	1.6%
	Prec=PoorOvtkDriver	Inappropriate overtake by driver	7	0.7%
	Prec=PoorOvtkOther	Inappropriate overtake by other driver or rider	23	2.1%
	Prec=PoorMnvrDriver	Inappropriate turn or manoeuvre by driver	80	7.5%
	Prec=PoorMnvrOther	Inappropriate turn or manoeuvre by other driver or rider	63	5.9%
Driver injury (OTS injury level of the respective driver)	DrvInj=Uninjured	Driver suffered no injury	576	53.8%
	DrvInj=Slight	Driver was slightly injured	445	41.6%
	DrvInj=Serious	Driver was seriously injured	42	3.9%
	DrvInj=Fatal	Driver was fatally injured	7	0.7%

Area (around the crash location)	Area=Rural	Rural area (countryside, fields and only sparse housing)	368	34.4%
	Area=Urban	Urban area (at least one side of the road built up)	702	65.6%
Horizontal geometry (Qualitative assessment of curvature of road)	HorizGeom=Left	Left curve	22	2.1%
	HorizGeom=LeftSharp	Left sharp curve	4	0.4%
	HorizGeom=LeftSlight	Left slight curve	51	4.8%
	HorizGeom=Right	Right curve	25	2.3%
	HorizGeom=RightSharp	Right sharp curve	9	0.8%
	HorizGeom=RightSlight	Right slight curve	77	7.2%
	HorizGeom=Straight	Straight (no curve)	882	82.4%
Lighting (Light conditions at the time of the crash)	Light=DarkNSL	Darkness: no street lighting	50	4.7%
	Light=DarkSLUnk	Darkness: street lighting unknown	11	1.0%
	Light=DarkSL	Darkness: street lights lit	188	17.6%
	Light=DayNSL	Daylight: no streetlighting present	571	53.4%
	Light=DaySLUnk	Daylight: streetlighting unknown	243	22.7%
	Light=DaySL	Daylight: streetlights present	7	0.7%
Road type (on which the crash occurred)	RdType=DualCgw	Dual carriageway	161	15.0%
	RdType=OneWayStr	One way street	26	2.4%
	RdType=SingCgw	Single carriageway	883	82.5%
Speed limit (posted at the crash location)	SpdLim<=20mph	20mph and less	1	0.1%
	SpdLim=30mph	30mph	584	54.6%
	SpdLim=40-50mph	40 or 50mph	270	25.2%
	SpdLim=60mph	60mph	159	14.9%
	SpdLim=70mph	70mph	56	5.2%
Surface (Road surface condition due to weather at the crash location)	Surf=Dry	Dry surface	673	62.9%
	Surf=Flood	Flooded surface	9	0.8%
	Surf=Icy	Icy surface	6	0.6%
	Surf=Snowy	Snowy surface	3	0.3%
	Surf=Wet	Wet surface	379	35.4%
Traffic control (Type of traffic control at the location of the crash)	TrfCtrl=None	No active or static yield instruction	582	54.4%
	TrfCtrl=GW	Static give-way instruction	245	22.9%
	TrfCtrl=Stop	Static stop instruction	14	1.3%
	TrfCtrl=Light	Traffic light control	229	21.4%

**The count represents the number of samples on a driver level, i.e. the number of drivers related to the respective attributes.*

Samples with at least one unknown attribute value were removed as part of the data processing steps. This happened at two instances, namely 1) before computing the cluster with level-1 data and 2) before computing the rules with level-2 attributes for the data in each cluster. The first removal of unknowns resulted in a final sample size of $n = 1325$ for clustering, including $n = 930$ for T-junctions, $n = 368$ for crossroads and $n = 27$ for other or no junctions. The frequencies of the attributes are given on the right-hand side in Table 9 and * STATS19 is the National UK protocol for information to be collected whenever an injury crash is reported to the Police (see section 2.3.1)

Table 11 gives all additional second-level attributes including the short name and a description of the attribute. Originally, it was planned to use the collision code (letter + number) as the attribute to distinguish between the different sub-categories of collisions (see Appendix B). However, the collision number was found to be erroneous due to a coding problem in the OTS database error. At the time when the study results were compiled, the data error was not eliminated yet, which made it necessary to use the collision letter (supercategory) only. Nevertheless, enough information can be derived from other attributes such as "manoeuvre" and "first point of impact".

. The second removal of unknowns was done on the extended level-2 dataset. Therefore, the final overall sample size (n=1070) of the dataset used for the association rules is different to the clustering dataset.

5.3 Clustering with *k*-medoids

Having discussed which attributes are used, the following sections address the clustering method chosen, how it was applied to the accident dataset and which results were achieved.

5.3.1 Introduction to *k*-medoids

Partitioning methods organise all samples into k partitions, i.e. clusters. One of the most commonly used partitioning method is k-means (Lloyd, 1982; MacQueen, 1967; Steinhaus, 1956), which treats each data sample as an object having a location in Euclidean space. It finds a partition, in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. This is usually based on a Euclidean distance measure. The spatial centre of each cluster is called "centroid" that is the mean of all objects within the cluster, and that is the point to which the sum of distances from all objects in that cluster is minimized. Those centroids are initially placed in space, randomly in the simplest case, but ideally far from each other. Each sample or object is iteratively assigned to the cluster that has the closest centroid. When all objects have been assigned, the positions of the centroids are recalculated. This process is repeated until the centroids no longer change.

However, k-means has some limitations. For example, with fewer samples, the initial configuration of the centroids will influence the cluster results significantly, not leading to the global optimum (Äyrämö and Kärkkäinen, 2006). The number of k's must be determined beforehand, and due to the distance principle, only circular or spherical shapes are clustered, although some applications would require arbitrary shapes. Also, k-means is sensitive to outliers, and such outliers may have a disproportionate impact on the final cluster configuration (Chawla and Gionis, 2013).

One of the main limitations regarding this study is the fact that the standard k-means algorithm is not applicable to categorical data. Euclidean distance functions and mean values would not make sense for the given OTS dataset. Therefore, a derivate of k-means is used, called k-medoids. The method uses objects called medoids instead of centroids. Instead of using the mean as the centre of the cluster, a member of the cluster is chosen as the centre, whose average dissimilarity to all the objects in the cluster is minimal. In other words, the medoid is the most centrally located point in the cluster. Thus it is more robust to outliers, because it does not minimize a sum of squared Euclidean distances, as k-means does. Furthermore, k-medoids allows clustering categorical data, where a mean is impossible to define. For this reason, alternative dissimilarity measures can be applied, such as the "Hamming distance" (Hamming, 1950; Wegner, 1960) or the "Jaccard coefficient" (Jaccard, 1901). See Section 5.3.2 for details on the dissimilarity measure used for this study.

One of the most powerful and commonly used algorithm for k-medoids is PAM (Partitioning Around Medoids) proposed by Kaufman and Rousseeuw (1990). It proceeds in two steps as follows:

Build step:

(1) Choose k objects to become the medoids, or in case these objects were provided use them as the medoids
(2) Calculate the dissimilarity matrix if it was not given
(3) Assign every object to its closest medoid

Swap step:

(4) Within each cluster, each object is tested as a potential medoid by checking if the sum of within-cluster distances gets smaller using that object as the medoid. If so, the object is defined as a new medoid.
(5) If at least one medoid has changed, go to (3), else end the algorithm.

Whatever distance or dissimilarity measure is used, the absolute-error criterion of k-medoids can be defined as

$$E = \sum_{i=1}^{k} \sum_{p \epsilon C_i} dist(p, o_i) \tag{5}$$

where E is the sum of the absolute error for all objects p in the dataset, k is the number of clusters and o_i is the representative object (medoid) of the cluster C_i. The PAM algorithm works effectively for small data sets such as the underlying OTS dataset. For larger datasets, alternative k-medoids algorithms should be used, such as CLARA (Clustering Large Applications, Kaufman and Rousseeuw, 1990). The k-medoids clustering was implemented in MATLAB by using the in-built library.

5.3.2 Clustering parameters

In MATLAB, the following parameters were chosen for the k-medoids implementation:

- Algorithm: "PAM"
- Dissimilarity measure: "Hamming"
- Number of maximum k: 15
- Number of replicates: 5
- Initialization method: "Plus"
- Validity index: Silhouette value

The PAM algorithm (see Section 5.3.1) was chosen, because it is most appropriate for the given sample size. The algorithm can produce better solutions than other k-medoids algorithms in some situations, but the computation times can be longer.

MATLAB provides many different distance measures, among which two are applicable to categorical, binary-coded data, namely Hamming and Jaccard. The Hamming distance, originally used for the detection of errors in information transmission (Hamming, 1950), was chosen as the distance measure, because it treats 0's and 1's equally, i.e. it does not prefer 1's over 0's. In contrast to that, the Jaccard distance only counts the differences of non-zero elements. The Hamming distance gives the number of mismatches and is calculated by

$$d_{Hamming}(p, o) = \sum_{i=0}^{n-1} [y_{p,i} \neq y_{o,i}] \tag{6}$$

with p and o being the two objects compared, and i as the index of the respective attribute value y out of the total number of variables n.

A crucial issue of k-medoids is the initial selection of the medoids, i.e. the starting configuration of the PAM algorithm. An easy way of doing this is to randomly select them, but it was preferred to use the k-means algorithm for cluster centre initialization (called "plus"). However, a poor selection of medoid positions would result in poor partitioning. Therefore, MATLAB provides the option to define a certain number of replicates, i.e. the number of times to repeat clustering using new initial medoid positions. The best replicate is then chosen for the final output. For this study, a number of 5 replicates was defined.

The best number of clusters k was achieved by iteratively stepping from $k_{min} = 2$ to $k_{max} = 15$ clusters. Experiments with the dataset showed that a $k_{max} > 15$ does not result in any more change of the error function, as the curve flattens. The results from each k were compared to find the best k, i.e. the one with the lowest average silhouette value (see Section 5.3.3). Actually, finding the best k is one of the most debated problems in the clustering community. In literature, various validity metrics can be found to compute a cluster's performance in partitioning, among which are the Akaike's Information Criterion (AIC, Akaike 1974), the Bayesian Information

Criterion (BIC, Schwarz, 1978), Calinski-Harabasz (Calinski and Harabasz, 1974) or Davies Bouldin index (Davies and Bouldin, 1979). For the scope of this study, the Silhouette value was used, developed by Rousseeuw (1987) as a graphical display for validating clusters. The so-called silhouette plot was chosen for this study, because it provides a simple visualization to decide on the number of k. The entire clustering is displayed by combining the silhouettes into a single plot, as seen in the example in Figure 45. The plot consists of the silhouette values $s(i)$, ranked in decreasing order for all objects i within a cluster. Considering $a(i)$ as the average dissimilarity of an object i to all other objects within its own cluster A, and $b(i) = \min d(i, B)$ as the minimum average dissimilarity of i to all objects of a neighbouring cluster B, then $s(i)$ is calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i),\ b(i)\}} \tag{7}$$

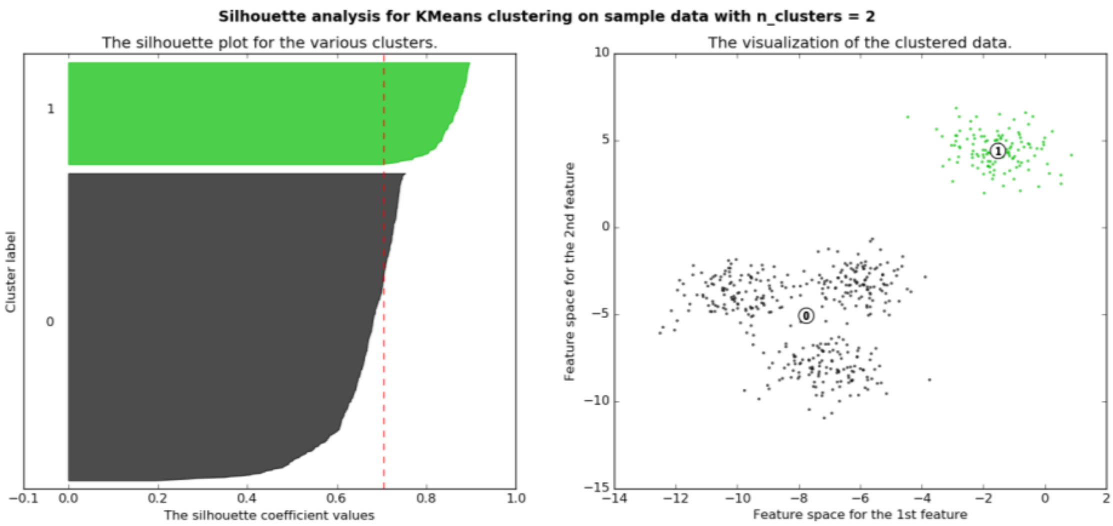


Figure 45: Example of a silhouette plot (left) compared to the corresponding feature space (right) (scikit, n.d.)

Ultimately, $s(i)$ measures how well object i has been allocated to the right cluster. The range of $s(i)$ reaches from -1 to +1, where a value close to 1 means that i has been perfectly classified, and a negative value indicates that i might have been falsely allocated. The vertical width of the silhouette represents the cluster size. For evaluating the best k, the average silhouette value of all objects is calculated and compared to the others. In the following subsection, the clustering results are presented and discussed.

5.3.3 Clustering results

The crash dataset was divided into the two main junction types: (1) Three-legged T-junctions and (2) four-legged crossroads. For other types of junctions (e.g. private drives, pedestrian crossings), the sample size was too small ($n = 27$) to compute clusters. This partitioning prior to clustering was done due to the scope of the study, namely to provide targeted scenarios and parameter variations for virtual vehicle simulations. The goal was not to find clusters characterized by junction types, but by driving situations, manoeuvres and injury outcome (see level-1 attributes). Furthermore, the number of intersection legs was found to be a significant variable to model intersection crashes (Abdel-Aty et al., 2006) and was used to group intersection crashes in various studies (e.g. Abdel-Aty and Haleem, 2011; Arndt, 2003; Persaud and Nguyen, 1998; Vogt and Bared, 1998).

Before interpreting the computed clusters, the results had to be processed as follows: For both the T-junction and the crossroads dataset, a table was created giving the frequency of each attribute per cluster (see Table 12 and Table 13). A χ^2-goodness-of-fit test (Pearson, 1900) was applied to the data, by testing the 99.5% significance of each distribution of attributes within a cluster (e.g. maximum injury = Uninjured, Slight, SeriousFatal) compared to the overall (expected) distribution of this attribute group. The null hypothesis, i.e. the statement that there is no relationship between the two distributions, was confirmed or rejected according to the p-value. To get valid results, only the attribute groups where all expected frequency values are greater or equal to 5 were tested. By doing so, it could be determined whether a cluster is significantly different to the total distribution of attributes. If the null hypothesis is rejected for an attribute group, and if an attribute within this group is higher than the expected value (the frequency of this attribute in the overall population), then it is highlighted as over-represented.

5.3.3.1 Clusters found for three-legged junctions

The silhouette plot in Figure 46 (left) shows the average silhouette values (cluster validity) for all k's. In general, the higher the number of clusters the higher the silhouette values get. A higher number of clusters might lead to over-fitting and a lower number of clusters to under-fitting. To find the best k, a compromise between cluster size and cluster validity had to be found. Association rules, which are computed for each cluster in the next step, were originally made for large-scale data. Hence, the goal was to avoid very small clusters, i.e. results with clusters containing less than 30 samples are disregarded ($k = 14$ and $k = 15$). Since $k = 13$ has the highest average silhouette value with 0.383, the lowest number of samples that were allocated to the wrong cluster, and overall, the lowest percentage of clusters with negative silhouette values, it was chosen as most valid k.

Figure 46 (right) depicts the silhouette plot for each of the thirteen clusters, with one horizontal bar per sample within the cluster. Samples with a negative silhouette value might be assigned to the wrong cluster. However, the number of those samples is considerably low, except for cluster 4, where the average silhouette value suffers compared to the other clusters. Cluster 4 must therefore be treated carefully when interpreting the results.

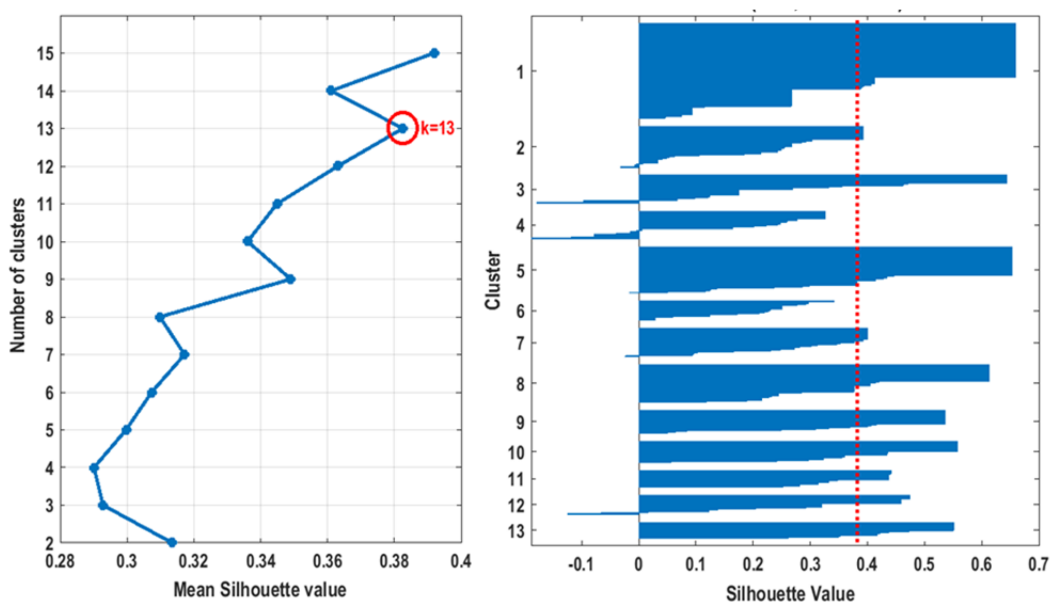


Figure 46: Mean silhouette values for all *k*'s (left) and silhouette plot for *k*=13 (right) for T-junctions

Table 12: Cluster results for T-junctions (*k* = 13, *n* = 930)

Level-1 Attributes	T-C1	T-C2	T-C3	T-C4	T-C5	T-C6	T-C7	T-C8	T-C9	T-C10	T-C11	T-C12	T-C13
Sample size	**212**	**90**	**62**	**62**	**102**	**43**	**63**	**83**	**52**	**46**	**38**	**42**	**35**
MaxInj=Uninjured	0	11	7	8	15	9	63	7	0	0	4	2	5
MaxInj=Slight	212	69	52	42	78	30	0	68	45	0	29	0	25
MaxInj=SeriousFatal	0	10	3	12	9	4	0	8	7	46	5	40	5
JctShp=T-minLeft	195	0	1	0	0	0	58	3	51	41	0	1	0
JctShp=T-minRight	0	0	58	62	102	0	0	0	0	0	38	14	35
JctShp=T-termMaj	17	90	3	0	0	43	5	80	1	5	0	27	0
1stIntAct=Car	183	60	53	40	81	24	53	60	37	33	27	0	23
1stIntAct=LGV-HGV	18	6	5	4	5	2	7	2	5	5	4	3	2
1stIntAct=PTW	3	10	3	10	6	4	0	14	3	2	3	35	6
1stIntAct=Other	4	4	1	3	4	2	2	2	4	2	1	1	0
1stIntAct=Cycle	1	8	0	5	2	10	0	5	2	1	1	2	1
1stIntAct=Pedestrian	3	2	0	0	4	1	1	0	1	3	2	1	3
Manvr=GoingAheadOther	201	0	17	6	97	0	50	7	45	43	27	1	0
Manvr=Other	8	11	1	4	5	0	4	2	2	3	11	2	0
Manvr=TurnL	2	0	0	1	0	43	7	0	4	0	0	2	0
Manvr=TurnR	1	75	5	51	0	0	1	69	0	0	0	36	35
Manvr=WaitTurnR	0	4	39	0	0	0	1	5	1	0	0	1	0
1stImpact=Back	25	9	55	0	0	5	10	0	0	4	0	1	0
1stImpact=Front	162	68	1	0	102	18	35	0	0	37	0	11	35
1stImpact=Nearside	0	13	1	48	0	6	10	0	52	0	0	1	0
1stImpact=Offside	25	0	5	14	0	14	8	83	0	5	38	29	0

The frequencies of each attribute within each cluster were compiled in a table to present the results at a glance (see Table 12). Cells shaded in grey indicate that the distribution of numbers for the given field is significantly different from the distribution in the whole population (χ^2-test with significance $\alpha = 0.05$) and that the particular number highlighted is over-represented. Due to values lower than 5 in the expected frequency table, the χ^2-test could not be applied to all observations. For example, Cluster 2 contains 90 samples and all of these occurred at "JctShp=T-termMaj" compared to a distribution of 350 "JctShp=T-minLeft", 309 "JctShp=T-minRight" and 271 "JctShp=T-termMaj" in the overall 930 population. The probability that this would happen by chance is less than 0.05% and the number 90 is over-represented, as it is higher than the expected value.

In what follows, the respective clusters are explained in detail:

Cluster T-C1: "The car hits another car or LGV with its front, while going straight on a road with a minor roads joining from the left."

Cluster T-C1 is the largest cluster with a size of 212 crashes, from which all resulted in slight injury. More than 90 percent of the accidents occurred at T-junctions with a minor road joining from the left. „1stImpact=Front" and „1stImpact=Back" are over-represented as well as „Manvr=GoingAheadOther". There is no clear indication on the collision type of this cluster, thus association rules are used for further analyses.

Cluster T-C2: "The car hits another car or PTW with its front, while turning right into a major road."

The third largest *cluster T-C2* clearly groups collisions while turning right, with a highly significant representativeness of frontal and nearside impacts, all of which occurring at roads terminated by a major road. Powered two-wheelers (PTW) and bicyclists have relatively high frequencies, but the car is still the dominant crash partner.

Cluster T-C3: "The car is hit on its back, while waiting to turn right into a minor road."

Cluster T-C3 with 62 samples represents car-to-car collisions at roads with minor roads joining from the right, mainly resulting in slight injury. Since there are mainly impacts on the back of the car, this cluster can be seen as rear-end crash group.

Cluster T-C4: "The car is hit on the nearside, while turning right into a minor road."

Cluster T-C4 occurred on a road with a minor road joining from the right, with nearside impacts in 77 percent of the cases and high frequencies for „Manvr=TurnR" and „1stIntAct=Car".

Cluster T-C5: "The car hits another car with its front, while going straight over a T-junction with a minor roads joining from the right."

The second largest *cluster T-C5* indicates rectangular collisions with another car crossing the car's trajectory from the right, although this assumption will be validated by association rule mining.

Cluster T-C6: "The car collides with another road user, while turning left into a major road."

Cluster T-C6 is characterized by a left turn into a major road, which results in a collision mainly with another car. This cluster has a relatively high number of bicycle crashes (10).

Cluster T-C7: "The car collides with another car, while going straight over a T-junction with a minor road joining from the left."

All 63 accidents in *cluster T-C7* resulted in no injury for any of the participants. This is clearly a minor risk cluster mainly with cars and goods vehicles involved, with „Manvr=GoingAheadOther“ having a high frequency.

Cluster T-C8: "The car is hit by another car or PTW on its offside, while turning right into a major road."

Cluster T-C8 represents slight injury collisions with mainly other cars or PTW. Offside impacts were found over-represented, while turning right into a major road.

Cluster T-C9: "The car is hit by another car on its nearside, while going straight over a T-junction with a minor road joining from the left."

Cluster T-C9 involves nearside collisions only, which happened on a T junction with a minor road joining from the left, while the car was going straight. 45 of the 52 cases resulted in slight injury.

Cluster T-C10: "The car hits another car with its front, while going straight over a T-junction with a minor road joining from the left."

Cluster T-C10 represents a group of high-risk collisions with serious or fatal injuries in all 46 cases. Front impacts are over-represented and „Manvr=GoingAheadOther“ and „1stIntAct=Car“ have high frequencies. Association rules will be used to analyse this cluster in more detail.

Cluster T-C11: "The car is hit by another car on its offside, while going straight over a T-junction with a minor road joining from the right."

In comparison to T-C9, *cluster T-C11* involves offside collisions only, which happened on a T junction with a minor road joining from the right, while the car was going straight or made another manoeuvre. Five of the 38 cases resulted in serious or fatal injury.

Cluster T-C12: "The car collides with a PTW, while turning right into minor or major road."

Cluster T-C12 is a PTW cluster, with 40 out of 42 collisions resulting in serious or fatal injury. In 85 percent of the cases, the car was turning right. Association rules will be used to analyse this cluster in more detail.

Cluster T-C13: "The car hits another car or PTW with its front, while turning right *into a minor road."*

The smallest *cluster T-C13* is characterised by right-turns into a minor road, with „1stImpact=Front" in all cases. Five of the 35 cases resulted in serious or fatal injury, which is most likely due to the six cases involving PTW. Association rules will be used to analyse this cluster in more detail.

5.3.3.2 Clusters found for four-legged junctions

For the crossroads dataset with 368 samples, $k = 6$ was found to be most valid for separating the clusters, because it has a high average mean silhouette value of 0.395. The silhouette plot in Figure 47 (left) shows the average silhouette values for all k's. Although larger values were computed for higher k's (10-15), they were disregarded due to their small cluster sizes (<30) and possible overfitting.

Figure 47 (right) depicts the silhouette plot for each of the six clusters, with one horizontal bar per sample within the cluster. The total mean silhouette value is higher and the number of samples with a negative value is lower compared to the T-junction dataset. This means that for the attributes and for the k chosen, the four-legged junction dataset seems to be better separated.

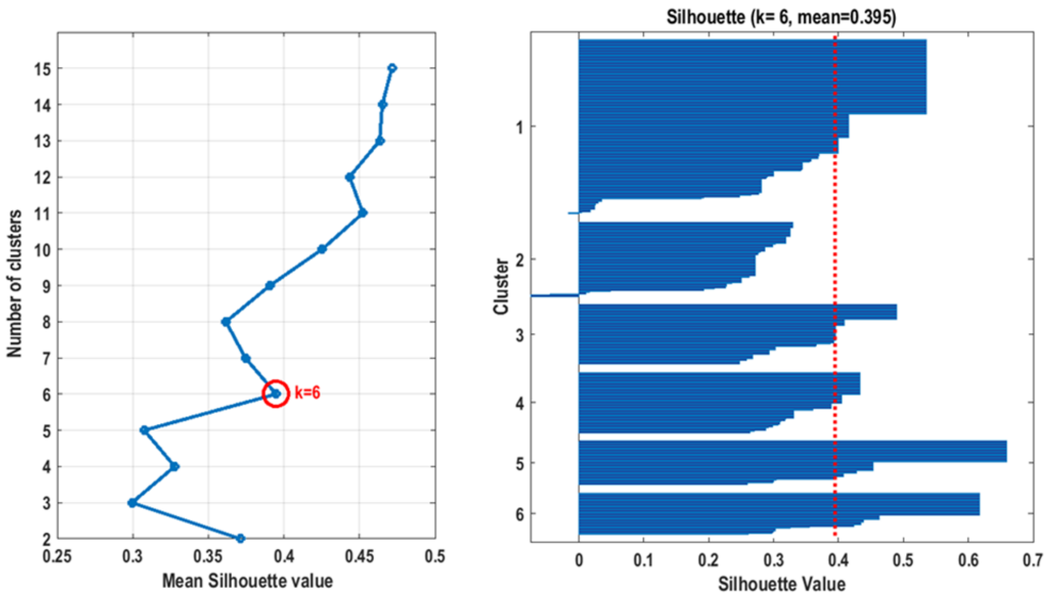


Figure 47: Mean silhouette values for all *k*'s (left) and silhouette plot for *k*=6 (right) for four-legged junction clusters

As for the T-junction dataset, the frequencies of each attribute within each cluster were compiled in a table to present the results at a glance (see Table 13). Cells shaded in grey indicate that the distribution of numbers for the given field is significantly different from the distribution in the whole population (χ^2-test with significance $\alpha = 0.05$) and that the particular number highlighted is over-represented.

Table 13 shows that the four-legged junction dataset is mainly separated by the type of junction and first point of impact. Experiments with varying parameters, such as initial medoid configuration or including the missing values did not result in different partitions. Including more attribute groups resulted in a decrease of the average silhouette value. For all clusters, the χ^2-test was not applied to the attribute groups "1stIntAct" and "Manvr" due to expected frequency values lower than 5. For the attribute group "1stImpact", only cluster X-C1 had sufficient frequency values for a χ^2-test. The distributions for injury level ("MaxInj") do not significantly differ in any cluster from the total population in their attribute group.

Table 13: Cluster results for Crossroads ($k = 6$, $n = 368$)

Level-1 Attribute	X-C1	X-C2	X-C3	X-C4	X-C5	X-C6
Sample size	**142**	**60**	**48**	**49**	**35**	**34**
MaxInj=Uninjured	22	13	8	10	4	4
MaxInj=Slight	98	39	35	29	28	24
MaxInj=SeriousFatal	22	8	5	10	3	6
JctShp=X-minJoin	142	0	48	0	0	34
JctShp=X-brkMaj	0	60	0	49	35	0
1stIntAct=Car	118	44	38	39	30	28
1stIntAct=LGV-HGV	9	4	4	6	2	4
1stIntAct=PTW	3	7	1	3	1	0
1stIntAct=Other	3	1	1	0	0	2
1stIntAct=Cycle	2	2	4	1	1	0
1stIntAct=Pedestrian	7	2	0	0	1	0
Manvr=GoingAheadOther	116	32	25	35	25	29
Manvr=Other	4	0	0	0	0	1
Manvr=TurnL	5	9	2	2	1	1
Manvr=TurnR	15	19	21	12	9	3
Manvr=WaitTurnR	2	0	0	0	0	0
1stImpact=Back	12	5	0	0	0	0
1stImpact=Front	130	55	0	0	0	0
1stImpact=Nearside	0	0	48	0	35	0
1stImpact=Offside	0	0	0	49	0	34

The clusters for crossroads can be further described as follows:

Cluster X-C1: "The car hits another road user with its front, while going straight over a crossroad with minor roads joining left and right."

Cluster X-C1 is the largest cluster with 142 samples, which mainly include rear-end collisions, as the clusters X-C3 to X-C6 have no samples for „1stImpact=Back" and cluster 2 has only 5.

Cluster X-C2: "The car hits another road user with its front, while crossing or turning into a major road."

Cluster X-C2 groups situations on crossroads broken by a major road, with high numbers for turning left or right as well as „1stImpact=Front". Cars and PTWs were mostly involved.

Cluster X-C3: "The car is turning right into a minor road, when being hit on its offside by another vehicle."

All situations in *Cluster X-C3* occurred on a road with minor roads joining from the left and right, and in all situations, the car was hit on its nearside.

Cluster X-C4: "The car is crossing or turning into a major road, when being hit on its offside by another vehicle."

All situations in *Cluster X-C4* occurred on a road broken by a major road passing the car's path, and in all situations, the car was hit on its offside.

Cluster X-C5: "The car is going straight on a road broken by a major road, when being hit on its nearside by another vehicle."

All situations in *Cluster X-C5* occurred on a road broken by a major road passing the car's path, and in all situations, the car was hit on its nearside, mainly by another car. As for the previous clusters, there is no statistical significance given for the manoeuvre, interaction or injury level distribution.

Cluster X-C6: "The car is going straight over a junction with minor roads joining from the left and right, when being hit on its offside by another car or goods vehicle."

The smallest *Cluster X-C6* represents collisions on roads with minor roads joining from left and right, where the car was hit on its offside, while going straight over the junction.

5.4 Association rule mining

As illustrated in the methodology flowchart in Figure 20, the obtained clusters are further analysed by association rule mining. This section gives an overview on the principle of association rules and describes how the rules help to derive scenario parameters for the simulation.

5.4.1 Introduction to association rules

The k-medoids partitioning resulted in 13 and 6 distinct clusters, for T-junctions and crossroads respectively. For each of the clusters, it is now interesting to identify the attributes that often occur together, in order to understand the clusters. Association rule mining is a method to discover associations between attributes, also called "frequent itemset mining". A popular example of association rules is the market basket analysis. Retailers can get insights into which items are frequently purchased together, so that marketing strategies and product shelving can be optimized. For example, if a customer buys "beer", then he/she also buys "crisps". This would be expressed as "beer $\rightarrow$ crisps", where the item "beer" is called the antecedent and the item "crisps" the consequent. One itemset I can contain multiple items. Applying the association rules terminology to the OTS dataset, then each sample is called a transaction $\{t_1, t_2, \dots, t_n\} \in T$, and each attribute is an item $\{i_1, i_2, \dots, i_m\} \in I$. An association rule can be written in the following mathematical form: $X \rightarrow Y$ where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. Each rule is characterised by its support and its confidence:

$$supp(X) = \frac{|\{t \in T; X \subseteq t\}|}{n} = P(X) \tag{8}$$

$$conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} = P(Y|X) \tag{9}$$

For itemsets, the *support* value gives the proportion of transactions t in the dataset, which contains the itemset X. For rules, the support is defined as the support of all items in the rule, i.e. $supp(X \rightarrow \mathrm{Y}) = supp(X \cup Y) = P(X \wedge Y)$.

Equivalently, the *confidence* measures the strength of the rules and gives the conditional probability of the consequent Y given the antecendent X. In other words, it is the proportion of the transactions that contain X, which also contain Y. To explain the difference between the two measures, it is important to mention that two rules with flipped antecedent and consequent would both have the same support value. However, they would not have the same confidence, because the direction is taken into account.

The most common implementation of association rules was proposed by Agrawal et al. (1993), who called their method the Apriori algorithm. Accordingly, finding association rules involves two steps:

1. Find all frequent itemsets
2. Generate association rules from the frequent itemsets

The algorithm necessitates two parameters, namely a minimum support threshold, and a minimum confidence. By definition, if an itemset is below the minimum support threshold, then it is not frequent. If so, all its subsets must also be infrequent and can be pruned. In contrary, any subset of a frequent itemset must be frequent. By following this principle iteratively, the number of possible itemset configurations can be reduced tremendously with a simple algorithm.

The second step is to generate rules from the frequent itemsets found in step 1. Here, the minimum confidence threshold comes into play: For each frequent itemset I, all nonempty subsets are generated. For every nonempty subset s of I, create the rule $s \rightarrow (I - s)$ if the minimum confidence for this rule is given. Since the rules are generated from frequent itemsets, each one also satisfies the minimum support. In this way, strong association rules can be found.

Depending on the data dimensionality, and on how low the minimum support and confidence thresholds have been set, the algorithm might produce millions of rules. In literature, numerous rule pruning and post-processing methods to identify the rules of most interest were published. It was found that the confidence measure is a rather poor measure to discover the dependence of the consequent with respect to the antecedent (Guillaume et al., 1998; Silverstein et al., 1998). In this study, a metric called lift, also known as "interestingness" is chosen, which is calculated as follows:

$$lift(X \rightarrow Y) = lift(Y \rightarrow X) = \frac{supp(X \cup Y)}{supp(X) \cdot supp(Y)} = \frac{P(X \wedge Y)}{P(X) \cdot P(Y)} \tag{10}$$

If the lift value is less than 1, then the occurrence of X is negatively correlated with the occurrence of Y, meaning that the occurrence of one likely leads to the absence of the other one. If the resulting value is greater than 1, then X and Y are positively correlated, meaning that the occurrence of one implies the occurrence of the other. If the lift equals 1, then X and Y are independent (Han et al., 2011). By setting an appropriate minimum lift value greater than 1, only high-lift rules can be extracted for interpretation.

As a popular and simple data mining technique, various accident researchers used this approach to discover patterns in their data (Kumar and Toshniwal, 2015; Mirabadi and Sharifian, 2010; Montella, 2011; Pande and Abdel-Aty, 2009; Weng et al., 2016). In this study, association rules are applied to the discovered clusters to get more information on the underlying patterns of accident attributes. Association rule mining was implemented in R by using the "arules" package (Hahsler et al., 2017, 2005).

5.4.2 Parameters for the algorithm

In MATLAB, the following parameters were chosen for the association rules implementation:

- Algorithm: "Apriori"
- Minimum support: 0.03
- Minimum confidence: 0.75
- Minimum lift: 1.25

The choice of the minimum support and confidence depends on the application and the expected outcome of the study. In theory, it is desirable to obtain rules with high support, high confidence and a lift value much greater than 1. The idea of this study implies the analysis of certain accident situations and characteristics, which can be very rare (Montella et al., 2012). After experimenting with different values, a minimum support of 0.03 was chosen, so that all itemsets occurring in less than 3 percent of the samples are disregarded. Choosing a lower threshold results in an increase of computation time and rules, which would all have to be interpreted. Choosing a higher support value might disregard relevant information about the clusters.

There are different approaches in literature on the choice of a minimum confidence value. For example, Montella (2011) chose a threshold with $conf = 0.1$ for their powered-two-wheeler study, which is much lower than usual. However, in this study it is preferred to obtain rules, where the probability of the consequent given the antecedent is higher than 75 percent. Additionally, only rules with a $lift > 1.25$ are considered for the results.

To further reduce the number of rules obtained, redundant rules were excluded according to the following procedure: A rule is redundant if a more general rule with the same or a higher lift exists. That is, a more specific rule is redundant if it is only equally or even less correlated than a more general rule. A rule is more general if it has the same consequent but one or more items removed from the antecedents. Formally, a rule $X \rightarrow Y$ is redundant if for $X' \subset X: lift(X' \rightarrow Y) \geq lift(X \rightarrow Y)$ (Hahsler et al., 2017).

5.4.3 Results and derived scenarios

For each identified cluster, association rules were computed using the parameters given in Section 5.4.2. In total, the analysis of each cluster resulted in 42 different crash scenarios comprising various parameters, which are all given in Appendix D (for T-junctions) and Appendix E (for four-legged junctions). This means that some clusters have two or more subgroups of scenarios, because the analysis revealed differences in the crash circumstances, even within the clusters. The next sections will explain how those scenarios were derived and give a description of each scenario, including a simplified illustration of the crash. Simplified means that not all circumstances could be visualised in the sketch, such as road types (minor/major), number of lanes, horizontal alignment, speed limits, vehicle types or injury levels. The red dots in the collision sketches are the points of impact (i.e. front, back, offside or nearside).

5.4.3.1 High-injury scenarios

For each identified cluster, association rules were computed using the parameters given in Section 5.4.2. In total, the analysis of each cluster resulted in 42 different crash scenarios comprising various parameters. Only high-risk scenarios, which resulted in serious or fatal injury, are presented in this section, as they provide a set of safety-critical situations. More precisely, the further scenarios include crash situations from the T-junction clusters T-C4, T-C10, T-C12 and T-C13, and from the four-legged junction clusters X-C1, X-C2, X-C4 and X-C6.

As an example, Cluster T-C10 is selected for further explanation. The following procedure was applied to every cluster in the same manner. All association rules obtained for each cluster are given in Appendix F. Given the distributions in Table 12, the cluster T-C10 can be described as follows: “The car hits another car with its front resulting in serious or fatal injury, while going straight on a road with a minor road joining from the left.”

A useful attribute to give a clearer indication about the crash circumstances is the collision type, indicated by letters A to Q in the OTS data specification (see Appendix B). For cluster T-C10, the collision types L („Right Turn Against“) and J

(„Crossing with Vehicle Turning“) were found to be the most frequent. Therefore, all rules containing those attributes within their items were further analysed to see which other attributes are associated with them.

Table 14 gives the 2-item and 3-item rules for T-C10 and collision type L, sorted by the five highest support values. The rules are sorted by the support to obtain the attributes that are often associated with each other. It can be seen that this collision type is associated with single carriageways (rule nr. 1) as well as with no traffic control ("TrfCtrl=None", see rule nr. 2, 4 and 11) and going straight ("Manvr =GoingAheadOther", see rule nr. 3). Another car as collision partner is already defined by the cluster, but the rules reveal that "Coll_L=RightTurnAgainst" and "FirstIntAct=Car" are associated with dry surface (see rule nr. 5), uninjured driver of the ego car (see rule nr. 10), a fail to give way by the other car driver (see rule nr. 12), daylight (see rule nr. 13), 40-50 mph speed limit (see rule nr. 9) and urban area (see rule nr. 22).

Table 14: 2-item and 3-item rules obtained for cluster T-C10 with collision type L, sorted by the five highest support values

Nr	Antecedent	Consequent	Supp	Conf	Lift
1	Coll_L=RightTurnAgainst	RdType=SingCgw	0.237	0.818	1.413
2	Coll_L=RightTurnAgainst & TrfCtrl=None	RdType=SingCgw	0.237	1.000	1.727
3	Coll_L=RightTurnAgainst & Manvr=GoingAheadOther	RdType=SingCgw	0.237	0.900	1.555
4	Coll_L=RightTurnAgainst & RdType=SingCgw	TrfCtrl=None	0.237	1.000	1.357
5	Coll_L=RightTurnAgainst & Surf=Dry	FirstIntAct=Car	0.184	1.000	1.357
6	Coll_L=RightTurnAgainst & Surf=Dry	RdType=SingCgw	0.158	0.857	1.481
7	Coll_L=RightTurnAgainst & Area=Rural	RdType=SingCgw	0.158	0.857	1.481
8	Coll_L=RightTurnAgainst & DrvInj=Uninjured	RdType=SingCgw	0.132	1.000	1.727
9	Coll_L=RightTurnAgainst & SpdLim=40=50mph	FirstIntAct=Car	0.132	1.000	1.357
10	Coll_L=RightTurnAgainst & DrvInj=Uninjured	FirstIntAct=Car	0.132	1.000	1.357
11	Coll_L=RightTurnAgainst & DrvInj=Uninjured	TrfCtrl=None	0.132	1.000	1.357
12	Coll_L=RightTurnAgainst & Prec=FailGiveWayOther	FirstIntAct=Car	0.132	1.000	1.357
13	Coll_L=RightTurnAgainst & Light=DayNSL	FirstIntAct=Car	0.132	1.000	1.357
14	Coll_L=RightTurnAgainst & Light=DaySLUnk	RdType=SingCgw	0.105	1.000	1.727
15	Coll_L=RightTurnAgainst & SpdLim=40=50mph	Surf=Dry	0.105	0.800	1.448
16	Coll_L=RightTurnAgainst & DrvInj=Uninjured	Surf=Dry	0.105	0.800	1.448
17	Coll_L=RightTurnAgainst & Prec=FailGiveWayOther	Surf=Dry	0.105	0.800	1.448
18	Coll_L=RightTurnAgainst & Light=DayNSL	Surf=Dry	0.105	0.800	1.448
19	Coll_L=RightTurnAgainst & Light=DaySLUnk	TrfCtrl=None	0.105	1.000	1.357
20	Coll_L=RightTurnAgainst & Area=Urban	FirstIntAct=Car	0.105	1.000	1.357
21	Coll_L=RightTurnAgainst & Light=DaySLUnk	HorizGeom=Straight	0.105	1.000	1.267
22	Coll_L=RightTurnAgainst & Area=Urban	HorizGeom=Straight	0.105	1.000	1.267
23	Coll_L=RightTurnAgainst & Surf=Wet	HorizGeom=Straight	0.105	1.000	1.267

Table 15 gives the 2-item and 3-item rules for the cluster T-C10 and collision type J, sorted by the five highest support values. It can be seen that this collision type is associated with a failure to give way by the other driver (see rule nr. 1). This combination is further associated with another car as collision partner (see rule nr. 5), no traffic control (see rule nr. 6), wet surface (see rule nr. 10), single carriageway (see rule nr. 11), rural area (see rule nr. 12), serious driver injury (see rule nr. 20) and 40-50 mph speed limit (see rule nr. 35). Taking a deeper look into the serious driver injuries, it can be noted that they are further associated with 40-50 mph speed limit (see rules nr. 27 and 28), wet surface (rule nr. 29) and single carriageway (rule nr. 30). However, this set of rules show that there is no clear indication on some attributes, such as the road type, as "RdType=DualCgw" is among the frequent items (see rules 42 to 45). Also, the driver can be uninjured or seriously injured or the area can be urban or rural. Those varying attributes could be used as varying parameter in the simulation, while the others constitute the "static" environment and situation.

While the rules in the tables are relatively easy to interpret, this is no more the case with 4-, 5- or 6-item rules, also due to the high number of obtained rules. Therefore, each set of rules (comprising 2- to 6-item rules) was further visualised by directed graphs that were created from adjacency matrices of the associations found between all attributes. The graph was then reduced to the edges that direct to a certain consequent, represented by edge tables including source, target and weight of the edges. In this case, the targets (or consequents) were the collision types L (see Figure 48) and J (see Figure 49), and the sources were all remaining attributes. The weight or thickness of each edge represents the number of associations identified between the respective antecedent node and the given consequent in the centre. In other words, nodes with thick edges indicate dominant crash attributes and thus define the scenario. For antecedent nodes that are not present in the graph, there were no associations found in the rules. Thus they can be considered negligible for the respective scenario. Note that the graph does not reflect support, confidence or lift.

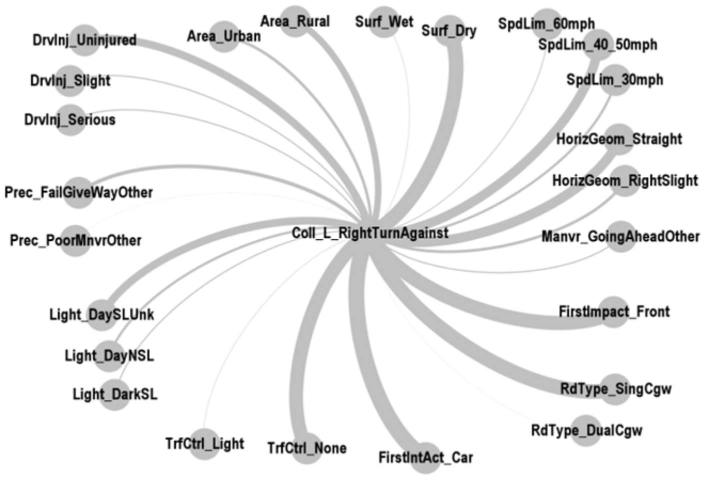


Figure 48: Weighted, directed graph obtained from all association rules for cluster T-C10 having collision type L as consequent

Table 15: 2-item and 3-item rules obtained for cluster T-C10 with collision type J, sorted by the five highest support values

Nr	Antecedent	Consequent	Supp	Conf	Lift
1	Coll=J	Prec=FailGiveWayOther	0.211	0.800	2.338
2	Coll=J	Light=DayNSL	0.211	0.800	1.520
3	Coll=J & Light=DayNSL	HorizGeom=Straight	0.211	1.000	1.267
4	Coll=J & HorizGeom=Straight	Light=DayNSL	0.211	1.000	1.900
5	Coll=J & FirstIntAct=Car	Prec=FailGiveWayOther	0.184	0.875	2.558
6	Coll=J & TrfCtrl=None	Prec=FailGiveWayOther	0.184	0.875	2.558
7	Coll=J & Area=Rural	Light=DayNSL	0.184	1.000	1.900
8	Coll=J & Area=Rural	HorizGeom=Straight	0.184	1.000	1.267
9	Coll=J & Prec=FailGiveWayOther	Surf=Wet	0.158	0.750	1.781
10	Coll=J & Surf=Wet	Prec=FailGiveWayOther	0.158	1.000	2.923
11	Coll=J & RdType=SingCgw	Prec=FailGiveWayOther	0.158	0.857	2.505
12	Coll=J & Area=Rural	Prec=FailGiveWayOther	0.158	0.857	2.505
13	Coll=J & Surf=Wet	Light=DayNSL	0.158	1.000	1.900
14	Coll=J & Light=DayNSL	Surf=Wet	0.158	0.750	1.781
15	Coll=J & Surf=Wet	Area=Rural	0.158	1.000	1.407
16	Coll=J & Area=Rural	Surf=Wet	0.158	0.857	2.036
17	Coll=J & Surf=Wet	HorizGeom=Straight	0.158	1.000	1.267
18	Coll=J & HorizGeom=Straight	Surf=Wet	0.158	0.750	1.781
19	Coll=J & TrfCtrl=None	RdType=SingCgw	0.158	0.750	1.295
20	Coll=J & DrvInj=Serious	Prec=FailGiveWayOther	0.105	1.000	2.923
21	Coll=J & SpdLim=30mph	Area=Urban	0.079	1.000	3.455
22	Coll=J & Area=Urban	SpdLim=30mph	0.079	1.000	3.800
23	Coll=J & SpdLim=30mph	Surf=Dry	0.079	1.000	1.810
24	Coll=J & Surf=Dry	SpdLim=30mph	0.079	0.750	2.850
25	Coll=J & SpdLim=30mph	RdType=SingCgw	0.079	1.000	1.727
26	Coll=J & SpdLim=30mph	TrfCtrl=None	0.079	1.000	1.357
27	Coll=J & DrvInj=Serious	SpdLim=40=50mph	0.079	0.750	2.375
28	Coll=J & SpdLim=40=50mph	DrvInj=Serious	0.079	1.000	3.455
29	Coll=J & DrvInj=Serious	Surf=Wet	0.079	0.750	1.781
30	Coll=J & DrvInj=Serious	RdType=SingCgw	0.079	0.750	1.295
31	Coll=J & Area=Urban	Surf=Dry	0.079	1.000	1.810
32	Coll=J & Surf=Dry	Area=Urban	0.079	0.750	2.591
33	Coll=J & Area=Urban	RdType=SingCgw	0.079	1.000	1.727
34	Coll=J & Area=Urban	TrfCtrl=None	0.079	1.000	1.357
35	Coll=J & SpdLim=40=50mph	Prec=FailGiveWayOther	0.079	1.000	2.923
36	Coll=J & SpdLim=40=50mph	Surf=Wet	0.079	1.000	2.375
37	Coll=J & SpdLim=40=50mph	Light=DayNSL	0.079	1.000	1.900
38	Coll=J & SpdLim=40=50mph	Area=Rural	0.079	1.000	1.407
39	Coll=J & SpdLim=40=50mph	HorizGeom=Straight	0.079	1.000	1.267
40	Coll=J & DrvInj=Uninjured	Light=DayNSL	0.079	1.000	1.900
41	Coll=J & DrvInj=Uninjured	HorizGeom=Straight	0.079	1.000	1.267
42	Coll=J & RdType=DualCgw	Light=DayNSL	0.079	1.000	1.900
43	Coll=J & RdType=DualCgw	Area=Rural	0.079	1.000	1.407
44	Coll=J & RdType=DualCgw	FirstIntAct=Car	0.079	1.000	1.357
45	Coll=J & RdType=DualCgw	HorizGeom=Straight	0.079	1.000	1.267
46	Coll=J & Surf=Dry	RdType=SingCgw	0.079	0.750	1.295

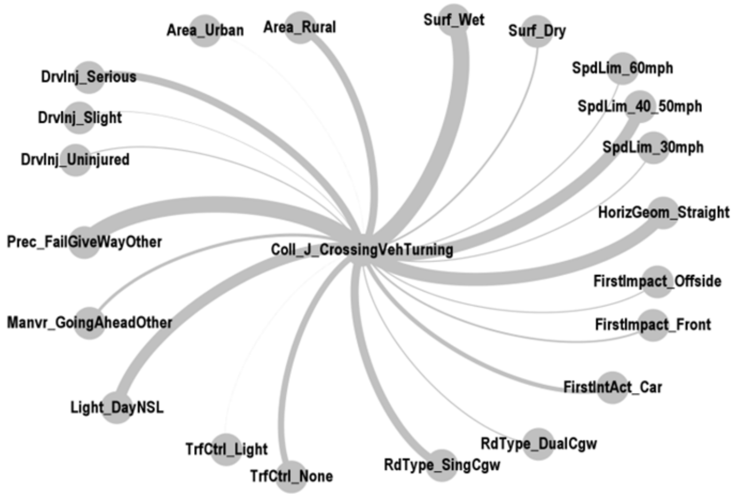


Figure 49: Weighted, directed graph obtained from all association rules for cluster T-C10 having collision type J as consequent

By visually inspecting the graphs and rules tables, the scenarios for this cluster can be described as follows (note that all crashes in the data occurred on UK roads with left-hand traffic):

Scenario T-10.1 (related to collision type L): Car A goes straight on a major road and hits another car B with its front, which is coming from the opposing direction and is turning right into a minor road. This happens on a single carriageway with a speed limit of 40 mph or 50 mph at a non-signalised junction and is caused by B failing to give way. The surface is dry and B suffers serious or fatal injury.

Scenario T-10.2 (related to collision type J): Car A goes straight on a major road and hits another car B, which is emerging from a minor road on the left with the intention to turn right. This happens on a single carriageway in a rural area with a speed limit of 40 mph or 50 mph at a non-signalised junction and is caused by B failing to give way. The surface is wet and A suffers serious injury.

The same procedure was applied to the other clusters and their respective collision types. Table 16 gives all descriptions for the identified high-injury scenarios including the count, followed by simplified illustrations in Figure 50 for T-junctions and Figure 51 for four-legged junctions to better understand the descriptions in the text. The red dots in the figures are the points of impact (i.e. front, offside or nearside). Surface conditions, area (rural/urban), speed limits, vehicle types and injury levels are not shown, but described in the following from the perspective of car A, i.e. the ego car associated with each sample.

The described scenarios build the foundation for further research on testing automated vehicle technologies. It can be observed that there are no rear-end collisions included in the set of high-risk scenarios. This is because the injury outcome was found to be lower for rear-end collisions than for angle collisions.

Table 16: High-injury scenario descriptions

Scenario	Count	Description
Three-legged junctions		
T-4.1 (coll. type L)	39	Car A is turning into a minor road and is hit by a PTW B on its nearside, which is going straight in the opposing direction. A is travelling on a single carriageway with 40-50 mph speed limit without active or static yield instruction. It is caused by A failing to give way or manoeuvring inappropriately.
T-10.1 (coll. type L)	13	Car A is going straight on a major road and hits another car B with its front, which is coming from the opposing direction and is turning right into a minor road. A is travelling on a single carriageway with a speed limit of 40 mph or 50 mph without active or static yield instruction, and it is caused by B failing to give way. The surface is dry and B suffers serious or fatal injury.
T-10.2 (coll. type J)	11	Car A is going straight on a major road and hits another car B, which is emerging from a minor road on the left with the intention to turn right. A is travelling on a single carriageway in a rural area with a speed limit of 40 mph or 50 mph without active or static yield instructions, and it is caused by B failing to give way. The surface is wet and A suffers serious injury.
T-12.1 (coll. type J)	20	Car A is turning right into a major road and is hit by a PTW B on the offside, which is going straight on the crossing path. A is travelling on a rural single carriageway controlled by a static give-way sign and it is caused by A failing to give way. The surface is wet and B suffers serious or fatal injury.
T-12.2 (coll. type G)	9	Car A is turning right into a minor road and is hit on the offside by a PTW B, which is overtaking. A is travelling on an urban single carriageway with 30 mph speed limit without active or static yield instruction, and it is caused by an inappropriate overtake from B.
T-12.3 (coll. type M)	6	Car A is turning left into a major road and is hit by a PTW B on its offside, which is going straight on the major road from the right. A is travelling on an urban single carriageway with 30 mph speed limit controlled by give-way signs, and it is caused by A failing to give way. B suffers serious or fatal injury.
T-13.1 (coll. type L)	25	Car A is turning right into a minor road and hits a PTW B with its front, which is going straight in the opposing direction. A is travelling on a rural single carriageway with 30 to 50 mph speed limit without active or static yield instruction, and it is caused by A failing to give way or manoeuvring inappropriately. The surface is wet and B suffers serious or fatal injury.
Four-legged junctions		
X-1.1 (coll. type H)	59	Car A is going straight on a major road and hits another car B with its front, which is crossing the path from the left. A is travelling on a rural single carriageway with 60 mph speed limit without active or static yield instruction and it is caused by B failing to give way.
X-2.1 (coll. type H)	25	Car A is crossing a four-legged junction and hits another car or PTW B with its front, which is crossing the path from the right. A is travelling on a rural single carriageway road with 40-50 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. A remains uninjured or suffers slight injury. Max. injury of B: Serious or fatal. No clear indication on the surface condition.
X-4.1 (coll. type J)	12	Car A is turning right into a major road and is hit by a car or LGV B on the offside, which is going straight on the major road from the right. This happens on a rural dual carriageway with 40-50 mph speed limit controlled by static give-way signs and is caused by A failing to give way. The surface is wet and A suffers serious or fatal injuries.
X-6.1 (coll. type H)	15	Car A is going straight on a major road and is hit by car B on the offside, which comes from a minor road and crosses the path from the right. A is travelling on a single carriageway road with 30 mph speed limit controlled by traffic lights, and it is caused by B failing to give way. The surface is wet and B suffers serious or fatal injuries.
X-6.2 (coll. type L)	7	Car A is going straight on a major road and is hit by car B on its offside, which turns right from the opposing direction. A is travelling on a road with 60 mph speed limit controlled by traffic lights, and it is caused by B losing control of the vehicle. B suffers serious or fatal injuries.

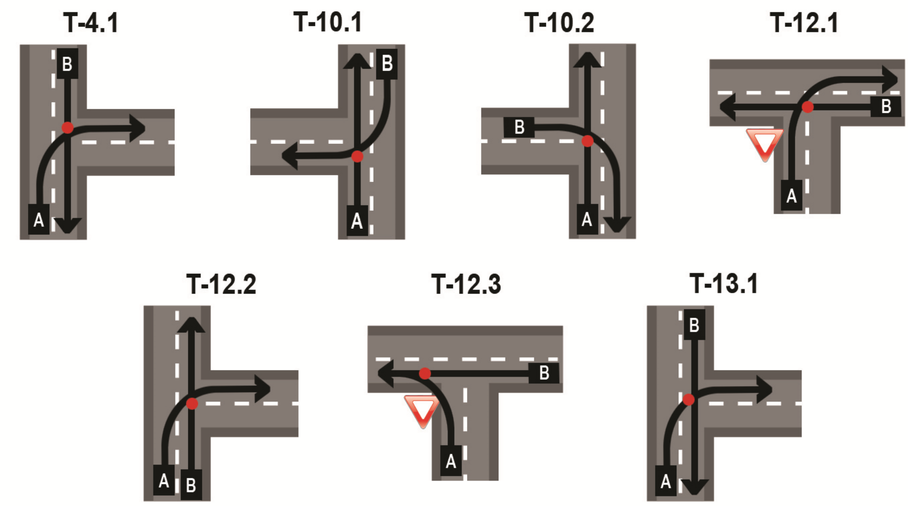


Figure 50: Simplified illustrations of all high-injury scenarios identified for T-junctions

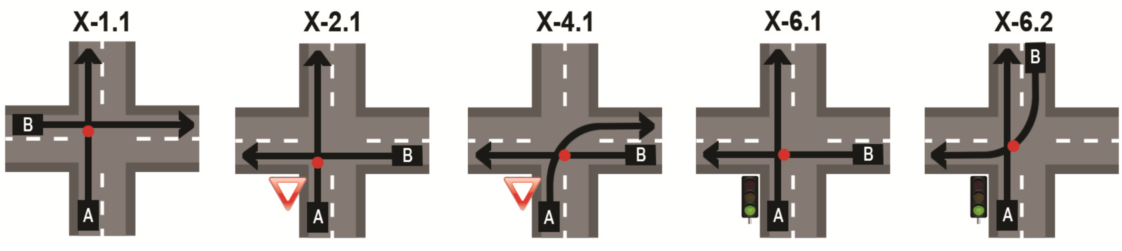


Figure 51: Simplified illustrations of all high-injury scenarios identified for four-legged junctions

There was no high-injury scenario found that involves car-pedestrian or car-bicycle collisions only. This can be explained by the low number of pedestrians (2.2%) and cyclists (2.5%) among all involved persons in the given accident dataset, compared to car occupants (78.5%), motorcyclists (6.3%) or goods vehicle occupants (8.6%). Although there is no doubt about the importance of vulnerable road user safety, neither the cluster analysis nor the association rule method resulted in a distinct pedestrian or cyclist scenario. Considering the frequency of certain crash types at junctions, car-pedestrian and car-cyclist collisions are discounted, which might not be true if injury frequencies were taken into account.

5.4.3.2 High-frequency scenarios

This section presents the scenarios that were found to be most frequent, independent on the injury outcome. Figure 52 and Figure 53 depict the top five high-frequency scenarios for three-legged and four-legged junctions, respectively, i.e. the scenarios with the highest number of crashes included. Note that the traffic control type is given for the path of A.

Table 17 shows the crash counts for each high-frequency scenario including a short description. Some of the three-legged junction scenarios were combined due to their similarities. For example, T-2.1 and T-8.1, which were derived from two different clusters, were grouped. This was also done for the second and third most frequent scenario groups for three-legged junctions. The count column in the table gives the number of crashes within the respective cluster that are allocated to the particular collision type. For example, the 44 samples for T-1.1 are the collisions of type F (rear-end) within cluster T-C1.

Figure 52: Simplified illustrations of the five most frequent scenarios identified for T-junctions

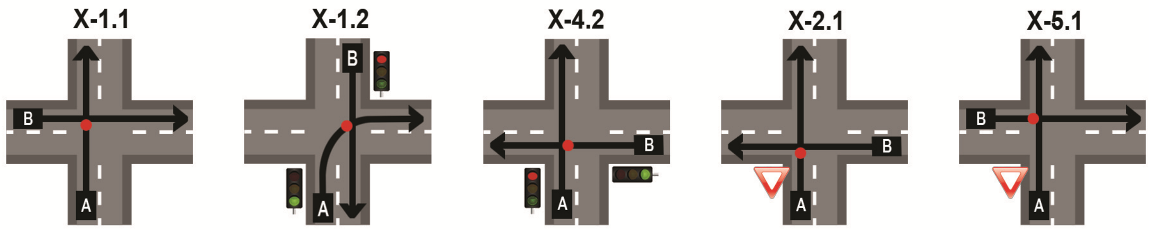


Figure 53: Simplified illustrations of the five most frequent scenarios identified for four-legged junctions

It can be observed that the top five most frequent scenarios at four-legged junctions do not include rear-end collisions. This finding corresponds to the crossing-path scenarios identified by Najm et al. (2001), which are primarily angle crashes. Furthermore, there is no particular scenario involving car-pedestrian or car-bicycle collisions only. This can be explained by the low number of pedestrians (2.4%) and cyclists (3.6%) as collision partners, compared to other cars (71.8%), PTW (9.7%) or goods vehicles (7.8%).

The frequency of collision types per cluster is given in Table 18 for T-junctions and in Table 19 for four-legged junctions. Grey cells represent high-frequency scenarios, while the cells with red font are the high-injury scenarios. Note that there are many different collision types per cluster with smaller sample sizes, which can be explained by the limitations of clustering. As shown in Section 5.3.3, some samples are wrongly allocated to a neighbouring cluster and do not necessarily belong to the key scenarios for the respective partition.

It can be seen that the high-frequency scenarios at three-legged junctions do not include any of the high-injury scenarios. The high-frequency scenarios for three-legged junctions include two rear-end collisions (T-5.1/5.2 and T-1.1), which are not included in the high-injury scenarios. This is because the injury outcome was found to be lower for rear-end collisions than for angle collisions, which was also reported by Beck (2015).

Table 17: High-frequency scenario descriptions

Scenario	Count	Description
Three-legged junctions		
T-2.1/8.1 (coll. type J)	99	Car A is turning right into a major road and hits another car with its front, which is coming from the right. A is travelling on an urban single carriageway with 30 to 50 mph speed limit controlled by a static give-way sign and it is caused by A failing to give way, leading to no or slight injury.
T-1.2/7.4/9.1 (coll. type J)	69	Car A is going straight on a major road and hits another car B with its front, which turns right from a minor road joining from the left. A is travelling on an urban single carriageway with 30 to 50 mph speed limit without active or passive yield instruction, and it is caused by B failing to give way. The surface is dry and A suffers slight injuries.
T-5.1/5.2/11.2 (coll. type F)	55	Car A is going straight on a major road and hits car B at the rear-end, which is going straight. A is travelling on a single or dual carriageway with 30 to 60 mph speed limit without active or static yield instruction. It is caused by A failing to stop or to avoid, leading to no or slight injury.
T-1.1 (coll. type F)	44	Car A is going straight on a major road and hits another car at the rear end. A is travelling on a rural dual carriageway with 70 mph speed limit without active or passive yield instruction, and it is caused by A failing to stop or failing to avoid or by other precipitating factors from B. The surface is dry and A remains uninjured or suffers slight injury.
T-1.3 (coll. type L)	42	Car A is going straight on a major road and hits another car B with its front, which is turning right into a minor road. A is travelling in dark light conditions on an urban single carriageway with 40 to 50 mph speed limit controlled by traffic lights, and it is caused by B failing to give way or manoeuvring inappropriately. A suffers slight injuries.
Four-legged junctions		
X-1.1 (coll. type H)	47	Car A is going straight on a major road and hits another car B with its front, which is crossing the path from the left. A is travelling on a rural single carriageway with 60 mph speed limit without active or static yield instruction and it is caused by B failing to give way.
X-1.2 (coll. type L)	28	Car A is turning right into a minor road and hits another car B, which is coming from the opposing direction. A is travelling on an urban road with 40 to 50 mph speed limit controlled by traffic lights, and it is caused by B violating the red light. Max. injury: Slight.
X-4.2 (coll. type H)	24	Car A is crossing a major road and is hit by another car B on its offside, which is crossing from the right. A is travelling on an urban single carriageway road controlled by traffic lights, and it is caused by A failing to give way. Max. injury: Serious or fatal. There is no clear indication of the surface condition.
X-2.1 (coll. type H)	21	Car A is crossing a four-legged junction and hits another car or PTW B with its front, which is crossing the path from the right. A is travelling on a rural single carriageway road with 40-50 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. A remains uninjured or suffers slight injury. Max. injury of B: Serious or fatal. No clear indication of the surface condition.
X-5.1 (coll. type H)	21	Car A is crossing a major road and is hit by another car B on its nearside, which is crossing from the left. A is travelling on an urban or rural single carriageway road with 30 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. There is no clear indication of the injury severity.

Table 18: Frequency of collision types for the T-junctions clusters (red font: high-injury scenarios, grey fill: high-frequency scenarios)

Collision letter	T-C1	T-C2	T-C3	T-C4	T-C5	T-C6	T-C7	T-C8	T-C9	T-C10	T-C11	T-C12	T-C13
A: Overtaking and lane change	10	0	1	4	4	3	6	3	6	2	3	2	0
D: Cornering	8	1	0	1	2	3	2	2	2	0	0	0	0
F: Rear end	64	7	43	0	48	7	18	1	1	7	6	1	1
G: Turning vs. same direction	3	0	13	7	13	0	3	0	3	0	1	9	0
H: Crossing no turns	14	6	2	3	10	10	8	9	5	5	7	1	2
J: Crossing vehicle turning	47	55	3	3	0	4	7	60	24	11	3	20	3
L: Right turn against	47	8	0	39	7	0	11	6	2	13	0	2	25
M: Manoeuvring	13	10	0	5	15	15	7	2	6	1	15	6	1
P Pedestrian other	3	2	0	0	3	1	1	0	1	3	2	1	3
Other (B, C, E, K, N, Q)	3	1	0	0	0	0	0	0	2	4	1	0	0
SUMME	**212**	**90**	**62**	**62**	**102**	**43**	**63**	**83**	**52**	**46**	**38**	**42**	**35**

Table 19: Frequency of collision types for the four-legged junctions clusters (red font: high-injury scenarios, grey fill: Top 5 high-frequency scenarios)

Collision letter	X-C1	X-C2	X-C3	X-C4	X-C5	X-C6
A: Overtaking and lane change	3	0	0	1	0	2
D: Cornering	3	0	0	0	0	0
F: Rear end	17	7	0	0	1	0
G: Turning vs. same direction	4	1	3	0	0	2
H: Crossing no turns	59	25	17	30	23	15
J: Crossing vehicle turning	8	5	8	12	0	1
L: Right turn against	34	11	18	4	6	7
M: Manoeuvring	7	8	2	2	4	7
P Pedestrian other	7	2	0	0	1	0
Other (B, C, E, K, N, Q)	0	1	0	0	0	0
SUMME	**142**	**60**	**48**	**49**	**35**	**34**

5.5 Conclusions and outlook

This chapter presented a novel approach on how to extract pre-crash scenarios from accident data, which was applied to three-legged and four-legged road junctions in the UK. The clustering method k-medoids was found to be most appropriate for the given dataset, since it is robust against outliers and can cope with categorical data. The study resulted in thirteen crash clusters for T-junctions and six crash clusters for four-legged junctions. Association rules were computed for each cluster and revealed associated crash characteristics, which were the basis for the scenario descriptions. Considering the clusters with high injury outcome, twelve pre-crash scenarios were identified, which constitute the core population of critical driving situations in the given dataset. Failure to give way and inappropriate manoeuvres are among the main precipitating factors. There was no scenario found involving car-pedestrian or car-bicycle collisions, which can be explained by the low number of vulnerable road users among all involved persons in the given accident dataset. In summary, however, the results support existing findings about junction safety and add further definition to the clusters identified. For example, as indicated in the literature, higher injury levels coincide with powered two-wheelers involved as well as higher speed limits.

The study was preparatory research for the development of a sub-microscopic simulation framework, which is described in the chapter that follows. The scenarios obtained help to reduce the possible number of simulation parameter variations, such as vehicle trajectories, velocities as well as road and junction parameters. The crucial question to answer is how to transfer the scenarios into simulation and how to enhance them with representative variations of real-world effects.

6 Study 3: Developing a simulation and evaluation framework

As described in the previous chapter, study 2 resulted in a set of safety-critical scenarios at road intersections, derived from historical junction crash data. The third study aims to transfer the derived collision scenarios to a sub-microscopic traffic simulation environment, where the safety performance of automated driving functions can be evaluated. While the second study is based on crashes involving human drivers only, this study presents a novel methodology that adds the automated driving functionality to understand their impact on intersection safety. The study results and findings deliver answers to the research questions 4 and 5.

6.1 Problem definition

Numerous research projects have been conducted to virtually assess the safety performance of automated vehicles (Beglerovic et al., 2017; Olivares et al., 2016; Pütz et al., 2017; Rodarius et al., 2015; Roesener et al., 2017). Virtual testing can decrease costs in the development cycle, because it can comprise high numbers of scenarios with combinations of varying factors, which would not be feasible in real-world tests. To the author's knowledge, there was no research undertaken and published, which particularly defined testing scenarios and procedures for evaluating the safety performance of ADS at junctions.

The scenarios obtained in study 2 help to reduce the possible number of simulation model parameter variations, such as vehicle manoeuvres, trajectories, velocities as well as road and junction parameters. However, not all parameters for the simulations can be derived from the clustering analysis. Some data elements are simply not included in the accident database, such as exact driving trajectories or driving behaviour. The proposed methodology fills this gap by enhancing the collision scenarios with representative variations of real-world conditions, sampled by a Monte Carlo approach.

Automated driving functions are influenced by various factors that can be divided into behaviour of other traffic participants, road and junctions design as well as adverse weather and light conditions. It is assumed that the sensor functionality works reliably under "normal" laboratory conditions, but real-world effects and events can compromise it. For example, vision sensors struggle with limited visibility due to adverse weather conditions or sight obstructions and the motion planning might be influenced by unexpected behaviour of other road users. This limited functionality poses a risk not only for the automated vehicle's occupants, but also for other road users. There is a need to investigate the influence of such factors, henceforth referred to as "criticalities", on the safety performance in various scenarios at road junctions.

Besides the normal automated driving function such as lateral and longitudinal control, collision avoidance and mitigation systems play an important role for safe operation of ADS at road junctions (see review in Section 2.4). Common systems such as forward collision avoidance (or automated emergency braking) cannot cope with all accident or conflict situations at junctions. The automotive industry strives to develop reliable systems able to avoid side or angle collisions, including vehicular communication technologies. However, there are no common testing procedures and the high number of possible combinations of testing conditions pose a challenge. This thesis leads to recommendations for testing and validating automated driving systems at junctions, with focus on virtual vehicle testing as a pre-stage or parallel activity to field operational tests on public roads, including static (e.g. road design and layout) and dynamic content (e.g. involved road users and vehicles, their trajectories and behaviour). In summary, the research questions addressed by this study are RQ4 and RQ5:

- How can those collision scenarios be represented and enhanced for sub-microscopic simulation to evaluate the safety performance of intersection assistance systems?
- What general recommendations can be made for the safety performance indicators to be considered in virtual testing of ADS at junctions?

6.2 Ontology for simulation experiments

Before describing the simulation framework in detail, this section defines a set of important terms used in the further sections. In literature, there is no consistent terminology for scenarios and vehicle simulation of automated and assisted driving. Kienle et al. (2014) proposed a fundamental ontology to simplify communication and the exchange of findings in this research field. This thesis uses an adapted version of this ontology, as given in Figure 54. The figure depicts a hierarchical structure to illustrate the different levels in the simulation study. In the following sections, the terminology is defined in detail.

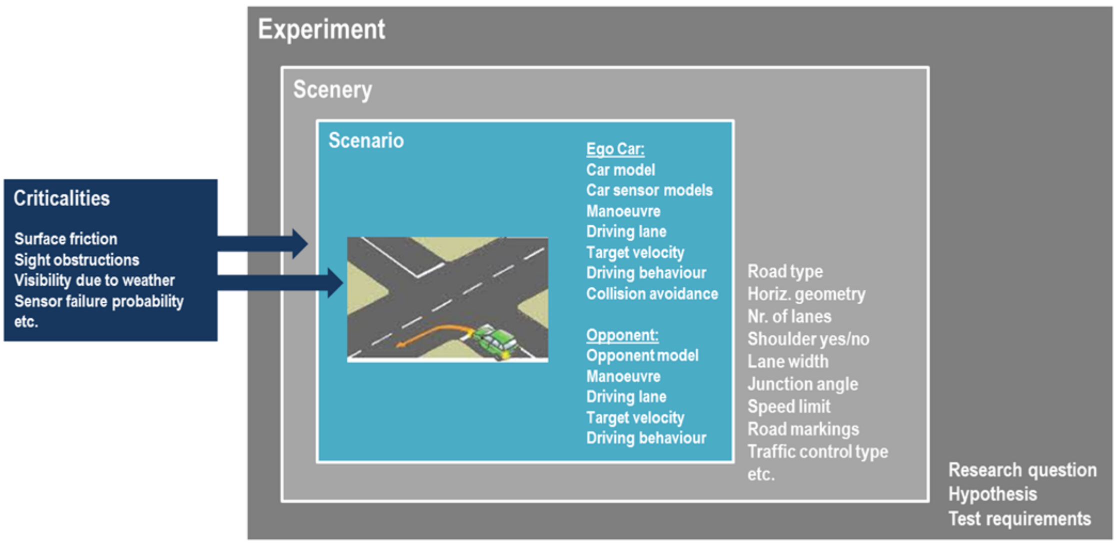


Figure 54: Ontology and hierarchy for simulation experiments (adapted from Kienle et al. (2014))

6.2.1 Experiments

As the top level, an *experiment* is defined as a systematic procedure carried out to validate a hypothesis. It is further specified by the requirements for the tests, such as the vehicle or road models to be included or the study duration and computational constraints to be considered. The applicability of the simulation and evaluation framework is demonstrated by an experiment presented in Section 6.6.

6.2.2 Sceneries

The term *scenery* describes all static elements of the surrounding area. In other words, it is defined by the layout and geometry of the road section to be simulated, including static single elements such as road signs, markings or traffic lights. The road type could be a single or dual carriageway with a specific number of lanes with or without hard shoulder. The horizontal geometry can be a left or right curve or straight, the junction shape can be three- or four-legged, controlled by a traffic light, static yield instruction or uncontrolled. Furthermore, the junction can be equipped with an infrastructure-based intersection assistance system. The parameters derived from the association rules in study 2 help to specify a scenery to be simulated.

The following basic sceneries are examples that can be used for the simulation experiments and are based on the Highways England design guidelines on priority junctions ("TD 42/95 - Geometric Design of Major/Minor Priority Junctions," 1995).

- *Simple T-junction* (see Figure 55): without any ghost or physical islands in the major road, without channelising islands in the minor road approaching and with road markings.
- *Ghost island T-junction* (see Figure 56): with a physical island on the minor road and ghost island markings on the major road to direct traffic movement.
- *Single-lane dualling T-junction* (see Figure 57*):* with a physical island on the minor road and central reservation islands on the major road to direct traffic movements.
- *Dual-carriageway T-junction* (see Figure 58): with a widening of the central reserve on the major road to provide an offside diverging lane and waiting space for vehicles turning right into the minor road.
- *Simple four-legged junction* (see Figure 59), with static give-way instruction, two roads crossing at right angle and road markings

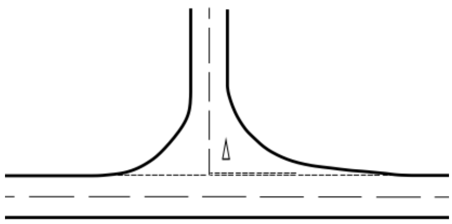


Figure 55: Scenery T_{simple}: Simple T-junction

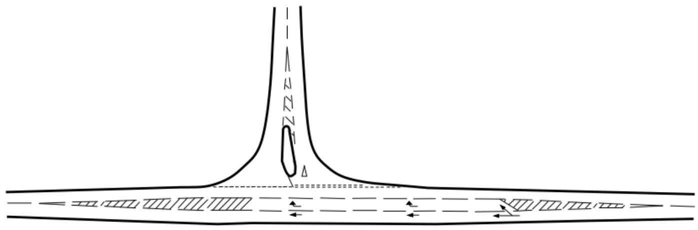


Figure 56: Scenery T_{ghost}: Ghost island T-junction

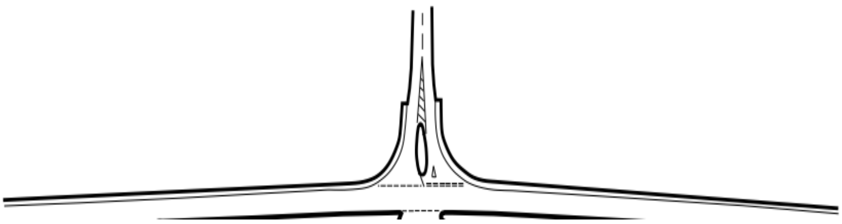


Figure 57: Scenery T_{island}: Single-lane dualling T-junction

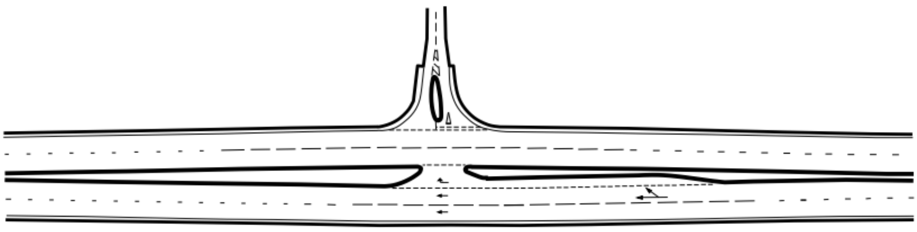


Figure 58: Scenery T_{dual}: Dual-carriageway T-junction

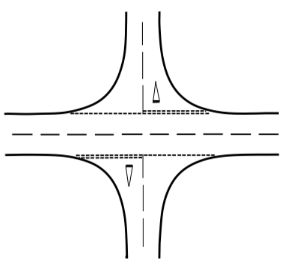


Figure 59: Scenery X_{simple}: Simple non-signalized crossroads

6.2.3 Scenarios

While the scenery describes all static elements of the road environment, a *scenario* introduces dynamic traffic objects, i.e. moving road users. For each scenery, a certain set of scenarios is simulated according to the results of study 2. A scenario thus defines the interaction between the ego car and its opponent, including their manoeuvres, their exact trajectory and velocity (driving behaviour) to conduct the manoeuvre and their position, i.e. which lane they are driving. Furthermore, the vehicle models are specified, along with the sensor models of the ego car.

6.2.4 Ego car

The sub-microscopic simulation environment requires a primary car as the main subject of investigation. In this study, the *ego car* is the automated car equipped with modern sensors and collision avoidance technologies. In contrast to the *opponent* vehicles, the ego car has detailed physical models, driving dynamics and physical sensor models.

6.2.5 Criticalities

The term *criticality* defines a certain factor or condition that may influence the safety performance of automated collision avoidance systems. It is assumed that a criticality has a negative impact on the performance, but this needs to be studied. For example, poor surface friction influences the braking performance in critical situations. Poor visibility due to weather influences the sensor perception and sight obstructions pose challenges for object detection. An additional criticality is sensor failure or inaccuracy, which can occur due to interference, e.g. for radar.

6.3 Requirements and selection of simulation software

Section 2.5.2 introduced common sub-microscopic simulation software used for virtual vehicle testing. The following simulation platforms were selected as candidates to be used for the study, because they provided ready-to-use models for the vehicle and sensors as well as driving behaviour:

- PreScan, by Tass International
- CarMaker, by IPG Automotive

Besides others such as dSPACE, VTD and Pro-SiVIC, those two platforms are widely used in the field of pre-crash simulation and ADAS development. They are designed to test vehicle components, either in SIL, MIL, HIL or VIL, but have limitations in modelling physical vehicle deformation, rebound effects and occupant injury. For this purpose, other software tools such as PC-Crash, Virtual CRASH or MADYMO are used, primarily for accurate accident reconstruction. However, this thesis focuses on the simulation of pre-crash situations and not on post-impact effects. Therefore, a list of requirements was collected, which are necessary to complete this study. For both tools, test licences were provided with full functionality for a maximum of one month. During this test period, the functionality of the tools was reviewed with regards to the requirements. The resulting checklist for CarMaker is given in Table 20, the one for PreScan in Table 21. Note that the requirements were collected in February 2016 with the software releases that were available at that time. Some of the functionalities have been updated during the development of this thesis.

Table 20: Requirement analysis of CarMaker

Requirement	Fulfilled (✓/~/x)	Comment
Environment sensor models to simulate camera, LIDAR, radar and others under varying conditions	~	Recognition sensors included, but as ideal models. Realistic models are included from Version 6, but not in combination with Simulink.
Road and roadside models to create a virtual road environment including junction features, which can also be perceived by the sensor models	✓	Road and roadside models can be added and are also perceivable by the sensors.
Realistic vehicle models to simulate different vehicles with varying mass, shape and handling	✓	Available with a detailed modelling of engine, tire, steering, suspension, brake and more.
Pedestrian and bicycles models, including dark clothing etc.	~	Pedestrian and bicycles can be added as traffic objects. Textures for certain objects could be added but they do not influence the sensors.
Realistic driver or manoeuvre modelling to allow reconstructing crash scenarios and road user movements	✓	Certain driver models are available, also ABS, ESP implementations are available, realistic simulations through proper modelling.
Capability to simulate adverse weather conditions (perceived by sensor models)	x	Weather conditions (fog is currently available) can be added to the simulation environment but they do not influence the sensors.
Scripting and automation functionality and links to other platforms, e.g. to MATLAB/Simulink	✓	Matlab/Simulink integration and scripting automation support is available, TestManager is implemented in the GUI
Capability to simulate limited visibility, e.g. darkness and realistic headlight beam modelling to simulate visibility at night	~	Darkness and headlight beams can be added to the simulation movie, but will not affect the detection rate of the sensor. However, sensor parameters can be manually estimated for different conditions.
Capability of simulating imperfections of traffic signs and lane markings, which influences the sensor models	~	Not implemented, workaround through recognition of the sensor rate is possible.
Import of road infrastructure data and (semi-) automatic generation of road models	~	Automatical import of *.kml files is available, but the network can also take general settings then (lane width, friction stripes, etc.), automated generation of additional infrastructure (road markings, traffic signs, etc.) is currently not supported.
Interface to OpenCRG	✓	Available by adding a road marker to a road segment
Interface to OpenDRIVE	x	Not implemented, but planned for future versions.
Non-obligation to buy additional software tools (e.g. Matlab/Simulink)	✓	Works for simulation purposes standalone, Matlab/Simulink extension is optional.

Table 21: Requirement analysis of PreScan

Requirement	Fulfilled (✓/~/✗)	Comment
Environment sensor models to simulate camera, LIDAR, radar and others under varying conditions	✓	Realistic sensor models are available.
Road and roadside models to create a virtual road environment including junction features, which can also be perceived by the sensor models	✓	Virtual road environment can be built in detail. Due to predefined road segments a network can be set up easily and quickly. Sensor models can perceive objects.
Realistic vehicle models to simulate different vehicles with varying mass, shape and handling	~	Available with a predefined modelling of driveline (transmission, shifting strategies), engine and suspension. Simulink is necessary for modification.
Pedestrian and bicycles models, including dark clothing etc.	✓	Several human models available (children, men, woman, with/without raincoat etc.). Textures can be defined per object to change clothing or upper/lower body clothing colour can be set.
Realistic driver or manoeuvre modelling to allow reconstructing crash scenarios and road user movements	~	Driver models are available, but a data verification test showed that no ABS, ESP is implemented per se and that models are idealised.
Capability to simulate adverse weather conditions (perceived by sensor models)	✓	Different weather can be added and influences the recognition rate of sensors
Scripting and automation functionality and links to other platforms, e.g. to MATLAB/Simulink	✓	Matlab/Simulink integration, TestManager in the GUI implemented
Capability to simulate limited visibility, e.g. darkness and realistic headlight beam modelling to simulate visibility at night	✓	Darkness and headlight beams can be added to the simulation movie and affect the detection rate of the sensors.
Capability of simulating imperfections of traffic signs and lane markings, which influences the sensor models	✓	Imperfections are implemented for traffic signs and road markings and are sensor detectable.
Import of road infrastructure data and (semi-)automatic generation of road models	~	Only through OSM, but with limited capability for traffic signs (workaround by adding them in the OSM-File). Additional infrastructure (road markings etc.) must be added manually.
Interface to OpenCRG	✗	Not available, implementation is evaluated by TASS.
Interface to OpenDRIVE	✗	Not available, implementation is evaluated by TASS.
Non-obligation to buy additional software tools (e.g. Matlab/Simulink)	✗	The standalone simulation does not take dynamics, etc. into account. Matlab/Simulink availability is mandatory.

In summary, CarMaker has its strengths, when it comes to the availability of predefined models. For example, sensor and driver models are available from scratch and can be configured by the user. In contrast to that, PreScan requires MATLAB/Simulink to build one's own models for the driving control. This is a drawback in relation to this thesis, because ready-to-use driver models are not available.

The main disadvantage of CarMaker is that weather and visibility conditions can only be added as a visual feature in the video creation, but not as integrated part of the environment model. This means that the effect on sensor perception must be estimation by modifying the values for range and accuracy of the perception, e.g. reduced range in case of fog.

Based on the requirement analysis, taken into consideration the slightly higher price of PreScan, it was decided to purchase the CarMaker software.

6.4 Simulation framework

The following section will explain how the simulation framework was set up and which models are included. After explaining the data input configuration and the test automation method, the models are categorised into road environment, vehicle and sensors and driving behaviour.

6.4.1 Overview of simulation architecture

The software CarMaker distributed by the company IPG is a vehicle simulation software with various functionality and tools for investigating complex traffic scenarios. It allows virtual test driving of passenger cars and light utility vehicles, by simulating the interaction between vehicle, driver, other traffic and road environment. Apart from the classic vehicle dynamics simulation, CarMaker enables the testing of driver assistance systems, which is the main focus of this thesis, since automated driving can be seen as a further advancement of driving assistance. The simulation possibilities comprise model-in-the-loop (MIL) simulations, as done in this thesis, as well as hardware-in-the-loop (HIL) tests on specific electronic components (Holzmann, 2006).

The developed simulation architecture comprises five elements, as shown in Figure 60. The input for the simulations, i.e. the varying parameters and combinations of those are specified by an XML-structured configuration file (1). This allows the script file for the simulation automation to be kept slim without changing the code for each new test run. The automation script (2) is implemented in MATLAB and sets up the simulation environment, interprets the input XML-file and communicates with the CarMaker-Simulation instance. In Simulink, the CarMaker model blocks can be modified and input and output quantities can be specified. The CarMaker GUI (3) is then executed to perform the simulation runs and to deliver the specified output quantities as CSV-files for each run. Those output files are stored in the simulation database and are post-processed to compute the safety indicators (4), which are described in Section 6.4. Finally, the resulting indicators, including the information if it was a collision, near-miss or no conflict at all, are stored in the results database (5).

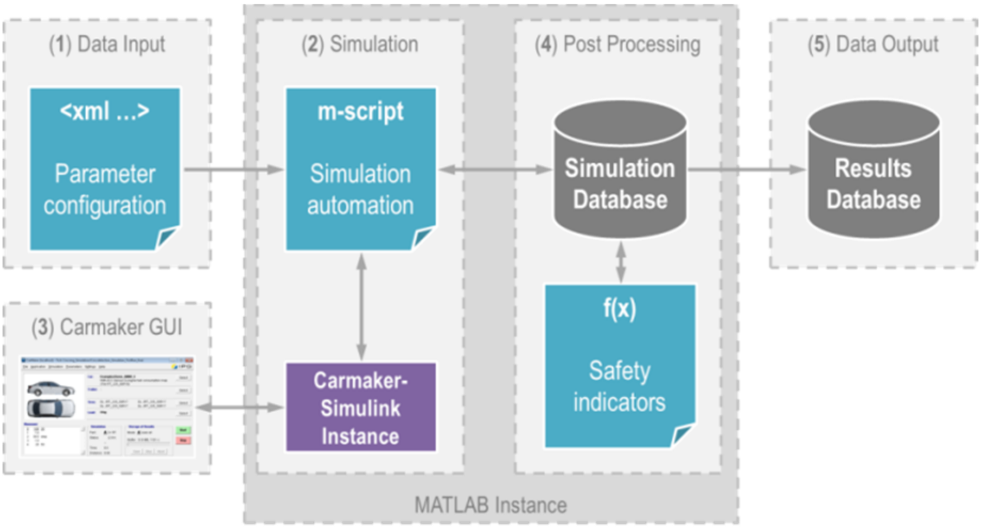


Figure 60: Simulation architecture

6.4.2 XML-based data input and parameter configuration

The number, values and ranges of the simulation parameters vary depending on the test requirements and objectives. Therefore, a graphical user interface was created in C#, which allows the user to configure the parameters and their variations and to automatically generate XML-files as inputs for the simulation automation script. The XML structure and the tag names are based on the hierarchy given in Figure 54 and allow to configure different experiments and scenarios.

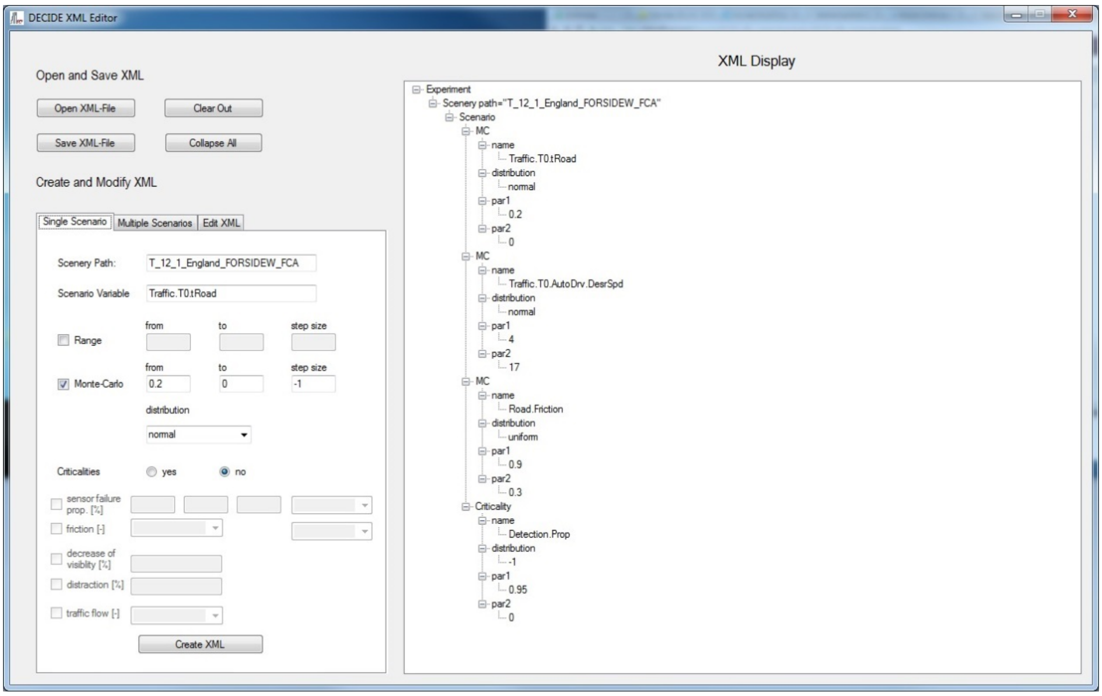


Figure 61: GUI for XML-based parameter configuration

Within a scenario, the user must define at least one parameter variation tag, which consists of a name (that must be the same as in the CarMaker parameter list) and its actual numerical variation. A parameter can either be a Monte Carlo parameter, where the input probability distribution is specified (see Section 6.7) or a parameter that is varied in a range of fixed values with a fixed interval. Figure 61 depicts a screenshot of the GUI with the parameter configuration of the demonstration experiment presented in Section 6.8.

The resulting XML-file (see below) is interpreted in MATLAB by the simulation test automation script, which is explained in the section that follows.


```xml
<?xml version="1.0" encoding="utf-8"?>
<Experiment>
  <Scenery path="T_12_1_England_FORSIDEW_FCA">
    <Scenario id="1">
        <MC>
            <name>Traffic.T0.tRoad</name>
            <distribution>normal</distribution>
            <par1>0.2</par1>
            <par2>0</par2>
        </MC>
        <MC>
            <name>Traffic.T0.AutoDrv.DesrSpd</name>
            <distribution>normal</distribution>
            <par1>4</par1>
            <par2>17</par2>
        </MC>
        <MC>
            <name>Road.Friction</name>
            <distribution>uniform</distribution>
            <par1>0.9</par1>
            <par2>0.3</par2>
        </MC>
        <Criticality>
            <name>Detection.Prop</name>
            <distribution>-1</distribution>
            <par1>0.95</par1>
            <par2>0</par2>
        </Criticality>
    </Scenario>
  </Scenery>
</Experiment>
```


6.4.3 Simulation test automation

A MATLAB script was written to prepare and to execute the simulations as well as to process and export the simulation results. Based on the input XML-file, the varying parameters are created and transferred to the CarMaker/Simulink instance.

The **simulation procedure** consists of the following main steps:

1. Manually create a CarMaker test run for each scenery, configuring all static parameters.
2. Import the inputs from the XML-file and add the variable parameters to the MATLAB workspace.
3. Create the output directory structure, following the hierarchy given in Figure 54 and based on the predefined parameter configuration (e.g. "…\ *ExperimentA \ Scenery1 \ Scenario1 \ ParameterVariation1*").
4. Load the Simulink model and execute two simulation runs for each parameter variation in CarMaker,

    a. one without an intervening manoeuvre from the vehicles ("reference run") and
    b. one with collision avoidance models activated

5. Save the output files of each simulation run into the predefined output directory.

Step 1 is the only manual task of the presented testing procedure, because the static parameters must be modelled in the CarMaker GUI. Therefore, a CarMaker *test run* is created for each scenery. Detailed information on the CarMaker GUI and the simulation models is given in Section 6.6. Besides the static road information, there may be other parameters that do not vary between the scenarios, such as the ego car model or its manoeuvres. For instance, the proposed demonstration experiment in this chapter varies the speed and lateral position of the opponent car, the road friction and the ego car's sensor systems. Hence, all other parameters such as vehicle and component models or the ego car's velocity and manoeuvres are also set static in the test run.

Step 2 imports the XML file and creates the respective variables in the MATLAB workspace. Plausibility checks are done to ensure the correct hierarchy structure, which should be *Experiment->Scenery->Scenario->Parameter variation*. This structure is also used for creating the output directories in Step 3.

Step 4 initiates the Simulink interface and executes the simulation. Each simulation run is defined by a certain combination of parameter variations sampled by a Monte Carlo approach (see Section 6.7). However, there are two simulation runs necessary for each parameter combination, one with all collision avoidance models deactivated (*reference run*) and one with activated collision avoidance. The *reference run* can be seen as an equivalent to the path and motion planning of automated vehicles, where the future trajectories of both the ego vehicle and the potential opponent road user are predicted. This "prediction" is replaced by the reference runs and can thus be seen as ideal without prediction errors. The calculation of some of the safety performance indicators require the future trajectory to imitate an automated vehicle's path and motion planning. As soon as an intervening manoeuvre is performed by one of the vehicles, e.g. an emergency braking, the actual trajectory differs due to vehicle instabilities such as skidding.

In Step 5, the simulation outputs are exported. The evaluation procedure, which is explained in the following section, requires the outputs listed in Table 22. This list of output quantities is used throughout the following sections to explain the evaluation procedure.

An overview of all MATLAB functions for the simulation test automation is given in Appendix G.

Table 22: Simulation output quantities for both vehicle A and B

Output description	Variable name
Global time	*t*
Distance covered over time for both vehicles	*vehc_A/B_Route_x*
Position of both vehicles over time in the global coordinate system	*vehc_A/B_Global_x*, *vehc_A/B_Global_y*
Velocity of both vehicles over time	*vehc_A/B_v*
Longitudinal accelerations of both vehicles over time	*vehc_A/B_ax*
Brake/decelerator activity for both vehicles over time	*vehc_A/B_brk*
Binary collision detection flag for each timestamp	*collision_detect*

6.5 Safety performance evaluation

The previous section described the overall simulation architecture and how a simulation run is configured and executed. Each simulation run has one output file including all quantities of Table 22, which is further analysed by an evaluation script to derive safety indicators. This section is devoted to the metrics used to quantify the safety performance of the automated vehicle in the given scenario. Basically, the safety performance analysis is similar to the traffic conflict technique, which aims to observe and record unsafe interactions between vehicles or road users (Perkins and Harris, 1968, 1967). Instead of analysing detected collisions only, near-misses are also taken into account. (Hydèn, 1987) published a safety pyramid (see Figure 62), which illustrates that collisions are only the tip of the iceberg regarding driving situations.

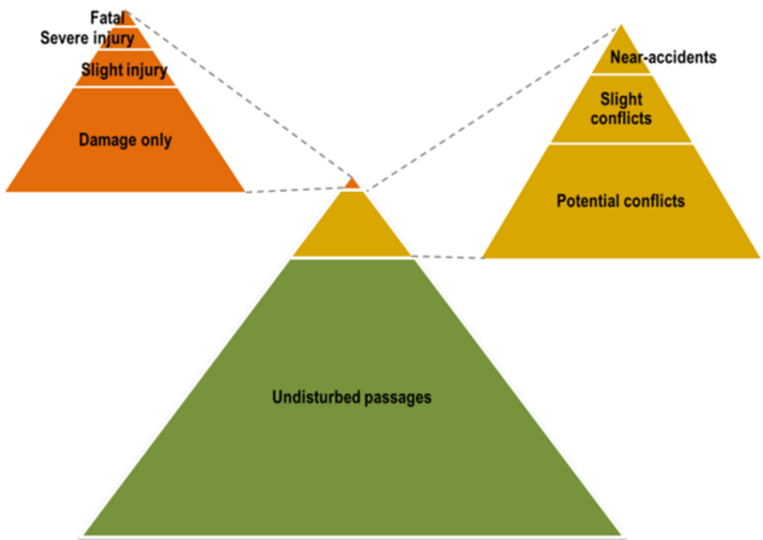


Figure 62: The safety pyramid adapted from (Hydèn, 1987)

A large number of events happening in traffic are not reported, such as conflicts, which are defined as "traffic events involving the interaction of two or more road users, where one or both drivers take evasive action such as braking or swerving to avoid a collision" (Parker and Zegeer, 1989). However, this definition can be extended, because there might be events where none of the involved road users takes an evasive action, but they almost collide. For example, this could happen if there is not enough time to react or the road users fail to react in time. There is consensus among traffic safety experts that detected conflicts can be an appropriate predictor for collisions, depending on factors such as traffic volumes (Sacchi and Sayed, 2016). In this simulation study, the frequency and severity of collisions as well as the frequency of conflicts help to quantify the safety performance of the system under test.

6.5.1 Overview of safety indicators used

Tolouei et al. (2013) mentioned two aspects of the safety performance of a vehicle in traffic. First, the *primary safety performance*, which refers to the risk of crash involvement of the vehicle in general. And second, the *secondary safety performance*, which is linked to the risk and severity of occupant injury, when the vehicle is involved in a crash. In the following subsections, selected safety performance indicators are explained, some of which having a certain threshold (see overview in Table 23). There are three uses of indicators in the study, namely 1) metrics used for the apriori collision avoidance algorithm that are computed during the simulation runtime, 2) metrics used for the a-posteriori detection and severity estimation of collisions and 3) metrics used for the a-posteriori detection of conflicts or near-misses. While the apriori indicators are the basis for the forward and side collision avoidance system of the ego car (see Section 6.8.3.1), the a-posteriori indicators are computed after each simulation run as safety performance metrics (see sections 6.5.2 and 6.5.3). Note that it is not the goal of this study to estimate the severity of conflicts, but the severity of collisions in terms of estimated injury risk. For the purpose of this thesis, it is sufficient to detect whether there was a conflict or not, in order to compute a conflict probability along with a collision probability. For a complete overview and explanation of available surrogate indicators, refer to Mahmud et al. (2017).

Table 23: List of safety performance indicators selected for the evaluation

Use	Indicator	(Threshold) Value
Collision avoidance (a priori)	Time-To-Intersection (TTI)	Continuous value
	Time-To-Disappear (TTD)	Continuous value
	Time-To-React (TTR)	0.01 s
	Gap acceptance time (GA)	1 s
Collision detection and severity estimation (a posteriori)	Collision detection	1
	Pre-impact velocity v_{imp}	see Eq. (11)-(19)
	Impact angle (α)	n/a
	Delta of velocity Δ_v	see Eq. (23)-(26)
Conflict detection (a posteriori)	Time-To-Collision (TTC)	$TTC < 1.5\,s$
	Time-To-Accident (TTA)	see Figure 72
	Post Encroachment Time (PET)	$PET < 1\,s$
	Longitudinal jerk (j_{long})	$j_{long} < -8\,m/s^3$

The choice of indicators depends on the collision scenario to be evaluated. For example, the PET is used for angle and turning collision scenarios, where both road users' trajectories do not end in the same path and direction, but it is not applicable to rear-end, merging or head-on situations. For those scenarios, the minimum TTC or its derivatives such as the time-integrated TTC are preferred (Mahmud et al., 2017).

Figure 63 depicts a flowchart for the indicator calculation proposed and applied to the demonstration experiment in Section 6.6. When a collision is detected, the aim is to estimate the severity in terms of injury risk. To do so, the impact velocities and the impact angle are calculated to obtain the Δ_v, along with the information which vehicle is the "bullet" vehicle. In case there is no collision detected, the aim is to find out whether the scenario is a critical conflict or an undisturbed situation. There are two decision points to answer this: First, did one of the vehicles perform an evasion manoeuvre, i.e. an emergency braking? Second, did the vehicles come close to each other, even if there was no braking involved. The first question is answered by the maximum jerk value combined with the TTA, while the PET or the TTC thresholds give an indication on the second question, depending on whether there was an angle (side) collision or another collision type. For head-on and rear-end collisions, the PET is replaced by the minimum TTC. If the minimum TTC is below 1.5 seconds, the situation is classified as conflict.

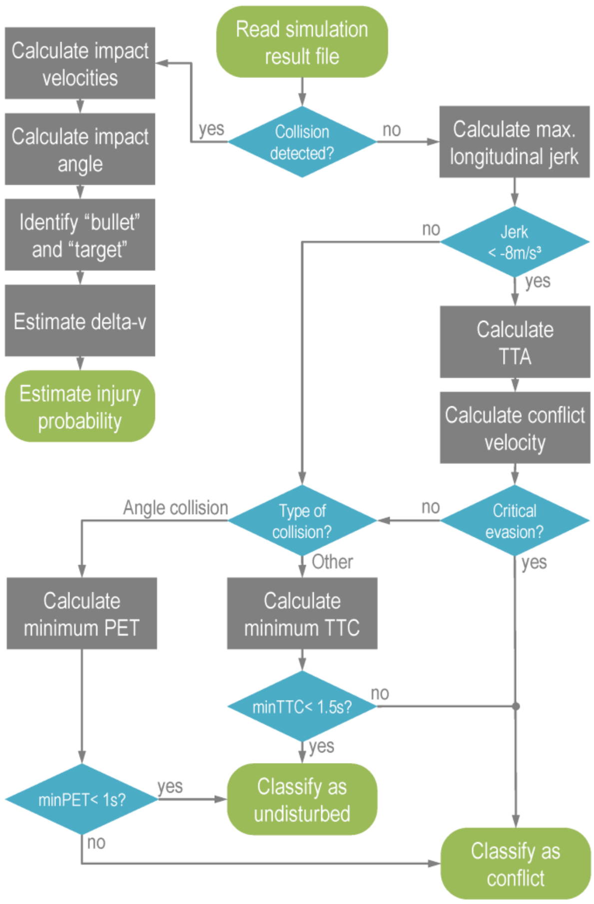


Figure 63: A-posteriori evaluation of the safety performance

The following sections explain how the safety indicators are calculated in detail, divided into collision detection/severity estimation and conflict detection.

Based on the information given above, the **evaluation procedure** consists of the following steps:

1. Import the simulation output files that were stored according to the directory structure defined by the simulation automation script.
2. Iterate through all output files and apply the safety evaluation algorithms.
3. Generate result plots and result files for each iteration (see Section 6.9).
4. Calculate a collision and conflict probability and generate distribution plots by processing all iterations at once.

6.5.2 Collision detection and severity estimation

As mentioned, CarMaker does not simulate the actual collision and deformation of the vehicles. Instead, collision detection is realised by a virtual sensor that delivers a binary output, namely whether the ego car intersects with another object or not. A cuboid or an extruded contour specify the detection area (see Figure 64 and Section 6.6.4 for more details).

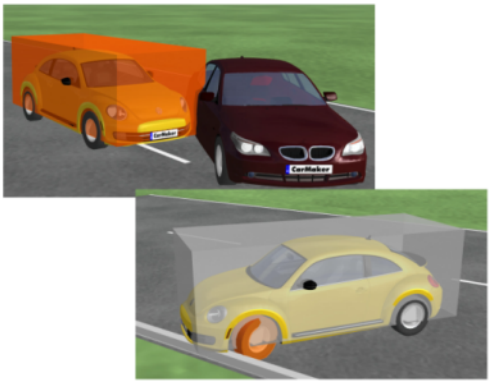


Figure 64: Collision detection in CarMaker with vehicle body and with wheels (IPG Automotive, 2016)

The velocities of the colliding vehicles are typically used to estimate the severity of a collision. One of the simplest indicators is the maximum speed v_{max} observed of either vehicle during a conflict event (Gettman and Head, 2003). This basic approach follows the "speed kills" principle, but does not take into account different collision types, vehicles masses or relative vehicle speeds. Hence, it is not a reliable estimator for injury severity.

Another simple approach is to take the pre-impact velocity v_{imp} as severity indicator, which works well for the injury estimation of vehicle-pedestrian or vehicle-cyclist collisions due to their biomechanical vulnerability (Jurewicz et al., 2015a). However, v_{imp} has its limitations for estimating the severity of vehicle-vehicle collisions, because vehicle masses as well as different collision types and impact angles should be considered. Critical impact speed thresholds for different collision types were published by Jurewicz et al. (2015b), as given in Table 24. The curves are based a study by Bahouth et al. (2014) on historical accident data from the American NASS/CDS database. Note that the values given in the table refer to the “bullet vehicle”, i.e. the vehicle that hits another “target vehicle”. The critical speed values in this table are derived from the injury probability functions depicted in Figure 65 by taking the 10 percent probability as a threshold for the respective collision types.

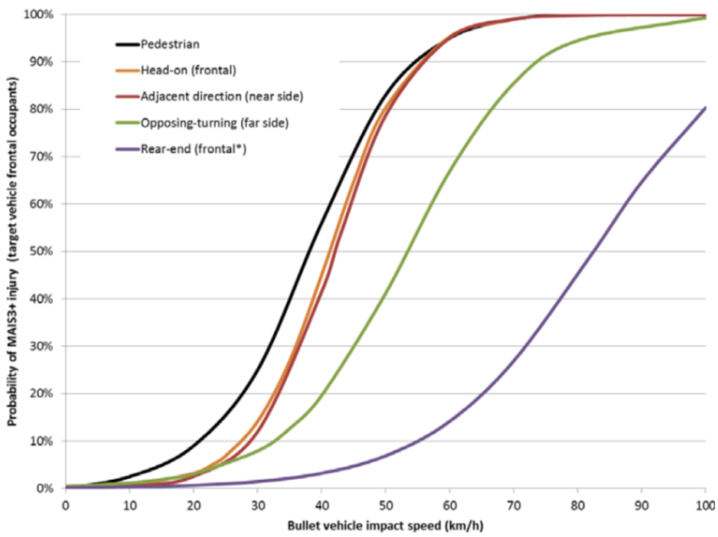


Figure 65: Serious injury risk as a function of v_{imp} (Jurewicz et al., 2015b)

Table 24: Critical pre-impact velocity thresholds for different collision types (Jurewicz et al., 2015b)

Collision type	Critical v_{imp} in km/h
Vehicle-pedestrian	20
Head-on	30
Adjacent direction	30
Opposing-turning	30*
Rear-end	55

** This value may vary, depending on the impact angle and turning vehicle speed.*

The functions are based on the MAIS (Maximum Abbreviated Injury Scale), which is an anatomical-based coding system to classify injury severity (Gennarelli and Wodzin, 2008). The AIS code for an injury has seven digits in the form of 12(34)(56).7, with

- digit 1 for the body region (e.g. 1-Head, 2-Face, 3-Neck etc.),
- digit 2 for the type of anatomical structure (e.g. 1-Whole area, 2-Vessels etc.),
- digits 3 and 4 for the specific anatomical structure (e.g. 02-skin abrasion, 04 contusion, 10-amputation etc.),
- digits 5 and 6 for specific injuries, and
- digit 7 as the severity from 1 to 6 (minor to maximum/fatal) and 9 as not further specified.

It is common in crash studies to use the single MAIS code, which defines the maximum injury severity of all body regions. MAIS1 is therefore defined as minor injury, MAIS3 as serious injury and MAIS6 as currently unsurvivable injury. Note that the underlying data from Bahouth et al. (2014) was taken from accidents in right-hand traffic and that the MAIS3+ probability refers to the injury of the "target" road users that have MAIS3 or higher.

Currently, there is limited research published on the severity estimation of cyclist or motorcycle collisions in relation to the impact speed. Jurewicz et al. (2015a) suggest extending the vehicle-pedestrian value to other vulnerable road users, until there is more evidence available.

It is agreed among experts that the change in velocity of a vehicle during a crash gives superior estimation than v_{imp}. For example, Stoff and Liers (2013) analysed data from 16,000 crashes with injured car occupants from the German GIDAS database ("GIDAS - German In-Depth Accident Study," n.d.) and developed injury risk functions. They found that the change of velocity is the most influencing measure to estimate injury severity, compared to average deceleration, angle velocity, shock pulse and Energy Equivalent Speed (EES). This delta of velocity Δ_v is simply calculated by:

$$\Delta_v = v_{post} - v_{imp} \tag{11}$$

with v_{post} as the post-impact velocity of the vehicle. From a simulation tool like CarMaker, v_{imp} can be easily obtained, but v_{post} is unknown, because the crash impact and vehicle deformations are not physically simulated.

The following analysis considers totally inelastic collisions, where the maximum amount of kinetic energy of a system is lost and the colliding vehicles stick together after the collision with an equal post-impact speed. It is significant to mention that this approximation is simplified, as other important factors such as the part of the vehicles hit, the vehicles' stiffness, the vehicles' rotation or the post-impact rebound are not included (Jurewicz et al., 2015b). However, according to Newton's mechanics, given m_1 and m_2 as the masses of vehicle 1 and vehicle 2, respectively, the momentum of both vehicles is conserved as follows:

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2) v_{post} \tag{12}$$

$$v_{post} = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} \tag{13}$$

For collision, where both vehicles travel in exactly the same direction, such as rear-end crashes, Eq. (11) can be rewritten as

$$\Delta_{v,1} = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} - v_1 = \frac{m_2 v_2 - m_2 v_1}{m_1 + m_2} = \frac{m_2}{m_1 + m_2}(v_2 - v_1) \tag{14}$$

$$\Delta_{v,2} = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} - v_2 = \frac{m_1 v_1 - m_1 v_2}{m_1 + m_2} = \frac{m_1}{m_1 + m_2}(v_1 - v_2) \tag{15}$$

For situations in which the two vehicles travel in opposite directions, such as in head-on collisions, one speed vector is considered negative and the other positive. Thus, attention must be given to the vector nature of momentum and v_2 becomes negative. However, most collisions involve a certain angle between the vehicles. For exact right-angle collisions, the Pythagoras theorem can be applied to compute the resulting vector after the collision:

$$\Delta_{v,1} = \frac{m_2}{m_1 + m_2}\sqrt{v_1^2 + v_2^2} \tag{16}$$

$$\Delta_{v,2} = \frac{m_1}{m_1 + m_2}\sqrt{v_1^2 + v_2^2} \tag{17}$$

Side impacts with a collision angle $\emptyset$ different to 90 degrees require the law of cosines to give the resulting magnitude of post-impact velocity, which changes the formula to:

$$\Delta_{v,1} = \frac{m_2}{m_1 + m_2}\sqrt{v_1^2 + v_2^2 - 2v_1v_2cos\emptyset} \tag{18}$$

$$\Delta_{v,2} = \frac{m_1}{m_1 + m_2}\sqrt{v_1^2 + v_2^2 - 2v_1v_2cos\emptyset} \tag{19}$$

Note again that, if vehicle 2 hit or was being hit by vehicle 1 at a greater angle than 90 degrees, then v_2 becomes negative when used in Eq. (18) and (19). To prove whether the calculated Δ_v is correct, it can be determined if the momentum change experienced by vehicle 1 equals in magnitude and opposite in direction to the momentum change experienced by vehicle 2:

$$m_1\Delta_{v,1} = -m_2\Delta_{v,2} \tag{20}$$

NHTSA uses the WinSMASH computer software to estimate Δ_v based on detailed measurements from the crash scene, vehicle damage and vehicle stiffness characteristics (Sharma et al., 2007). Their algorithm computes the energy absorbed by the vehicle, but also has a missing vehicle algorithm that is used to estimate Δ_v when the damage to one of the vehicles is unknown. The algorithms assume that the impact was instantaneous and that, at some point during the impact, both vehicles reached a common velocity. Due to these assumptions, WinSMASH has its limitations for some collision types, e.g. rollovers, sideswipes, over-ride/under-ride, multiple impacts to the same area, and towed trailer or vehicles.

Laureshyn et al. (2017) extended and operationalised the measure Δ_v within the context of traffic conflict observation. Their adaptation estimates the expected change of velocity experienced by a driver in the event that the conflict would have resulted in a crash. It takes into account both the proximity to a crash and the severity of its potential consequences. It would therefore be a promising candidate for a conflict indicator for the simulation framework of this thesis. However, since it was decided to disregard conflict severity estimation, it was not implemented. Furthermore, the proposed Δ_v extension was not yet sufficiently validated, i.e. further research is needed to show close correlations between conflicts and crashes.

Other studies break the velocity vector into lateral and longitudinal components of Δ_v (Bostrom et al., 2008; Digges et al., 2005; Kullgren, 2008). This has the advantage that side impact injury severity can be estimated more precisely in comparison to rear-end or head-on collisions. For the scope of this study, it is sufficient to estimate an approximate level of injury severity, which can be derived from a single Δ_v vector in combination with the impact angle. Numerous studies were conducted to estimate the relationship between Δ_v and injury severity for different collision types (Berg et al., 1998; Bostrom et al., 2008; Digges et al., 2005; Evans, 1994; Gabauer and Gabler, 2006; Gabler et al., 2005; Hassan et al., 1999; Kullgren, 2008; Meyer, 2016; NHTSA, 2001; Stigson et al., 2012; Thomas and Bradford, 1988; Thomas and Frampton, 1999).

For this simulation study, it was decided to implement the Δ_v calculation as given in Eq. (26) and (27). The following code shows the MATLAB implementation of the data preparation according to the impact angle and which vehicle was the bullet.

```
function [dvA, dvB, vpost, bulletIsA, bulletIsB] = calcDeltaV(safetyObj)

coll_index = find(safetyObj.coll_detect == 1); %check if a collision happened
bulletIsA = [];
bulletIsB = [];


if(isempty(coll_index)) %if no collision happened
    safetyObj.vehc_A_deltav = -1;
    safetyObj.vehc_B_deltav = -1;
else
    %if A entered the conflict zone first, then B is the bullet vehicle
    if safetyObj.t_enter_A_real <= safetyObj.t_enter_B_real
        bulletIsB = 1;
        bulletIsA = 0;
        %In case of rear-end case or angle < 90 deg:
        if safetyObj.impact_angle<90
            m1 = safetyObj.vehc_B_mass; %Vehicle mass
            m2 = safetyObj.vehc_A_mass; %Vehicle mass
            v1 = safetyObj.vehc_B_vimp; %Vehicle impact speed
            v2 = safetyObj.vehc_A_vimp; %Vehicle impact speed
            [dvB, dvA, vpost] = deltav(m1, m2, v1, v2, safetyObj.impact_angle);
        else
            %all other cases:
            m1 = safetyObj.vehc_A_mass; %Vehicle mass
            m2 = safetyObj.vehc_B_mass; %Vehicle mass
            v1 = safetyObj.vehc_A_vimp; %Vehicle impact speed
            v2 = safetyObj.vehc_B_vimp; %Vehicle impact speed
            [dvA, dvB, vpost] = deltav(m1, m2, v1, v2, safetyObj.impact_angle);
        end
    else
        %if B entered the conflict zone first, then A is the bullet vehicle
        bulletIsB = 0;
        bulletIsA = 1;
        %same for all cases:
        m1 = safetyObj.vehc_A_mass;
        m2 = safetyObj.vehc_B_mass;
        v1 = safetyObj.vehc_A_vimp;
        v2 = safetyObj.vehc_B_vimp;
        [dvA, dvB, vpost] = deltav(m1, m2, v1, v2, safetyObj.impact_angle);
    end
```

```
    safetyObj.vehc_A_deltav = dvA;
    safetyObj.vehc_B_deltav = dvB;
end

end
```

The following code shows the implementation of the actual Δ_v calculation:

```
function [dv1, dv2, vpost] = deltav(m1, m2, v1, v2, alpha)
%Calculates the delta-v for a totally inelastic collision


dv1 = (m2/(m1+m2))*sqrt(v1^2 + v2^2 - 2*v1*v2*cosd(alpha));
dv2 = (m1/(m1+m2))*sqrt(v1^2 + v2^2 - 2*v1*v2*cosd(alpha));

if alpha > 90 && alpha<=180
    v2 = -v2; %change speed vector direction for post impact speed calculation
end


vpost = (m1*v1+m2*v2)/(m1+m2); %Post-impact speed magnitude for both vehicles


%proof 1: Is the momentum change experienced by vehicle 1 equal in magnitude and
opposite in direction to the momentum change experienced by vehicle 2?
dv1 = -dv1;
tol = 0.0001; % A small value to set a tolerance for the equality check
equal = ismembertol(m1*dv1, -m2*dv2, tol); %are the momentums near to equal?
if ~equal
    error('Momentum changes not equal and opposite in direction');
end


%Go on with absolute values for injury severity estimation:
dv1 = abs(dv1);
dv2 = abs(dv2);


%proof 2: The heavier vehicle must have a lower delta-v
lower = [];
if m1>m2
    lower = dv1 < dv2;
elseif m1<m2
    lower = dv2 < dv1;
end
if ~lower && ~isempty(lower)
    error('Heavier vehicle has a higher delta-v.');
end

end
```

In one of the earlier studies, Evans (1994) investigated the injury and fatality risk for car-car accidents based on Δ_v in miles per hour. They analysed over 14,000 crashes from an American accident database from the years 1982 to 1991, and fit a generalized functional form to data for both injury prediction and fatality prediction. For belted occupants, the probability function of fatality risk resulted in:

$$P_{fatal} = \left(\frac{\Delta_v}{69.2}\right)^{4.57} \tag{21}$$

Equivalently, the probability function of injury risk is:

$$P_{injury} = \left(\frac{\Delta_v}{67.4}\right)^{2.62} \tag{22}$$

Gabauer and Gabler (2006) fitted binary logistic regression models to available data from in-vehicle event data recorders (EDR) to obtain a probability function for serious injury risk (MAIS 3+) based on Δ_v. The model results are given in Figure 66, without distinguishing into belted and unbelted occupants. A limitation of their study is that it solely includes frontal collisions and cannot be extrapolated to all collision modalities. For example, lateral velocity information would allow a more sophisticated prediction of injury risk for a broader set of crash modes.

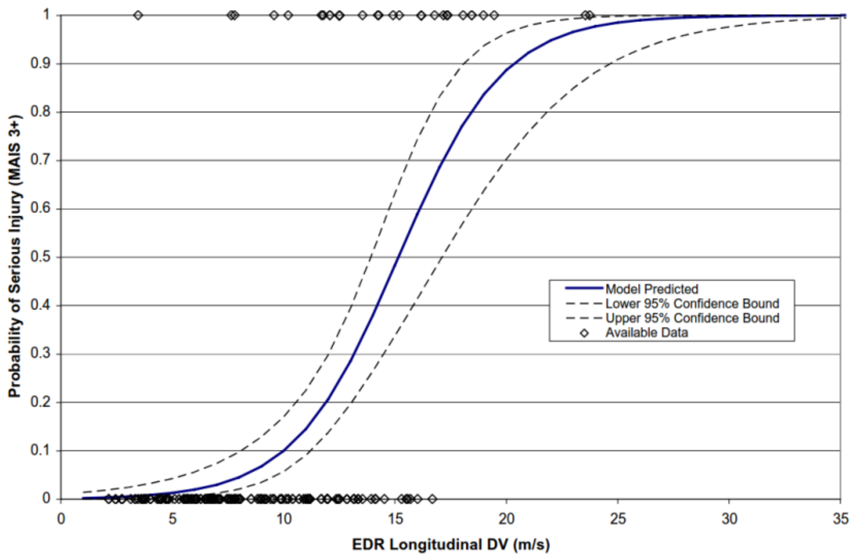


Figure 66: Serious injury risk as a function of Δ_v (Gabauer and Gabler, 2006)

In a study by NHTSA (2001), NASS-CDS[5] data from 1995 to 1999 was examined to derive a relationship between Δ_v and the probability of occupant injury along all different MAIS levels, as a basis for estimating the impact of improved stopping distance on vehicle safety. The percent probability risk of each MAIS injury level at each Δ_v was defined as the number of MAIS injury divided by the total number of occupants involved at a certain Δ_v. The risk prediction curves were derived using a mathematical modelling process with Δ_v as the independent variable and the probability risk as the dependent variable. The probability for MAIS 0 (no injury) was calculated as:

$$P_{MAIS\,0} = \begin{cases} 100 \cdot e^{-0.0807 \cdot \Delta_v}, \Delta_v \leq 35 \\ 0,\ \Delta_v \geq 36 \end{cases} \tag{23}$$

The probability for slight injury (MAIS 1+) was derived as follows:

[5] The American NASS (National Automotive Sampling System) is composed of two systems - the Crashworthiness Data System (CDS) and the General Estimates System (GES). Both are based on police crash reports, but CDS focuses on passenger vehicle crashes to investigate injury mechanics.

$$P_{MAIS\ 1+} = \begin{cases} 93.221 \cdot sin(0.0449 \cdot \Delta_v), \Delta_v \leq 35 \\ 100,\ \Delta_v \geq 36 \end{cases} \tag{24}$$

The serious injury probability (MAIS 3+) was calculated by:

$$P_{MAIS\ 3+} = 100 \cdot \frac{e^{0.1292 \cdot \Delta_v - 5.5337}}{1 + e^{0.1292 \cdot \Delta_v - 5.5337}} \tag{25}$$

Equivalently, a fatal injury (MAIS 6) has the following probability risk function:

$$P_{MAIS\ 6} = 100 \cdot \frac{e^{0.1524 \cdot \Delta_v - 8.2629}}{1 + e^{0.1524 \cdot \Delta_v - 8.2629}} \tag{26}$$

Similar to the study by Gabauer and Gabler (2006), these formulas have to be used with care, because they might not be applicable to all types of collisions. Furthermore, the study included only accidents, where at least one vehicle used the brakes. Another limitation is that the underlying accidents stem from the years 1995 to 1999. Considering that modern vehicle technologies and safety systems might have changed the overall injury probability in relation to the velocity, those functions could be replaced by more up-to-date data in the future. However, for this simulation study, their results provide a good estimation of injury probability, distinguished into no, slight, serious and fatal injury. In MATLAB, the injury estimation was implemented as follows:

```
function [P_inj] = calcPMAIS(safetyObj, vehc)

P_inj = zeros(7,1);
if ~isempty(vehc)
    if vehc == 'A'
        dv = safetyObj.vehc_A_deltav*3.6/1.6; %convert m/s to miles/hour
    elseif vehc == 'B'
        dv = safetyObj.vehc_B_deltav*3.6/1.6; %convert m/s to miles/hour
    else
        disp('Warning: P(injury) cannot be calculated. Please specify a vehicle.');
    end
    if dv<=35
        P_inj(1) = 100*exp(-0.0807*dv); %no injury
        P_inj(2) = 93.221*sin(0.0449*dv); %MAIS1+
    elseif dv>=36
        P_inj(2) = 100; %MAIS1+
    end
    P_inj(3) = 100*exp(0.1683*dv-5.0345)/(1+exp(0.1683*dv-5.0345)); %MAIS2+
    P_inj(4) = 100*exp(0.1292*dv-5.5337)/(1+exp(0.1292*dv-5.5337)); %MAIS3+
    P_inj(5) = 100*exp(0.1471*dv-7.3675)/(1+exp(0.1471*dv-7.3675)); %MAIS4+
    P_inj(6) = 100*exp(0.1516*dv-7.8345)/(1+exp(0.1516*dv-7.8345)); %MAIS5+
    P_inj(7) = 100*exp(0.1524*dv-8.2629)/(1+exp(0.1524*dv-8.2629)); %MAIS6+, FATAL
else
    disp('Warning: P(injury) cannot be calculated. Please specify a vehicle.');
end

end
```

6.5.3 Conflict detection

Apart from collisions, the proposed simulation framework is able to identify conflicts and near-misses. The following sections explain the conflict indicators used.

6.5.3.1 Time-To-Collision and conflict hexagon

A popular indicator to distinguish from critical to normal driving behaviour is the *Time to Collision (TTC),* which is used for traffic conflict analysis and collision avoidance systems. Hayward (1972) defined the TTC as the time required for two vehicles to collide if they were to continue their speed and path. Hence, it is a measure on how imminent a collision is. It is further a continuous parameter, which may be computed at every timestamp of a simulation or real-world observation. It was originally introduced for rear-end situations with two moving vehicles, where the TTC is calculated according to the following equation (Minderhoud and Bovy, 2001):

$$TTC_2(t) = \frac{X_2(t) - X_1(t) - l_1}{v_1(t) - v_2(t)} \;\; \forall\, v_2(t) > v_1(t) \tag{27}$$

with t as the current timestamp, X_1 and X_2 denoting the position of vehicle 1 (ahead) and 2 (following), v_1 and v_2 as the speed of vehicles and l_1 as the length of vehicle 1 (see Figure 67a). For the case of a potential head-on collision, the formular can be modified as follows (see Figure 67b):

$$TTC_2(t) = \frac{X_2(t) - X_1(t)}{v_1(t) + v_2(t)} \tag{28}$$

Computing the TTC for crossing path or turning against situations is not trivial, since the trajectories of the road users have to be predicted in order to find a *conflict point*. This means that a real-world implementation is complex and suffers prediction inaccuracy, either in a vehicle system for real-time collision avoidance, in a roadside intersection assistance system or as part of a retroactive traffic conflict analysis with roadside sensors and cameras.

Van der Horst (1990) calculates the TTC for angle collisions by taking into account the area around the conflict point instead of a single conflict point, which is further referred to as *conflict zone* (see Figure 67c). Presuming that the size of the conflict zone is based on the vehicle dimensions, the TTC is calculated as follows:

$$\begin{cases} TTC = \dfrac{d_2}{v_2}, if \; \dfrac{d_1}{v_1} < \dfrac{d_2}{v_2} < \dfrac{d_1 + l_1 + w_2}{v_1} \\ TTC = \dfrac{d_1}{v_1}, if \; \dfrac{d_2}{v_2} < \dfrac{d_1}{v_1} < \dfrac{d_2 + l_2 + w_1}{v_2} \end{cases} \tag{29}$$

with d_1 and d_2 as the distances from the fronts of the vehicles 1 and 2, respectively, to the beginning of the intersection area, v_1 and v_2 as the vehicle speeds, l_1 and l_2 as the vehicle lengths and w_1 and w_2 as the vehicle widths. The latter fraction terms are further denoted as *Time-To-Disappear* (TTD), as they define the expected duration until the road user will have left the conflict zone, if its velocity does not change. Equation (29) can therefore be simplified to:

$$\begin{cases} TTC = TTC_2, if\ TTC_1 < TTC_2 < TTD_1 \\ TTC = TTC_1, if\ TTC_2 < TTC_1 < TTD_2 \end{cases} \tag{30}$$

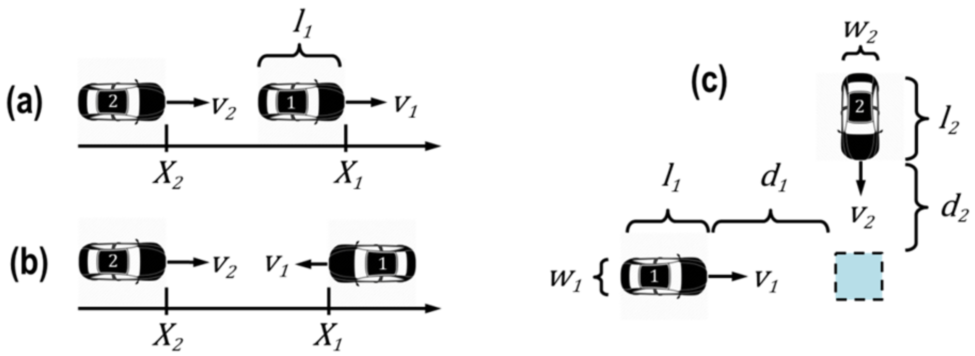


Figure 67: Calculation of TTC for (a) rear-end collisions, (b) head-on collisions and (c) angle collisions (adapted from Laureshyn et al., 2010)

In MATLAB, Equation (29) was implemented as follows, taking the values for *s_enter_A/B_theor* and *s_exit_A/B_theor* as inputs that correspond to the timestamp where the vehicles will theoretically enter or exit the conflict zone:

```
%Only distances and velocities are taken into computation:
TTC_A = (safetyObj.vehc_A_xCord_theor(safetyObj.s_enter_A_theor) - ...
safetyObj.vehc_A_xCord) ./ safetyObj.vA;
TTC_B = (safetyObj.vehc_B_xCord_theor(safetyObj.s_enter_B_theor) - ...
safetyObj.vehc_B_xCord) ./ safetyObj.vB;

%continuous expected time to disappear (TTD)
TTD_A = (safetyObj.vehc_A_xCord_theor(safetyObj.s_exit_A_theor) - ...
safetyObj.vehc_A_xCord) ./ safetyObj.vA;
TTD_B = (safetyObj.vehc_B_xCord_theor(safetyObj.s_exit_B_theor) - ...
safetyObj.vehc_B_xCord) ./ safetyObj.vB;

TTC_A(TTC_A<0)=0;
TTC_B(TTC_B<0)=0;
TTD_A(TTD_A<0)=0;
TTD_B(TTD_B<0)=0;

%Apply Van der Horst's formula (get only critical TTC for either A or B):
TTC_A_crit = TTC_A .* (TTC_A < TTD_B) .* (TTC_B < TTC_A);
TTC_B_crit = TTC_B .* (TTC_B < TTD_A) .* (TTC_A < TTC_B);
```

The drawback of the formula by Van der Horst (1990) is that it is only applicable to right angle collisions in its original form. Miller and Huang (2002) developed a different calculation method to obtain TTC for varying collision angles as essential part of their proposed car to car collision warning system. In their algorithm, the future path intersection is estimated first by using both vehicles' heading and locations in the global coordinate system. After the point of intersection is determined, the expected Time-To-Intersection TTX is estimated according to the velocities of both vehicles and the distance between the current position and the projected intersection point. Once a vehicle has cleared the intersection, its TTX becomes negative and there is no route contention. If there is a route contention, i.e. if $TTX_1$is close to TTX_2, then the TTX equals the TTC. The closeness of the TTX values can be defined by a so-called contention parameter, which takes into account the vehicle dimensions and uncertainties of velocity or heading angle. However, their approach does not consider exact vehicle dimensions and is not applicable to changing angles such as when driving a curve.

Laureshyn et al. (2010) presented another method to calculate the TTC for potential angle collisions. Their study included different collision types, since the vehicles may approach each other with different angles. Therefore, they introduced a point and line principle, assuming that the vehicles are rectangles and that it is always the corner (point) of one of the vehicles that hits the side (line) of the other vehicle. Their approach overcomes the problem that the conventional geometry-based definitions, when the vehicle trajectories do not cross at a right angle. Consequently, the conflict zone shape may change from a rectangle to a parallelogram, and the entrance and exit from the zone are no longer time moments but periods. Both vehicles can appear in the same zone without collision.

A TTC calculation method for the collision type "Right turn against" (assuming left-hand driving) was investigated by Sobhani et al. (2013), which elaborates on the principle of Laureshyn et al. (2010), but additionally includes the curvature function of the turning vehicle in the formula. For the particular application of pedestrian safety, Salamati et al. (2012) proposed an adaptated TTC formula, which allows the calculation for different lanes in a multi-lane road configuration. The conflict zone in their method is the crosswalk that might cross more than one lane.

In summary, it can be stated that for each collision type, a different TTC method must be applied. Hence, also the threshold values to measure the severity of a conflict vary accordingly. A table of published TTC thresholds reviewed from literature can be found in the survey of Mahmud et al. (2017). The minimum TTC lies between 0 and 2 seconds, but some papers also state a "desired" TTC ranging 0.9 to 4 seconds. In general, a TTC lower than the perception and reaction time should be considered critical. A fixed threshold of $TTC < 1.5s$ is typically chosen to distinguish between slight conflicts and serious conflicts, i.e. near-misses (Hydèn, 1987; Shbeeb, 2000).

In the proposed framework of this thesis, the conflict zone (as depicted in Figure 67c) is determined as part of the evaluation script, i.e. by analysing the outputs of the simulation run. The algorithm combines the approaches from Van der Horst (1990), Miller and Huang (2002) and Laureshyn et al. (2010) and takes into account the driving directions, curvature as well as the vehicles' dimensions and velocities. This is done by the following steps:

1. The theoretical collision point is calculated by finding the intersection point of both trajectories. Since the trajectories are not lines but points sampled with a frequency of 1 ms, the minimum Euclidean distance between both coordinate vectors is computed to find the closest coordinate pair.

2. For each vehicle path and timestamp, the current distance to the theoretical collision point is calculated.

3. To obtain the conflict entrance point, this distance is reduced by the half width of the conflict zone, which is henceforth denoted as d_{buffer}.
4. Equivalently, the conflict exit point of a vehicle is the distance to the collision point plus d_{buffer}. In this way, the conflict zone can be spanned.

For the case of precisely rectangular collisions (such as in Figure 67c), $d_{buffer} = w/2$. However, in practice the collision angles deviate from 90 degrees, which necessitates another way of calculating d_{buffer}. With other angles, the form of the conflict zone is no more a rectangle. It can be instead defined as *hexagon*, which is illustrated in Figure 68.

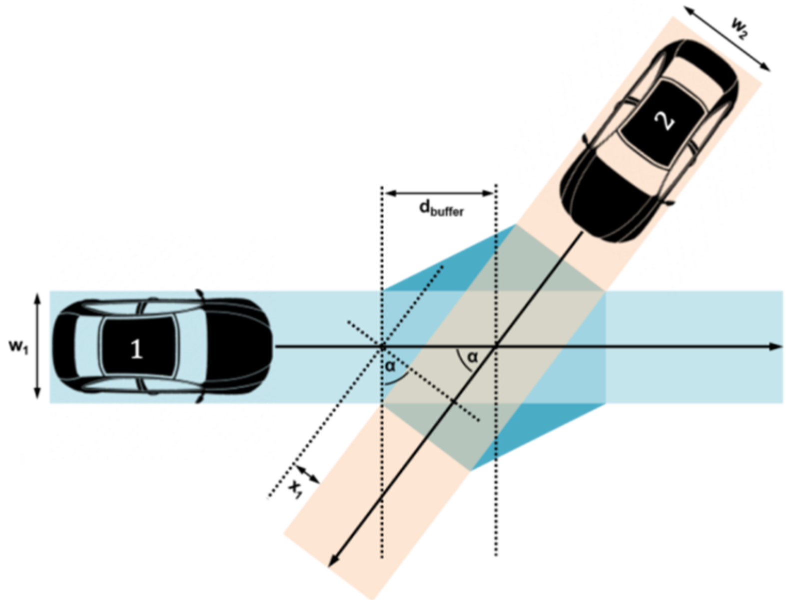


Figure 68: Principle of the conflict hexagon

In theory, the distance d_{buffer} can be determined by solving a trigonometric problem. Considering w_2 as width of vehicle 2 and α as the angle of the trajectory intersection, the buffer for vehicle 1 is calculated as follows:

$$d_{buffer} = \frac{x_1 + \frac{w_2}{2}}{sin(\alpha)} \tag{31}$$

Two triangles lead to the solution. Equation (31) gives the hypotenuse of the larger triangle with a given opposite leg, which is the sum of $w_2/2$ and x_1. The length of x_1 is the adjacent leg of the smaller triangle with a given hypotenuse of $w_1/2$ and the same angle α:

$$x_1 = cos(\alpha) \cdot \frac{w_1}{2} \tag{32}$$

By substituting x_1 in Equation (31), the buffer length for vehicle 1 can be rewritten as follows, henceforth named "the needle and pin formula"[6]:

$$d_{buffer} = \frac{w_2 + w_1 \cdot cos(\alpha)}{2 \cdot sin(\alpha)} \tag{33}$$

As the vehicle widths do not change within a simulation scenario, only the intersection angle varies according to the expected trajectories of the vehicles. Hence, the shape of the *conflict hexagon* changes with the angle, as Figure 69 depicts. Equation (33) can be proven by taking an α of 90 degrees, which results in a buffer of $w_2/2$ and therefore a conflict rectangle. However, note that this principle is only valid for straight trajectories.

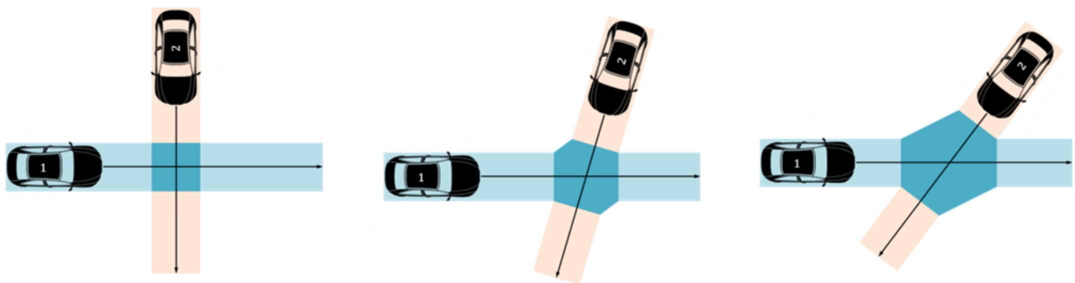


Figure 69: Different shapes of the conflict hexagon depending on the collision angle

Equation (33) necessitates the theoretical collision angle α, which can be calculated by

$$\alpha = arctan\left(\left|\frac{m_1 - m_2}{1 + m_1 m_2}\right|\right) \tag{34}$$

with m_1 and m_2 as the slopes of the two longitudinal centre line of vehicle 1 and 2, respectively (see Figure 70):

$$m_1 = \frac{Y_{1,exit} - Y_{1,enter}}{X_{1,exit} - X_{1,enter}} \tag{35}$$

$$m_2 = \frac{Y_{2,exit} - Y_{2,enter}}{X_{2,exit} - X_{2,enter}} \tag{36}$$

The principle of the angle calculation in Equation (34) is illustrated in Figure 70, showing the two trajectories as well as the two auxiliary lines to compute the slope.

[6] The trigonometric problem of the conflict hexagon, i.e. the formula to compute the buffer distances, was solved in a pub called „The Needle and Pin".

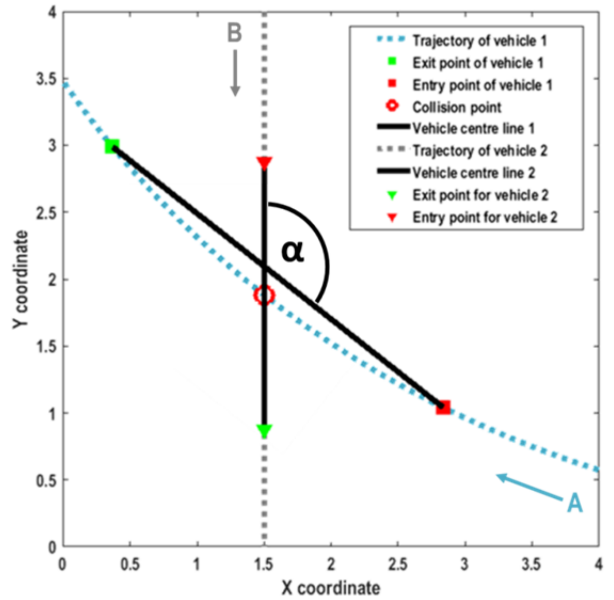


Figure 70: Illustration of the collision angle calculation

As mentioned earlier, the calculation has to be enhanced in the case of curved trajectories by creating a tangent on the curves at the point of intersection to obtain the angle. This still leads to an imprecise calculation of d_{buffer} especially at very low curve angles, because the distance w/2 is computed from the tangent and not from the curve. Therefore, another approach was chosen, which makes use of the fact that from the simulation outputs the precise trajectories of both vehicles are known and that the distances and heading angles are available for each timestamp. The trajectories of both vehicles are not only represented by vectors, but by "trajectory bands" spanned by the respective vehicle width. This necessitates the calculation of laterally displaced trajectory curves parallel to each other. Additionally, the tractrix curve of the turning vehicle must be taken into account to calculate the size of the conflict hexagon. This is realized by using two trajectory bands, namely one for the front of the vehicle and one for the rear.

For a situation as depicted in Figure 71, where one vehicle turns right and the other crosses the path from the right, the entrance and exit points of the zone are the intersections of the left-side trajectories of both vehicles or, vice versa, the intersection of the right-side trajectories. For example, the entry point for the turning vehicle is the intersection of its left trajectory from the front with the left trajectory of the approaching vehicle. Its exit point is the intersection of its right trajectory from the rear with the left trajectory of the approaching vehicle. The distance d_{buffer} is then determined by orthogonally projecting these points to the centre (front trajectory) line of the respective vehicle. In this way, the conflict hexagon is created.

It is important to mention that the hexagon entrance and exit points are used for both the TTC and the PET calculation, which is explained in Section 6.5.3.4. Both indicators are applied to assess the closeness of the conflicting vehicles, but the difference can be explained as follows: The PET is a single value calculated by the difference between the entrance and exit timestamps and considers braking manoeuvres that might have happened. In contrast, the TTC is a continuous value along the timeline and works as a virtual warning system, where the future trajectory is supposed to be unknown. The time until the conflict entrance (TTC) and exit (TTD) are theoretical assumptions given that the velocity of the vehicles will stay the same (see Eq. (29)).

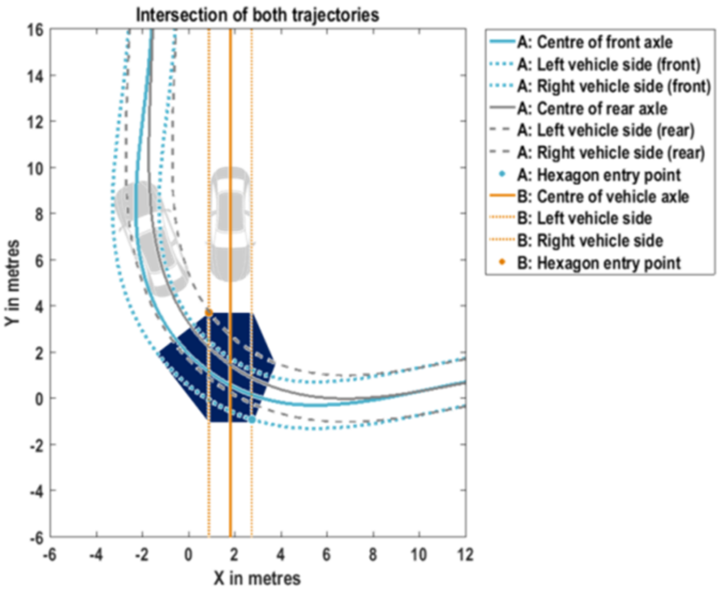


Figure 71: Principle of the conflict hexagon for curved trajectories

The MATLAB code to calculate the parallel curves and the hexagon is shown below:

```
v_threshold = 0.5; %Value in m/s, under which the trajectory coordinates are cropped

sA_real = safetyObj.vehc_A_xCord;
sB_real = safetyObj.vehc_B_xCord;
posA_real_fr = [safetyObj.vehc_A_Global_xCord_front
safetyObj.vehc_A_Global_yCord_front];
posA_real = [safetyObj.vehc_A_Global_xCord safetyObj.vehc_A_Global_yCord];
posB_real = [safetyObj.vehc_B_Global_xCord safetyObj.vehc_B_Global_yCord];

% Generate inner and outer parallels:
vAStill = find(vA<v_threshold);
posA_real(vAStill,:) = []; %remove samples while car is standing still
posA_real_fr(vAStill,:) = []; %remove samples while car is standing still

vBStill = find(vB<v_threshold);
posB_real(vBStill,:) = []; %remove samples while car is standing still

%Front trajectories (for A and B):
[posAx_inner, posAy_inner, posAx_outer, posAy_outer, ~, ~, ~, ~]= ...
parallel_curve(posA_real_fr2(:,1),posA_real_fr2(:,2),safetyObj.vehc_A_wid/2 + ...
corrBuffer,0,0);
[posBx_inner, posBy_inner, posBx_outer, posBy_outer, ~, ~, ~,
~]=parallel_curve(posB_real(:,1),posB_real(:,2),safetyObj.vehc_B_wid/2 +
corrBuffer,0,0);
posB_inner = [posBx_inner posBy_inner];
posA_inner = [posAx_inner posAy_inner];
posB_outer = [posBx_outer posBy_outer];

%Rear trajectories (only for A, because B is going straight):
```

```
[posAx_innerb, posAy_innerb, posAx_outerb, posAy_outerb, ~, ~, ~, ]= ...
parallel_curve(posA_real2(:,1),posA_real2(:,2),safetyObj.vehc_A_wid/2  + ...
corrBuffer,0,0);
posA_outerb = [posAx_outerb posAy_outerb];

[sA_enter, sB_enter, sA_exit, sB_exit, sA_enter_hex, sB_enter_hex] = ...
ComputeConflictHexagon( posA_inner, posB_inner, posA_outerb, posB_outer, ...
safetyObj.vehc_A_collIndex, safetyObj.vehc_B_collIndex, sB_real, ...
safetyObj.vehc_B_len, vAStill, vBStill );
safetyObj.s_exit_A_real = sA_exit;
safetyObj.s_exit_B_real = sB_exit;
safetyObj.s_enter_A_real = sA_enter;
safetyObj.s_enter_B_real = sB_enter;

%% Get real entry timestamps:
safetyObj.t_enter_A_real = safetyObj.time(safetyObj.s_enter_A_real);
safetyObj.t_enter_B_real = safetyObj.time(safetyObj.s_enter_B_real);

%% Get real exit timestamps:
safetyObj.t_exit_A_real = safetyObj.time(safetyObj.s_exit_A_real);
safetyObj.t_exit_B_real = safetyObj.time(safetyObj.s_exit_B_real);
```

In more detail, the function for the conflict hexagon calculation was implemented as follows:

```
function [ sA_enter, sB_enter, sA_exit, sB_exit, sA_enter_hex, sB_enter_hex ] = ...
ComputeConflictHexagon( posA_inner, posB_inner, posA_outerb, posB_outer, ...
collIndexA, collIndexB, sB_real, vehc_B_len, vAStill, vBStill )

if isempty(vBStill)
    vBStill = 0;
end
if isempty(vAStill)
    vAStill = 0;
end

%% Find index of hexagon entry point coordinates (intersection of the real inner
front trajectory of A and the real inner of B):
%in case the trajectories cross more than once: Take only the distance from the
theoretical collision point
if collIndexA>size(posA_inner,1)
    collIndexA = size(posA_inner,1);
end
%Calculate distance:
[~, minDistIndA, minDistIndB, ~, ~, ~, ~] =
GetDistance(posA_inner(1:collIndexA,:),posB_inner);
sA_enter_hex = minDistIndA(1); %Needed for plotting purposes
sA_enter = minDistIndA(1)+length(vAStill);
sB_exit = minDistIndB(1);

%% Find index of hexagon exit point coordinates (intersection of the real inner rear
trajectory of A and the inner of B):
%in case the trajectories cross more than once: Take only the
%distance around the theoretical collision point for B
cropLenB = 3000; %crop for better performance. Adapt for other simulation data
if cropLenB>collIndexB
    cropLenB = collIndexB-1;
end
cropLenA = 5000; %crop for better performance. Adapt for other simulation data
if cropLenA+collIndexA>size(posA_outerb,1)
    cropLenA = size(posA_outerb,1)-collIndexA;
end
%Calculate distance:
[~, minDistIndA, minDistIndB, ~, ~, ~, ~] = ...
GetDistance(posA_outerb(collIndexA:collIndexA+cropLenA,:),posB_outer(collIndexB- ...
cropLenB:collIndexB,:));
sA_exit = minDistIndA(1) + length(vAStill) + collIndexA-1; %add crop length
sB_enter_hex = minDistIndB(1) + collIndexB-cropLenB; %Needed for plotting purposes
%Reduce vehicle length to get actual entry point of B:
minDistIndB = minDistIndB(1) + length(vBStill) + collIndexB-cropLenB;
d = sB_real(minDistIndB);
dist = abs(sB_real - (d-vehc_B_len));
```

```
[~,sB_enter] = min(dist);

end
```

6.5.3.2 *Time-To-Accident*

In the proposed simulation framework, the TTC is applied to the cases where an evasion action, e.g. a braking manoeuvre, was taken by either of the road users. From the timestamp, where a critical braking manoeuvre was detected, the TTC is computed a posteriori. In literature, this application of TTC is also denoted *Time-To-Accident (TTA)* (Hydèn, 1987). Accordingly, the TTA is defined as "the time that remains to an accident from the moment that one of the road users starts an evasive action if they had continued with unchanged speeds and directions" (Svensson, 1998). The TTA can be related to the conflict speed in order to define conflict severity levels, as depicted in Figure 72. The conflict speed is defined as the vehicle's velocity at the moment the evasion action is taken. In comparison to the traditional approach with a fixed TTC/TTA threshold, it can be seen that e.g. a TTA of 1.5 seconds indicates a serious conflict at a speed of 40 km/h, but it is considered not severe at a speed of 30 km/h. Hence, the detection of a conflict is velocity-dependent, which is more reliable than a simple threshold.

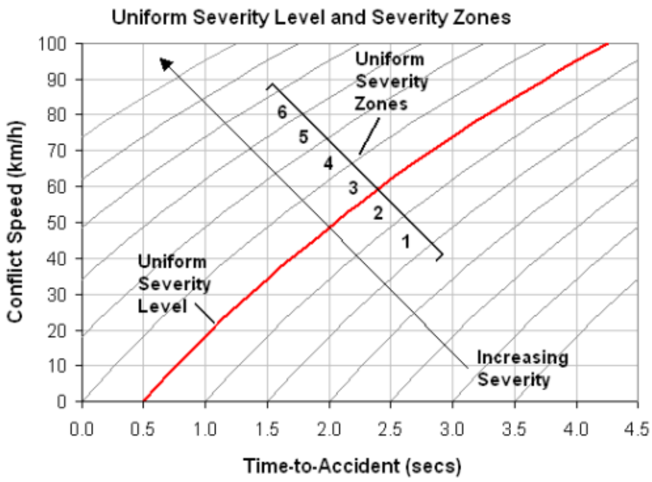


Figure 72: Severity levels based on TTA and conflict speed (based on Hydèn, 1987)

In MATLAB, the TTA is calculated as follows, taking the TTC values and the braking time (*t_brakeA/B*) as inputs.

```
dt=1000; % Time between two samples
%Conflict indicator based on TTA and conflict speed
xtta_orig = 0:0.5:4.5;
%Function based on Hyden (1987)
yvc_orig = [0 0 17.6 34.8 48.4 61.6 73.6 84.4 95.6 105]./3.6;
xtta = 0:0.01:xtta_orig(end); %new sampling
yvc = interp1(xtta_orig,yvc_orig,xtta); %interpolated function

%For A:
if ~isempty(t_brakeA)
    %do not calculate TTA if TTC is not critical
    if ismember(t_brakeA*dt-cropBegin, find(indTTC_A))
```

```
        TTA_A = TTC_A(t_brakeA*dt-cropBegin);
        dist = abs(xtta - TTA_A);
        [~, minIndex] = min(dist);
        vc_threshold_A = yvc(minIndex);
    else
        TTA_A = [];
        vc_threshold_A = -1;
    end
else
    TTA_A = [];
    vc_threshold_A = -1;
end
%For B:
if ~isempty(t_brakeB)
    %do not calculate TTA if TTC is not critical
    if ismember(t_brakeB*dt-cropBegin, find(indTTC_B))
        TTA_B = TTC_B(t_brakeB*dt-cropBegin);
        dist = abs(xtta - TTA_B);
        [~, minIndex] = min(dist);
        vc_threshold_B = yvc(minIndex);
    else
        TTA_B = [];
        vc_threshold_B = -1;
    end
else
    TTA_B = [];
    vc_threshold_B = -1;
end
```

6.5.3.3 Longitudinal jerk

When the driver model applies the brakes, it does not necessarily have to be a critical situation, since the driver models may decelerate when approaching an intersection. To distinguish critical braking events from non-critical ones, another indicator is applied to the proposed framework, namely the *longitudinal jerk* j_{lon}, which gives the derivative of the longitudinal deceleration when braking:

$$j_{lon} = \frac{\partial a_{lon}}{\partial t} \tag{37}$$

with a_{lon} as the longitudinal acceleration (negative in the case of braking) differentiated over time t. In comparison to the deceleration rate, which can also be found in various papers about traffic conflict analysis, the jerk helps to better differentiate sudden braking manoeuvres from other decelerations, which are not necessarily conducted to avoid a conflict (Zaki et al., 2014). (Wåhlberg, 2000) first investigated the relation of the composite g-force and speed to traffic accident frequency, for the case of city busses. With the rise of probe vehicle data and naturalistic driving studies, the g-force has become a common indicator to identify safety-critical events. While the *PET* does not consider evasion manoeuvres such as emergency brakings, the jerk is particularly used for those. The higher the negative jerk, the more critical is the situation. Zaki et al. (2014) found that the majority of vehicles at a selected intersection reacted to conflicts within a jerk range of -8 m/s³ and -10 m/s³. Although the relationship was found statistically not significant, vehicles with jerk values greater than 8 m/s³ tended to be involved in conflicts with $TTC < 2s$.

A common practice to obtain a smooth jerk profile is to apply a moving average filter to avoid outliers in the derivative values, which might be mechanically unrealistic

(Bagdadi and Várhelyi, 2013; Feng et al., 2017; Zaki et al., 2014). This is realised by applying a second-order Savitzky-Golay filter with a 1 s time windows to the acceleration signal in MATLAB. The numerical differentiation of the smoothed acceleration signal results in the jerk profile. If this profile includes a negative jerk greater than a given threshold, then the braking manoeuvre is considered critical and a conflict is given. Based on the study by Zaki et al. (2014), a threshold value of -8 m/s³ is chosen. The MATLAB function to calculate the jerk vector is given in the following:

```
function [Jerk, ax_sm] = calcJerk(safetyObj, ax, vehc)
  %Smooth ax:
  ax_sm = sgolayfilt(ax,2,999);
  %Differentiate smoothed ax to get the jerk:
  Jerk = diff(ax_sm)./diff(safetyObj.time(1:end-1));
  if vehc == ‚A‘
    safetyObj.vehc_A_jerk = Jerk;
  elseif vehc == ‚B‘
    safetyObj.vehc_B_jerk = Jerk;
  end
end
```

6.5.3.4 *Post Encroachment Time*

As a variation to the TTC, the *Post Encroachment Time (PET)* is an indicator that measures the temporal difference between two road users over a common spatial area (Archer, 2005), i.e. a conflict zone. In comparison to the TTC, it requires transversal trajectories of the road users, hence it is commonly used to estimate the severity of a junction conflict. In its original form, it is not applicable to situations where both road users' trajectories end in the same path and direction. For example, merging conflicts would require another indicator, such as the gap acceptance. However, the road users do not necessarily have to be on a collision course. The PET is defined as the time difference between a road user 1 exiting the conflict zone and road user 2 entering it:

$$PET = t_2 - t_1 \tag{38}$$

with t_1 as the time when the first road user exits the conflict zone and t_2 as the time when the second road user enters it. The smaller the PET, the higher is a potential conflict. Consequently, a negative value indicates that both road users are within the conflict zone. Figure 73 depicts the PET calculation on a junction example. In literature, common threshold values lie between 1 s and 1.5 s (Archer, 2005). For example, if $t_1 = 10s$ and $t_2 = 12s$, then the second vehicle has reached the conflict zone 2 seconds later than the first vehicle left it, i.e. the conflict zone was clear for 2 seconds and the situation could therefore be considered as safe.

There are different approaches for specifying the size of the conflict zone, which obviously affects the minimum PET threshold. In the example in Figure 73, the zone size corresponds to the vehicle dimensions, which implies a smaller "safety buffer" than specifying the conflict zone according to the junction size. The conflict zone can also be set to the whole junction area, which is not recommended for large junctions, or to the actual area of potential intersection, as proposed by this thesis, which requires the analysis of vehicle trajectories (see Section 6.5.3.1). Pirdavani et al. (2010) divided the junction into four equally sized areas to compute the PET separately. While commonly rectangular or squared areas are used, Killi and Vedagiri (2014) recommend a rhombus geometry further segmented into smaller grids. For the demonstration study explained in Section 6.6, it was decided to follow the conflict hexagon principle of the TTC calculation and to set the conflict zone according to the vehicle dimensions (see Section 6.5.3.1). The minimum PET is set to a value of one second, similar to Ismail et al. (2009).

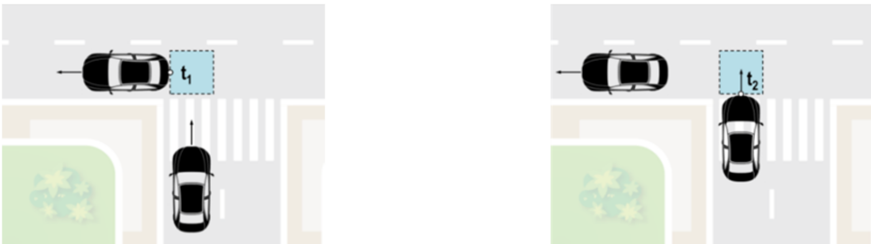


Figure 73: Illustration of the PET at t_1 (left) and t_2 (right)

Apart from the fact that the PET necessitates transversal trajectories, a common drawback of the PET is that it does not require velocities or distances for its calculation. This limits the enhancement of these indicators, e.g. by different speed levels, and the possibility to compare the relative severity of different post encroachment times (Archer, 2005). The difference between TTC and PET can be seen in Figure 74. It depicts a situation where vehicle 1 (blue line) and vehice 2 (grey line) are on a projected collision course, i.e. they are within the conflict zone at the same time. When vehicle 1 starts braking, the actual trajectory deviates and vehicle 1 enters the conflict zone after vehicle 2 has left it. This time difference is the PET. While the TTC gives the time until the projected arrival at the conflict zone (and equivalently, the TTD the time until the projected exit of the conflict zone), the PET is based on the actual timestamps measured retroactively. The PET is implemented as a separate MATLAB function as follows, taking the enter and exit timestamps for both vehicle class objects:

```
function [PET, t2, t1] = calcPET(safetyObj)
    %if A comes later than B, i.e. B exits the conflict zone first
    if safetyObj.t_exit_A_real >= safetyObj.t_exit_B_real
        t1 = safetyObj.t_exit_B_real; %B exiting with the rear
        t2 = safetyObj.t_enter_A_real; %A entering
    else
        t1 = safetyObj.t_exit_A_real; %A exiting with the rear
        t2 = safetyObj.t_enter_B_real; %B entering
    end
    PET = t2-t1;
end
```

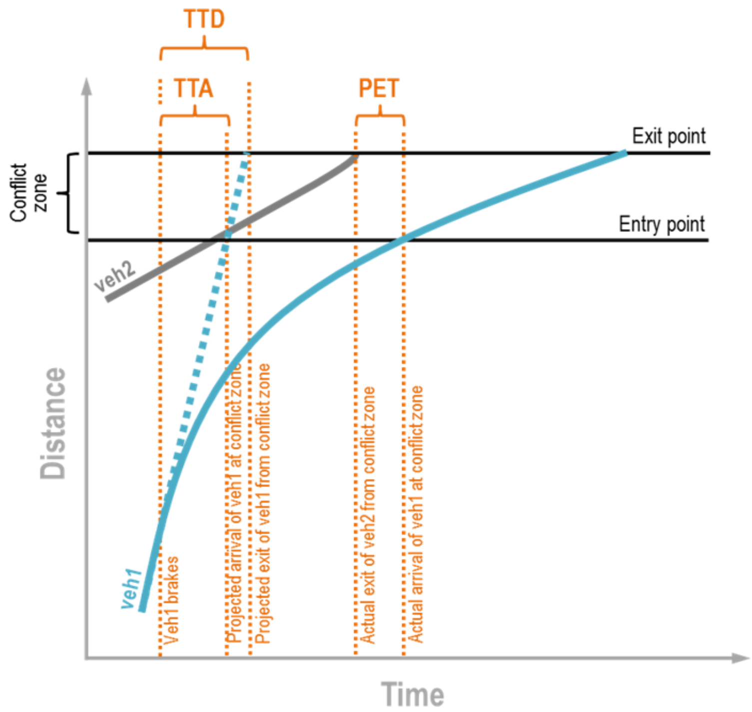


Figure 74: Illustration of *TTA*, *TTD* and *PET* in the time-distance diagram

6.6 Description of simulation models

CarMaker offers a GUI to parametrise the virtual vehicle environment and to start a simulation. As mentioned above, a so-called test run is created manually for each scenery, before the simulation automation script sets up the desired parameter variations and executes the simulation instance. In CarMaker, a test run consists of the modules Vehicle, Road, Manoeuvre (driver's tasks) and Driver, without which a simulation cannot be started. In addition, the user can define the modules Trailer, Tires, Traffic (other static or moving objects) and Environment (time and ambient conditions). All those modules have predefined models that can be parametrised by the user. The MATLAB/Simulink interface enables the user to use the signals that flow between the CarMaker models, either to read them for further calculations or to modify them. The following sections describe the CarMaker modules used for the proposed simulation framework (IPG Automotive, 2017).

6.6.1 Road environment modelling

The CarMaker Road module is a software library that defines the static road environment. The library can be used to build a road network that consists of several *nodes* connected by *links*, which can be further broken down into *lane sections*. A lane section can include several *lanes*. A brief definition of the different terms used in CarMaker is given in Table 25.

Besides the basic elements listed in the table, additional obstacles can be added such as bumps, beams, curbs or freely configurable pavement overlays. An interesting element to mention is the friction parameter, which can be set for a specific link. A change of friction can be realised by adding another link or by adding friction patches, e.g. to simulate worn pavement sections or aquaplaning.

Table 25: Terminology of the Road module in CarMaker

Name	Description
Link	Section of the road network, started and terminated by a node
Junction	Road element to smoothly join different links (up to 8)
Lane section	Longitudinal section of a link with a constant number of lanes
Lane	Lateral section of a lane section (up to 10), for which the lane width and the type (driving lane, road border, roadside, bicycle lane, pedestrian path, traffic island, parking area, bus lane, HOV lane or emergency lane) can be defined individually.
Elevation profile	Profile changes in lateral and longitudinal direction
Reference line	Road curvature, i.e. theoretical course along a specific link
Route	Reference line defining the course of the vehicle travelling along consecutive links (the route equals the reference line of the respective link)
Path	Actual course of the vehicle travelling on the road, taking into account corner cutting and driving on a specific lane

Accessories that can be added to a road include road markings, road paintings in the form of texture files, traffic signs, traffic signals, traffic barriers and guideposts. Additional scenery attributes are bridges, tunnels, roadside signs, trees and bushes as well as different geometry objects such as houses, street furniture, animals and people.

CarMaker provides a GUI called Scenario Editor[7] that allows the user to drag and drop road elements and to modify their parameters. Road sceneries created with the Scenario editor are stored as part of the test run, but can also be exported as individual files in the format ".rd5". Figure 75 depicts a screenshot of the Scenario Editor. While the left tab provides all the tools to build a road by adding road attributes, the right tab is used to parametrise different aspects of a selected road attribute.

As part of the road configuration, the start positions of the vehicles can be specified, along with their longitudinal and lateral offset, their orientation and their route. However, detailed manoeuvres are defined in the Manoeuvre module, as the following section explains.

[7] The term Scenario Editor might be misleading at this point, since this thesis defines a scenario as the movements of dynamic traffic objects within a scenery. Strictly speaking, it could be called a Scenery Editor.

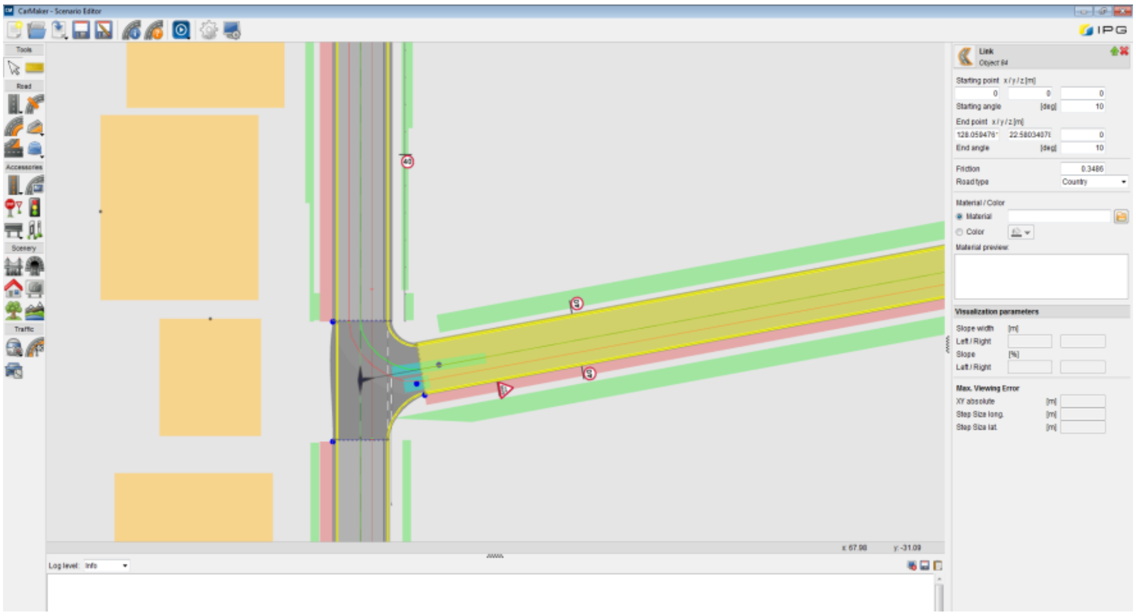


Figure 75: Screenshot of CarMaker's Scenario Editor

6.6.2 Manoeuvre modelling

In order to build a driving scenario, the user has to add and parametrise manoeuvres in CarMaker. A manoeuvre can consist of (1) longitudinal dynamics actions such as accelerating, braking or gear shifting, (2) lateral dynamics action such as steering and (3) additional actions defined by a list of special mini-manoeuvre commands using a script language. A manoeuvre may include several manoeuvre steps, each of which consists of a duration and a description of the driver's task. Each manoeuvre step must have an end condition, which can either be the time when the duration is reached, a distance covered or other user-defined end conditions. An alternative way to end a manoeuvre is to add a road marker at a given point on the road, which triggers the end of the current manoeuvre step.

The *longitudinal manoeuvres* can be specified in different ways. To accelerate and brake the vehicle according to its dynamic limits, the option IPGDriver is the default and recommended way (see Section 6.6.5). However, another option is to import real-world vehicle speed measurements to obtain a speed profile that the driver model must follow. Contrary to the IPGDriver module, a simple speed controller can be specified, which accelerates to or keep the desired speed, but cannot brake the vehicle. Alternatively, the position of the pedals (clutch, gas, brake) can be controlled manually along with the gear.

Also for the *lateral manoeuvres*, the IPGDriver module can be used to steer the vehicle according to its dynamics limits. There are alternative options such as a sinusoid input to the steering wheel or a manual change of the steering angle, which are however not applied to this study.

6.6.3 Vehicle modelling

CarMaker's Vehicle Data Set library provides detailed parametrisation of the ego vehicle, ranging from vehicle body characteristics to vehicle components such as engine, suspension, steering, tires, brakes or powertrain. Sensors are also part of the library, but are explained separately in the following section. If a real reference vehicle was available, the library would allow to create a validated virtual vehicle that replicates the real vehicle's behaviour to a great extent. If the user does not have access to such detailed information about the vehicle or if the test does not require a validated virtual vehicle, the Vehicle Data Set Generator can be used. This was done for this study. It creates an entire vehicle data set based on limited information, e.g. vehicle class, driving axle, vehicle dimensions and weight (see Figure 76 left). Additional settings include simple steering, engine, transmission, suspension and aerodynamics parameters as given in Figure 76 (right).

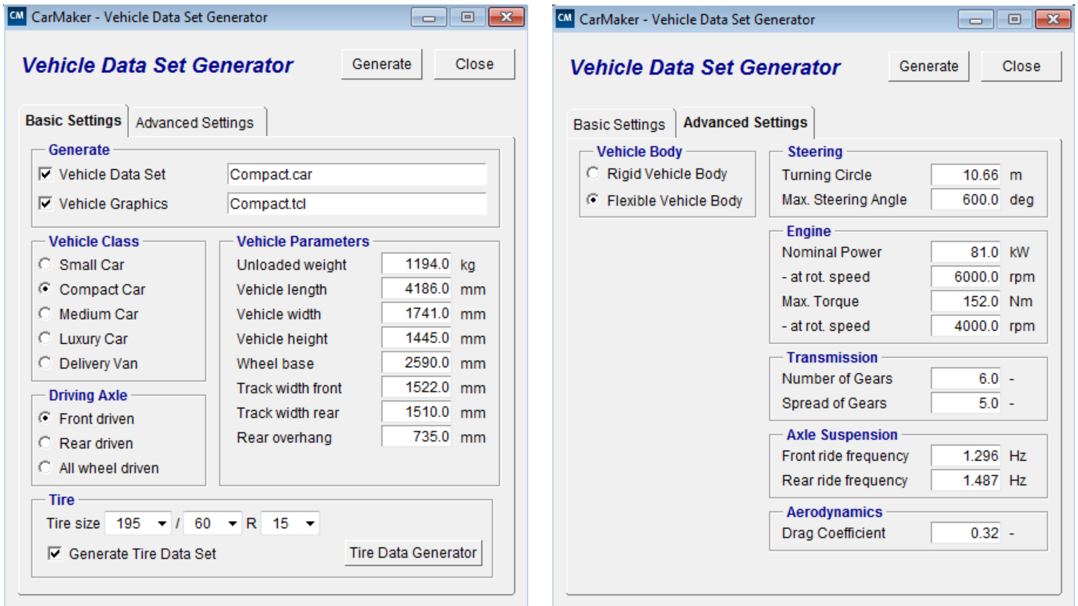


Figure 76: Screenshots of the basic (left) and advanced (right) parameters in CarMaker's Vehicle Data Set Generator

To understand the vehicle modelling, CarMaker's coordinate axis system (called frames) is essential to know. The software uses different axes depending on which object of the vehicle is specified. Figure 77 depicts that the virtual world coordinate frame is named *Fr0*, while *Fr1* denotes the frame of moving objects within the virtual world. The frame *Fr1* performs all movements of the object such as translations and rotations. Note that *X* points in forward driving direction, *Y* points to the left and *Z* is directed upwards. The origin *O* of the vehicle frame is located at the rear of the vehicle on the ground, and no part of the outer skin of the object is situated behind the (*O-Y-Z*)-plane. This is important to know when evaluating safety parameters such as TTC or PET. For every wheel, there is a mounting point *Mnt* located within the *Fr1* frame, which is the origin of the frame *Fr2* and hence the centre of the wheel. Again,

X points towards the driving direction, Y points to the left of the wheel spin axis and Z is directed upwards.

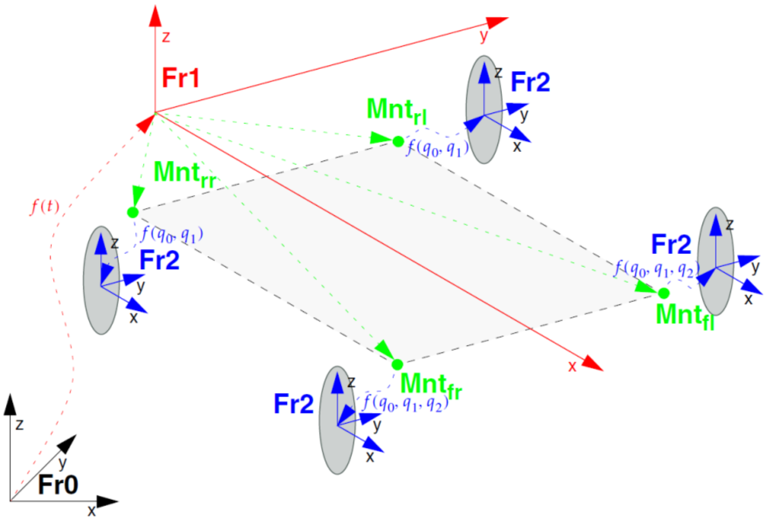


Figure 77: CarMaker's coordinate systems (IPG Automotive, 2017)

In the presented study, the default values created by the Vehicle Data Set Generator were sufficient to evaluate automated driving functionality. The basic vehicle parameters were set in a way to replicate a Tesla Model S for demonstration purposes (see Section 6.8.3.1). The most important parameters are related to the environment sensors, which are described in the following section.

6.6.4 Vehicle sensor modelling

In CarMaker, several vehicle sensors can be placed on the car to replicate safety systems and environment perception for ADAS. Table 26 gives an overview on the sensor models available in CarMaker, which can be distinguished into three groups of physical models.

First, *ideal* models replicate sensors for rapid prototyping or proof of concept that provide ground truth information concerning environment and object detection. They deliver a list of objects detected including all relevant information needed from those objects such as type, position, velocity and size.

Second, the *high-fidelity* sensors are useful for more advanced function development and testing. They replicate the physical boundaries and effects that real-world sensors encounter, such as reduced propagation, multi-path effects or erroneous signals. Currently, CarMaker provides high-fidelity sensors for radar and GNSS. Similar to the ideal sensors, they deliver an object list, but this might vary depending on the sensor characteristics.

Third, the *raw signal interface* sensors provide raw signals, such as the output of a video camera, instead of an object list. This allows the development and testing of signal processing and object tracking algorithms.

Table 26: Overview of CarMaker sensor models

Application	Physical model	Sensor name	Description
Vehicle dynamics	Ideal	Slip Angle Sensor	Measures a vehicle's side slip angle
	Ideal	Inertial Sensor	Measures inertial body movements such as velocity, acceleration and rotation
ADAS	Ideal	Object Sensor	Detects traffic objects and gives information
	Ideal	Free Space Sensor	Detects free and occupied spaces between traffic objects, equivalent to stereo video
	Ideal	Free Space Sensor Plus	Detects free and occupied spaces in the whole environment
	Ideal	Traffic Sign Sensor	Detects traffic signs along the road
	Ideal	Line Sensor*	Detects road markings such as lanes as well as guard rails, barriers etc.
	Ideal	Road Sensor	Provides road information as electronic horizon
	Ideal	Collision Sensor	Detects contacts of the test car with other traffic objects
	High Fidelity	Radar Sensor	Detects traffic objects based on the signal-to-noise ratio by a realistic radar replication
	High Fidelity	Global Navigation Sensor	Simulates GNSS satellites and their visibility for the vehicle receiver
	Raw Signal Interface	Camera RSI	Camera sensor providing picture information using IPGMovie

For this study, the vehicle dynamics sensors are disregarded and a selection of ADAS sensors is used, namely various object sensors, which are further explained in the following paragraphs. Those sensors were placed on the vehicle model in a way to replicate the sensing system of a Tesla Model S (see Section 6.8.3.1).

An *object sensor* works as an ideal sensor that detects all traffic objects within its range and field of view. The *free space sensor* is an extended module to the object sensor, as its sensor beams are subdivided into horizontal and vertical segments to scan the environment. Similar to the object sensor, the range and angle of view can be configured to mimic e.g. stereo-video or LIDAR sensors and the number of segments can be specified according to the desired level of resolution. Each segment of the free space sensor detects the nearest point of the surrounding traffic objects and delivers the detected objects' coordinates, approaching velocity and relative distance to the ego vehicle. Figure 78 (left) shows the GUI to set up the free space sensors, while Figure 78 (right) visualises the segment-based detection with green segments to highlight detected objects and grey segments to show the range of the sensor.

With CarMaker's new release version 6, the so-called *Free Space Sensor Plus* was made available for purchase, which is a high-fidelity sensor and further subdivides each segment into horizontal and vertical rays. Unlike the free space sensor, this sensor cannot only detect traffic objects, but all objects of the environment within the specified range and field of view. It thus represents a more realistic physical sensor model where the complete 3D environment can be reconstructed, similar to LIDAR. Due to the increased computational requirements, this sensor is calculated using the GPU. However, in the current release, the Free Space Sensor Plus can only be used in the CarMaker’s standalone GUI, but not in the Simulink environment. That is why this study uses the conventional object sensors instead, because the framework is based on Simulink.

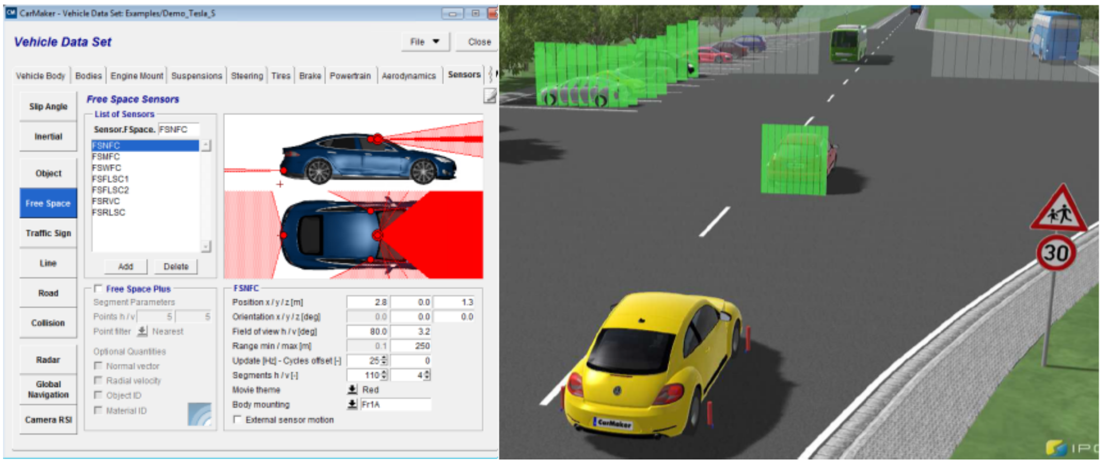


Figure 78: CarMaker’s free-space sensor (left: GUI for sensor parametrization, right: visualization of segments) (IPG Automotive, 2017)

The *traffic sign sensor* is used to detect pre-defined traffic signs (as road accessory) within its range and field of view. It outputs the total number of detected signs, the distance to each detected sign, the traffic sign type as well as the attribute values such as the speed limit. It verifies if the signs are facing to the sensor and sorts all detected signs by ascending distance. The traffic sign sensor can be used to influence the driver model behaviour, e.g. by placing a stop sign or reducing the speed limit.

The *line sensor* detects road markings and traffic barriers within its range and acts like an idealised camera. The sensor delivers all markings and barriers with ascending lateral distance to a specified point in the vehicle frame. Besides the lateral distance, the sensor provides the colour code, type, width and height of the detected marking or barrier. The line sensor adapts to the route the vehicle is supposed to take. Figure 79 depicts how the line sensor visualises its detection results (left) and in which way the sensor works (right). It creates a shape with seven points on the road (orange shape), which is determined by the intersecting plane of the road surface with five vertical and three horizontal areas. Within this shape, the road markings and barriers are detected.

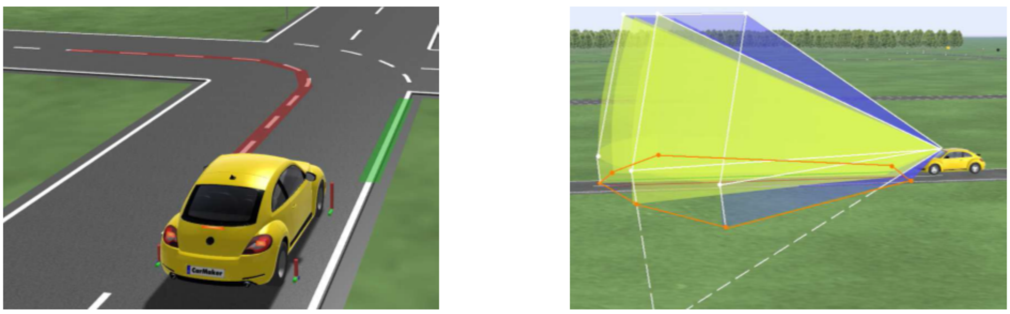


Figure 79: Illustration of CarMaker’s line sensor (IPG Automotive, 2017)

The *collision sensor* simply detects if the ego vehicle body or wheels gets in contact to another traffic object. Therefore, a cuboid or a user-defined contour defines the vehicles' envelope. The sensor then returns a counter for the number of detected collisions of the vehicle body or the wheels.

The *radar sensor* is implemented as a high-fidelity sensor, which means that it is able to consider occlusion effects and propagation losses. The object detection is affected by signal propagation latency and noise, but also by the radar cross sections defined for each object, which depends on the direction of incidence, statistical fluctuations, object occlusion and merging of objects (see example in Figure 80). Similar to real radar, an object is detected, if the signal-to-noise ratio exceeds a certain threshold. CarMaker provides a detailed GUI for modifying the characteristics of the antenna (gain map) to imitate specific radar sensor systems. Ultimately, the radar sensor model delivers the position, velocity, acceleration and relative course angle of detected objects, along with the objects' dynamics mode (standing, stopped, moving or oncoming), dimensions and probability of existence as well as a probability of the object being an obstacle on the path of the ego vehicle. Besides, the detected object ID is provided as a ground truth information.

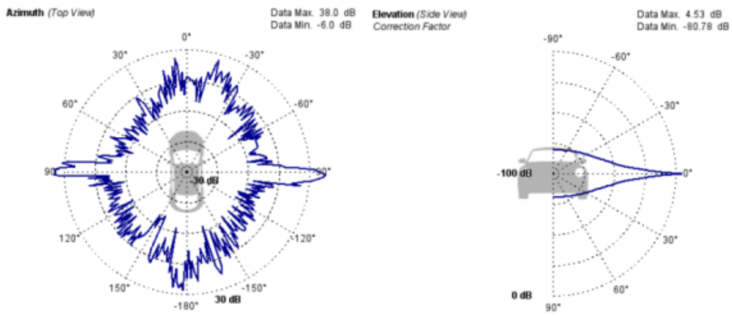


Figure 80: Example of a radar cross section (left: top view, right: side view) (IPG Automotive, 2017)

6.6.5 Driving behaviour modelling

A requirement for the driver model is that it behaves similar to a human in terms of acceleration, deceleration, steering and distance keeping. Researchers argue that it would increase the acceptability of assisted and automated driving functions if they behave human-like (Bifulco et al., 2008). The IPG Driver module enables the replication of realistic driving behaviour of the ego car. It consists of a controller for following a course and a speed controller on a given track, including the following actions:

- Choice of course within lane boundaries
- Choice of driving speed according to the course and vehicle behaviour
- Influence on speed by gas and brake pedal as well as by clutch operation and gear shifting
- Steering to stay within the lane boundaries

In the *initialisation phase*, i.e. before the simulation starts, the driver model calculates a static desired course and speed profile along the predefined road. This initial calculation allows that all marginal requirements, which are independent of the vehicle dynamics, are computed in advance. The course is determined by a spline approximation using the road lane information and the so-called corner cutting coefficient *ccc* (see Figure 81). The corner cutting is calculated by an optimisation algorithm that adapts the weights of the spline approximation. A corner cutting coefficient of one means that the total lane width is used to drive the curve considering the vehicle width, while a value of zero follows precisely the lane centre, i.e. the given course in the road model, which may not reflect realistic driving behaviour.

The second initialization step is the calculation of the static speed along the computed course, which can be seen as the intended velocity the longitudinal control model aims to maintain. Therefore, a kinematics model takes the cruising speed as well as the maximum lateral and longitudinal accelerations as input to determine a speed profile along the entire distance of the ego vehicle path.

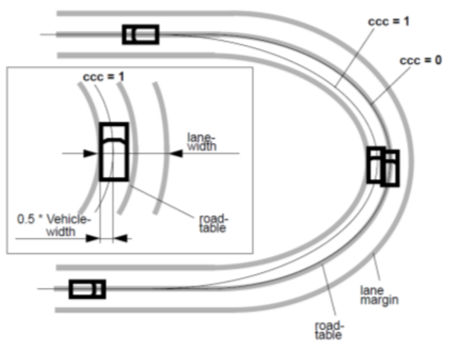


Figure 81: Principle of the corner cutting coefficient (IPG Automotive, 2017)

The desired static course and speed from the initialisation phase, along with the vehicle state (current vehicle motion including velocity, accelerations, side slip angle etc.) and the steering wheel torque, are the inputs for the driver model during the *simulation phase*. The output comprises the pedal positions (gas/brake/clutch), the gear shifter position and the steering wheel angle. The actions during the simulation can be grouped into steering and influence on speed (see Figure 82).

The *steering control* continuously recognises the motion and vehicle dynamics of the ego car and predicts the future course, which might deviate from the desired course computed in the initialisation phase. The distance, for which the future course is predicted, depends on the preview time, which is continuously adjusted according to the current vehicle dynamics. If the deviation between the predicted and the desired course exceeds a certain threshold, the driver model interferes by changing the steering behaviour. Besides the threshold, an important parameter for this steering action is the reaction time, which delays the model interference.

The controller to *influence speed* initially starts with desired speed that is suitable for the course and then continuously compares it to the actual speed. By doing so, the desired longitudinal acceleration or deceleration is determined, based on preset parameters such as the maximum longitudinal acceleration and deceleration. User-defined values to specify the dynamic behaviour of the controller include a tolerated speed deviation, reaction time, time to change pedals as well as a prognosis time, which is equivalent to the preview time of the steering controller. The controller outputs are the brake pedal force and the position of the accelerator pedal.

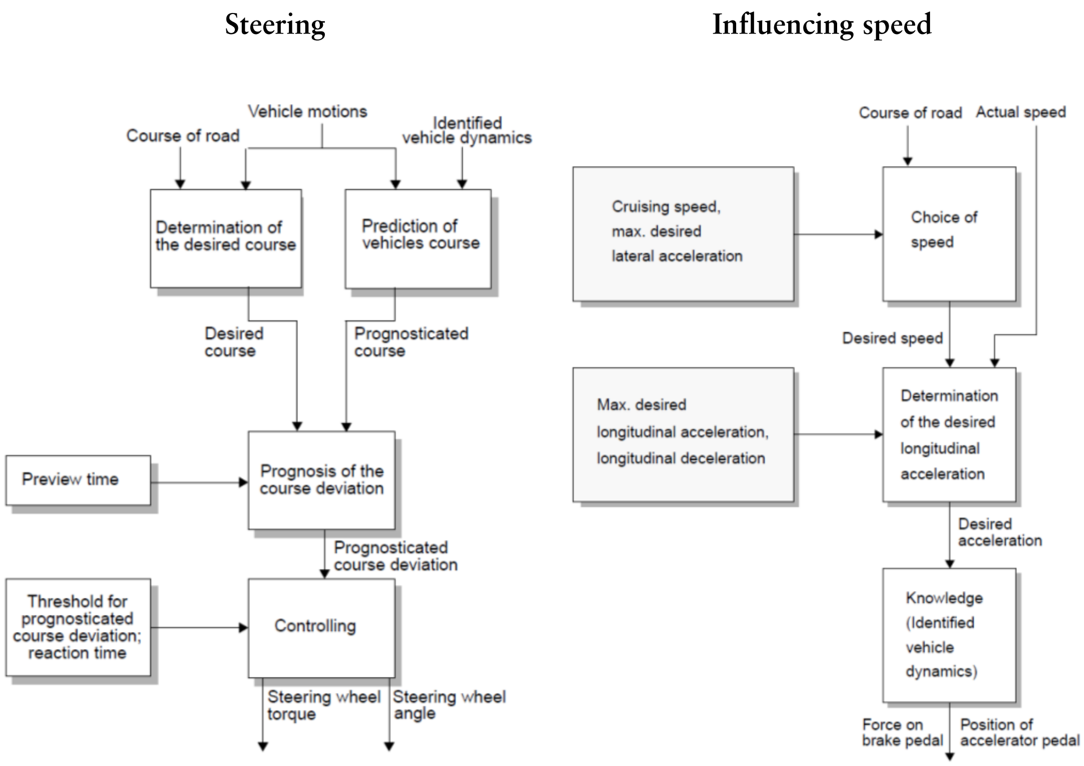


Figure 82: CarMaker's driver module to steering (left) influencing speed (right) (IPG Automotive, 2017)

6.6.6 Traffic modelling

Traffic as it is defined in CarMaker includes all stationary or movable objects other than the ego car. Movable *traffic objects* constitute the dynamic part of the test run, which can be parametrised to influence the simulation, e.g. parking vehicles, vehicles that change the driving lane, car that coincide at a junction, bicycle riders or pedestrians and animals. Stationary objects include e.g. residential or office buildings. Traffic objects are three-dimensional bodies that can have a certain geometry performing closed loop or open loop[8] driving manoeuvres. They can be used to test a driver's reaction or for ADAS validation, since the ego vehicle's sensors are able to detect traffic objects. The main difference of a movable traffic object to the ego vehicle

[8] An open-loop (feedforward) controller is independent of the process output, while a closed-loop (feedback) controller returns the output to control states of a dynamical system.

is that the vehicle dynamics are not calculated, i.e. they constitute much simpler vehicle models.

Basically, a traffic object's motion can be parametrised in two ways: 1) By defining manoeuvres in CarMaker including longitudinal and lateral movements or 2) by defining free motion, i.e. from external sources such as positioning data from real-world measurements. The longitudinal or lateral motion of a manoeuvre can be specified in several ways, e.g. by simply predefining a velocity or acceleration curve or by defining path points, which the traffic object must follow. In the demonstration experiment explained in Section 6.8, the opponent vehicle is modelled as traffic object with a predefined route and manoeuvre, which crosses the path of the ego vehicle. To this end, the option "Autonomous Driving" was chosen, which might lead to confusion given the context of the study. Actually, the traffic objects are supposed to be operated by human drivers, while the ego vehicle imitates an automated car. The option "Autonomous Driving" as function of CarMaker traffic objects models a realistic driving behaviour based on parameters such as maximum velocity, acceleration and deceleration or distance keeping characteristics for the ACC model (see Figure 83).

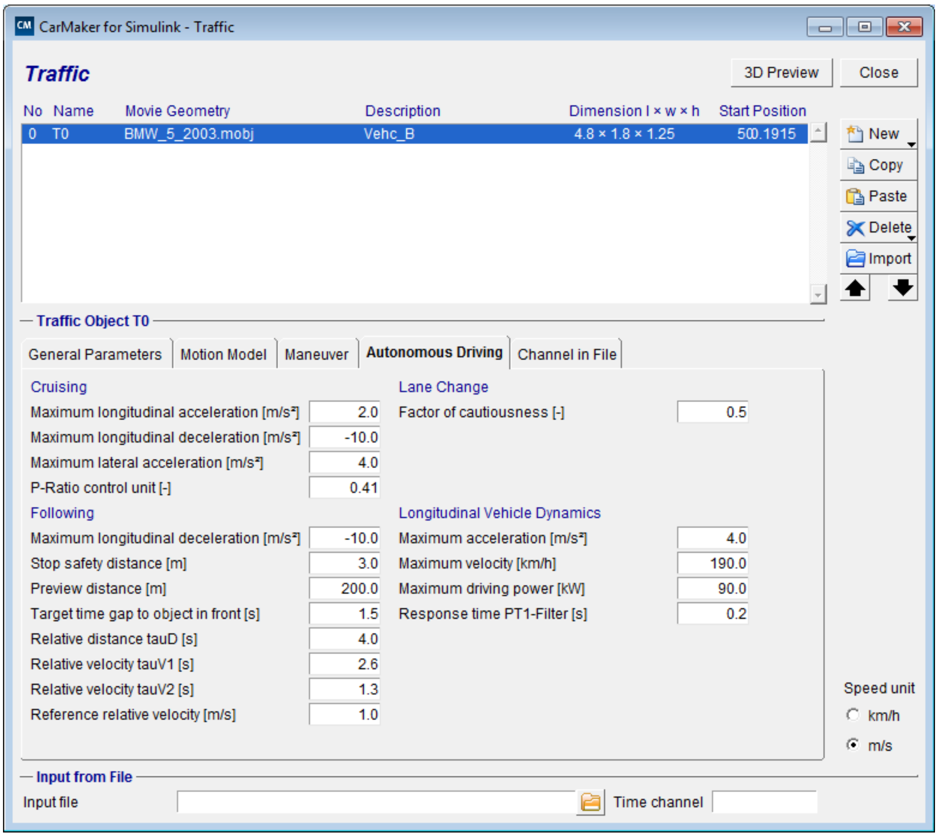


Figure 83: CarMaker's Autonomous Driving parameters for traffic objects

6.6.7 CarMaker/Simulink interface

The presented simulation framework uses the MATLAB/Simulink interface as a layer over CarMaker, to read and modify the signals that flow between the CarMaker models. This is necessary to build new controllers, e.g. the crossing and turning assistant or a smooth deceleration to the junction stop line.

The features of CarMaker are implemented into the Simulink environment using an S-function and the API functions that are provided by Matlab/Simulink. The CarMaker blocks are connected the same way as other Simulink blocks and existing Simulink models can be added to the CarMaker vehicle model. However, using the Simulink interface decreases the simulation performance in comparison to the standalone CarMaker software.

The CarMaker for Simulink simulation model consists of a subsystem containing a chain of individual blocks (see Figure 84). When Simulink executes CarMaker, all blocks of the CarMaker blockchain must be executed exactly once and in order. In order to replace existing CarMaker functionalities, the best way is to override signals transferred between the blocks. The blocks consist of a driving manoeuvre, vehicle control and vehicle model, which can be modified. The block, which is further modified for this study is the vehicle control model. Other functionalities of CarMaker, such as the configuration of traffic objects or sensors are included in the CarMaker simulation library and can be added when needed. The Simulink controllers for the intersection collision avoidance function are explained in the following sections.

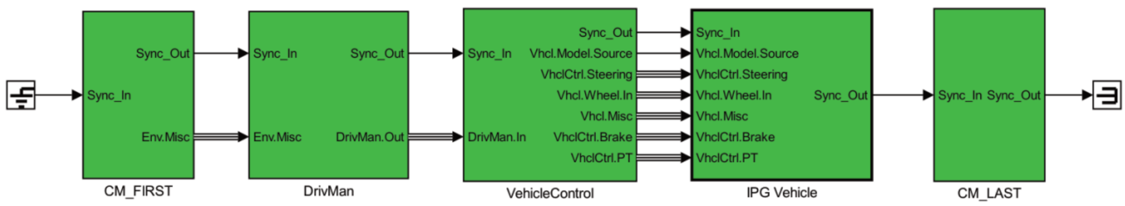


Figure 84: General structure of CarMaker in Simulink

6.6.8 Collision avoidance modelling

Avoiding collisions with other road users is one of the main tasks of an automated driving system. For an intersection situation, the vehicle must be able to predict and avoid potential collisions for a variety of scenarios (see Section 2.4) such as head-on, rear-end, turning or right-angle collisions, referred to as ICAMS in this thesis (see Section 2.4). The proposed simulation and evaluation framework uses two collision avoidance models, which are compared to each other:

1. *Forward collision avoidance (FCA)*, based on forward-facing vehicle sensors, also known as automated emergency brake, to sense other road users in front of the vehicle and to avoid collisions by evasion actions.

2. *Crossing and turning assistance*, based on both forward and sideward-looking sensors, designed to find a safe gap to cross or turn and to avoid side and angle collisions.

6.6.8.1 *Forward collision avoidance*

The forward collision avoidance model is predefined as *Autonomous Emergency Braking* (AEB) in the CarMaker vehicle control. Its task is to decelerate the vehicle to avoid a crash with the target object ahead. For this, the system compares the TTC with a time-threshold-brake to decide if a braking intervention is required. For a stationary or very slow-moving vehicle, the TTC is simply calculated as in Eq. (27). For a non-stationary target object, the formula is extended to:

$$TTC = \frac{\sqrt{v_{rel}^2 - 2 \cdot d \cdot D_{rel}} - v_{rel}}{D_{rel}} \tag{39}$$

with d as the relative distance, v_{rel} as the relative velocity and D_{rel} as the relative deceleration between ego car and target object.

The time-threshold-brake for a non-stationary target object is

$$t_{brake} = \tau_B \cdot \frac{v_{rel} + D_{rel} \cdot \tau_B}{2 \cdot (D_{max} - D_{obs})} \tag{40}$$

with τ_B as the brake loss time, D_{max} as the maximum allowed deceleration of the ego vehicle and D_{obs} as the actual target object deceleration detected by the ego car sensors. If $TTC < t_{brake}$, the AEB system activates a braking manoeuvre by setting D_{max} as target deceleration for the controller.

It must be noted that this AEB controller was designed to avoid forward collisions, primarily in rear-end situations. For the case of crossing and turning collisions, this system is assumed to have significant limitations. Therefore, it is compared to a crossing and turning assistant, which is explained in the following section.

6.6.8.2 *Crossing and turning assistance*

Collisions while crossing and turning are a relevant type of accidents at junctions, which pose a particular challenge for automated vehicles. To avoid those collisions, a crossing and turning assistant must detect approaching road users, predict their speeds and decide whether a safe turning or crossing manoeuvre can be made. There are several types of collisions that the assistant must avoid (see collision code sheet in Appendix B), such as:

- H: Crossing (no turns)
- J: Crossing (vehicle turning)
- K: Merging
- L: Right turn against
- N: Pedestrians crossing

The demonstration experiment explained in Chapter 6.6 includes the simulation of a crossing situation, where the ego car is waiting to turn right (type J). It waits for an appropriate gap between the crossing traffic, which in this case will be only one other opponent vehicle. Hence, a model to assist the ego car in finding a safe gap was developed as MATLAB/Simulink controller, which is illustrated in Figure 85. Similar to the work published by Dabbour and Easa (2014), it necessitates four apriori indicators (see Table 23):

- Time-To-Intersection (TTI)
- Time-To-Disappear (TTD)
- Time-To-React (TTR)
- Gap acceptance time (GA)

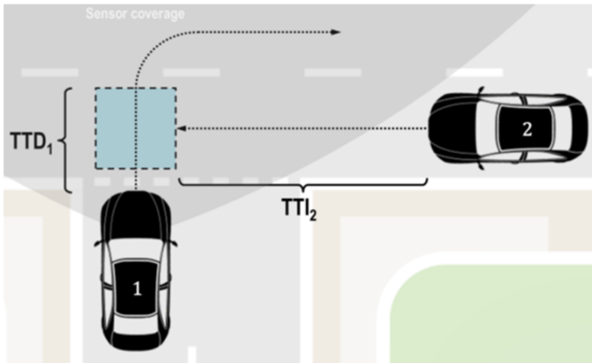


Figure 85: Illustration of the crossing and turning assistant model for collision type J "Crossing (vehicle turning)"

The TTI is a continuous metric and is defined as the estimated time a vehicle needs to reach the conflict zone, assuming that it does not change speed and direction of travel. The TTI equals the TTC, if both vehicles are on a collision course and the conflict zone is based on the vehicle dimensions (see Section 6.5.3.1).

The TTD is defined as the estimated remaining time a vehicle needs to exit the conflict zone, also assuming constant speed and direction. For the given case, the TTD is the time vehicle 1 needs to cross the trajectory of vehicle 2 in order to make the turn.

The TTR is the reaction time the vehicle needs before the turning manoeuvre is performed. In comparison to human drivers, automated driving systems will have a much faster reaction time, which basically comprises the duration of electrical signal transmission and mechanical acceleration.

The GA is the accepted time for a gap, i.e. an appropriate gap that does not make the car passengers feel uncomfortable or inpatient. As previously stated, the vehicle should behave as human-like as possible. Therefore, a gap acceptance time of 1 second is assumed, which equals the PET threshold defined in Section 6.5.3.4. This means that the time between one car leaving the conflict zone and the other one arriving at the conflict zone should be at least one second.

Ultimately, the model results in one of the two output states "remain waiting" or "turn", which are calculated as:

$$\begin{cases} Turn, & if\ TTI_2 - TTD_1 - TTR_1 > GA \\ Wait, & if\ TTI_2 - TTD_1 - TTR_1 \leq GA \end{cases} \tag{41}$$

As soon as the ego vehicle has reached the intersection and stopped to observe crossing traffic, Equation (41) is calculated at each timestamp. The initial state of this model is always "remain waiting". Obviously, Equation (41) only applies if the vehicle sensors have detected a vehicle within their range. Otherwise TTI could not be estimated. If there is no vehicle detected, the state changes from "remain waiting" to "turn".

From the moment the vehicle starts the turning manoeuvre, the crossing and turning assistant has done its job and the forward collision avoidance model becomes active again. Depending on the sensor range and angle coverage, the assistant will result in varying safe gap estimations. For example at very high speeds, the opponent vehicle might enter the sensor range when the vehicle has already entered the intersection (conflict zone), which may result in a critical situation. Note that the sensor coverage (grey area) in Figure 85 is illustrative. In the demonstration experiment, the sensor coverage is a varying simulation parameter in order to evaluate the safety performance differences between a forward-looking sensor system and an additional sideward-looking sensor system.

The crossing and turning assistant was implemented in MATLAB/Simulink by creating several modelling blocks:

1. The overall assistant input/output model, which takes the brake and gas pedal position as input and feeds them back to the vehicle control according to the assistant's output (see Figure 86).
2. The crossing and turning assistant block to control the manoeuvre for approaching the junction and or setting the actions according to the decision whether to remain waiting or to start accelerating (see Figure 87).
3. A block to decide whether to remain waiting or to start accelerating (see Figure 88), depending on the safe gap calculation.
4. A block to identify if there is a vehicle approaching (see Figure 89).
5. A block to calculate a safe gap to turn, according to Eq. (41) (see Figure 90)

As depicted in Figure 87, the *crossing and turning assistant* (2) is based on the input from the different object sensors mounted on the ego car. More precisely, a Boolean whether the respective object sensor identified the opponent vehicle or not is taken as input for the *junction entrance decision* block (3), which decides if the junction can be entered or not. Furthermore, this block controls the junction approach by calculating the optimal deceleration to stop smoothly before the junction.

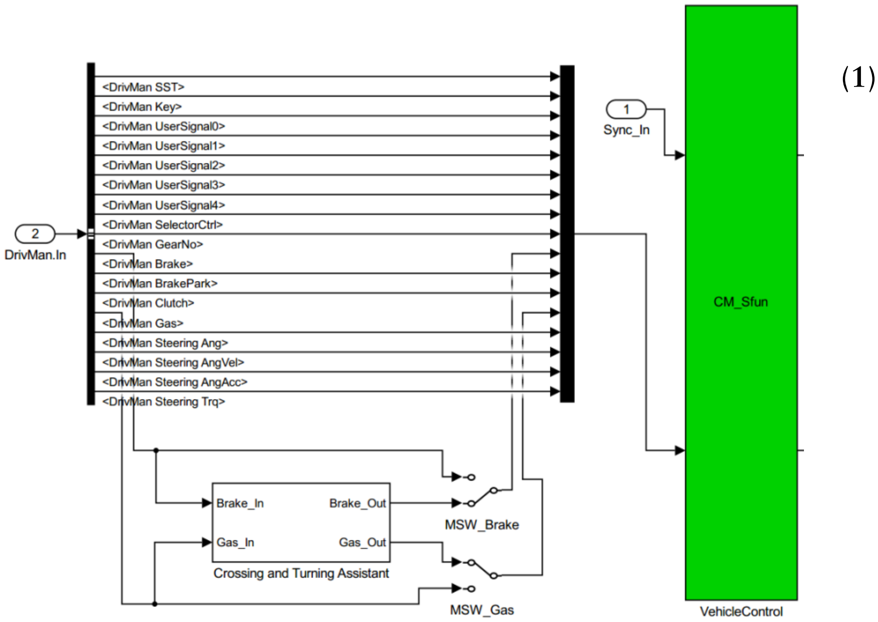


Figure 86: Simulink inputs and outputs for the crossing and turning assistant

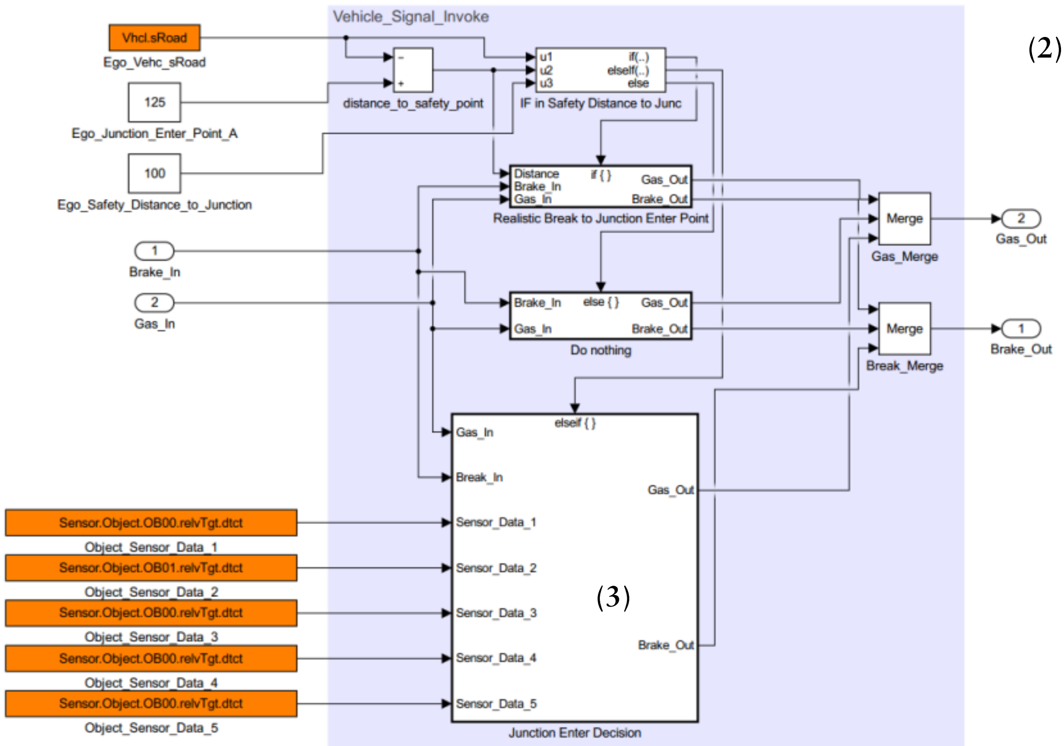


Figure 87: Simulink block for approaching the junction

The *junction entrance decision* block (3) is shown in Figure 88. The object sensor signals are transferred to the *object detection* block (4), which issues a warning if an object is within the range of the sensor. This warning is fed into the *safe gap calculation* block (5), which outputs a 1 if there is a safe gap and a 0 if the vehicle has to remain waiting.

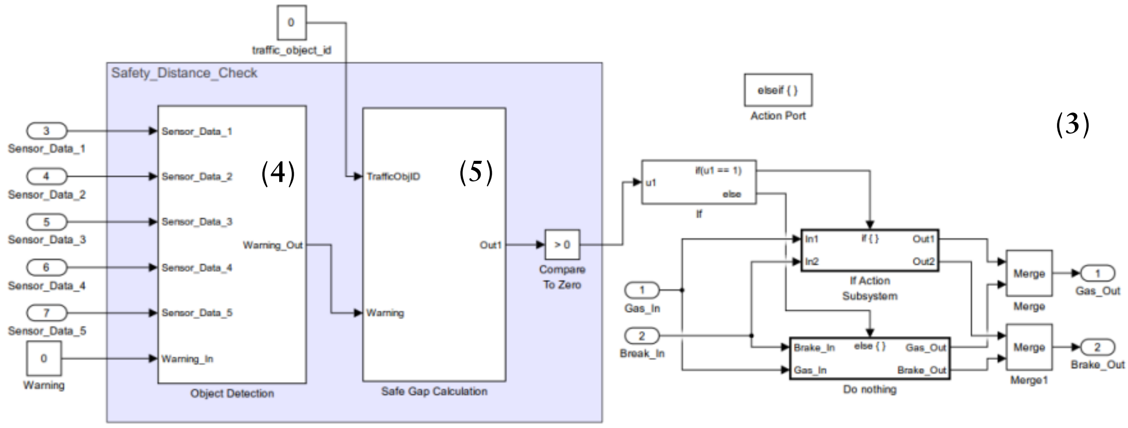


Figure 88: Junction entrance decision block in Simulink for the decision to enter the junction

Figure 89 depicts the *object detection* block (4), which issues a warning if one of the sensors has detected an object within its range and if the detection is activated for this run. As explained later in Section 6.8.4, one of the varying parameters for the simulations is the sensor detection probability. This probability is implemented by setting either 0 or 1 for the sensor activation of each run, randomly distributed among all runs according to the overall probability defined.

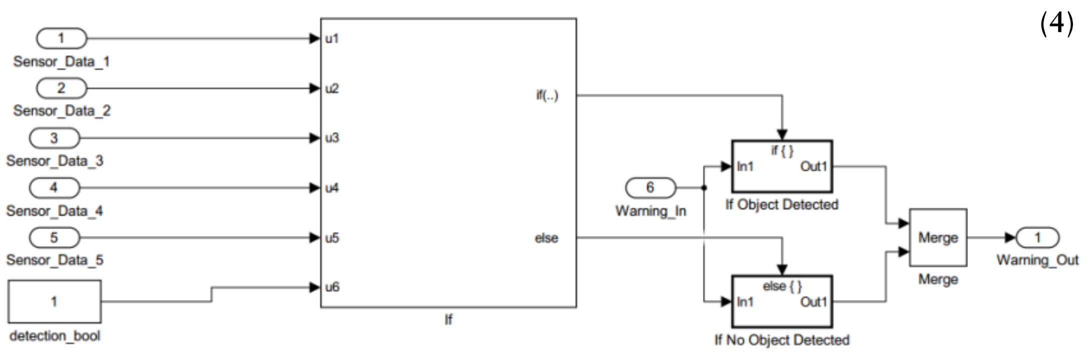


Figure 89: Object detection block in Simulink for identifying if there is an object within the range

Eq. (41) is then applied in the *safe gap calculation* block (5), as shown in Figure 90. First, the TTI of the opponent is computed by dividing its current distance to conflict by its current velocity. The distance to the conflict equals the conflict zone entrance point, which varies depending on the size of the conflict hexagon that is spanned by the projected impact angle and the vehicle dimensions. If $TTI_2 - TTD_1 - TTR_1 > 1s$, then the block outputs a 1, which triggers the ego car model to accelerate and turn. The TTR was set to a very low value of 0.01 seconds to correspond to the reaction time setting of the ego car's driving behaviour model defined later in Table 31.

(5)

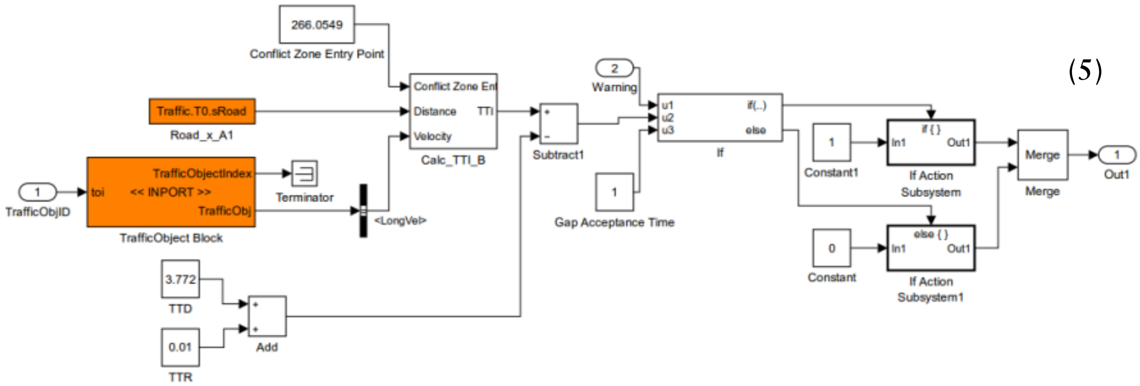


Figure 90: Safe gap calculation block in Simulink for calculating a safe gap to turn

6.7 Monte Carlo sampling

The proposed simulation and evaluation framework is designed to process a multidimensional input space consisting of varying parameters for the road configuration, driving scenario including vehicles and their sensors as well as driving behaviour. Each of those model inputs has a certain probability distribution, e.g. the velocity of vehicles might be normally distributed around the speed limit, while other variables might follow a uniform distribution. The basic tasks of the framework are (1) to generate samples from the given input distributions, (2) to run simulations for each set of the samples and (3) to perform statistical analyses on the simulation output values (see Figure 91).

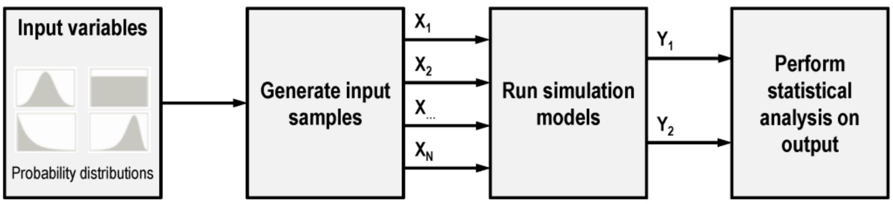


Figure 91: Principle of the Monte Carlo approach within the simulation framework

To this end, the framework applies a derivative of the *Monte Carlo method* (Metropolis and Ulam, 1949), which samples values from the input range using a (pseudo-)random number generator and then performs a large number of computations. The accuracy of the Monte Carlo method's output depends on the number of inputs used, because it approximates solutions to quantitative problems through statistical sampling. Hence, the exact solution would be found for an infinite number of samples. More precisely, the Monte Carlo method involves two computational steps:

1. Generate random variable samples $z = (z_1, z_2, \dots, z_N)$ that are uniformly distributed between 0 and 1
2. Transform the samples of the uniform variable z into a random variable x_i that follows the given distribution $F_{x_i}(x_i)$, as given by

$$x_i = F_{x_i}^{-1}(z_i), i = 1,2, \dots, N \tag{42}$$

where $F_{x_i}^{-1}$ is the inverse of the cumulative distribution function of x_i and N is the number of samples.

The Monte Carlo method is usually applied for computationally inexpensive models, because there are a high number of samples necessary to obtain satisfactory results. The approach results in much lower confidence bounds when the sample size is reduced. To cope with this problem, the framework uses a derivative of the Monte Carlo method called *Latin Hypercube Sampling* (LHS, McKay et al., 1979), which divides the cumulative probability curve into equal intervals on the scale and samples a random value from each interval of the input distribution (see Figure 92). Each output sample is constrained to match the input distribution very closely, which is why LHS is commonly used to reduce the number of runs in a Monte Carlo simulation. It takes even unlikely extremities into account as it is often desired. The number of segments equals the number of samples, i.e. a computation with 200 samples would split the probability range into 200 segments, each representing 0.5 percent of the total distribution. Figure 92 (left) illustrates the difference between the traditional Monte Carlo method and LHS and shows that the same number of samples might be distributed more irregularly.

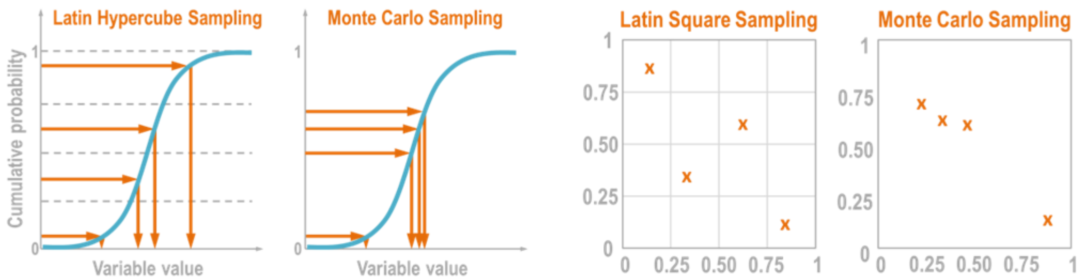


Figure 92: Difference between Latin Hypercube Sampling and Monte Carlo Sampling

The word Latin Hypercube stems from the *Latin Square* principle, which is defined as a square grid containing sample positions with strictly one sample in each row and each column (see Figure 92 right). Since the Latin Square represents two-dimensional input, the Latin Hypercube is the generalisation of this concept to a higher number of dimensions with each sample being the only one in its hyperplane segment. This means that LHS requires the knowledge about the position of the previously generated sample points, while the traditional Monte Carlo method is memoryless, as Figure 92 (right) depicts. In the demonstration experiment explained in the next section, LHS is applied to a three-dimensional input space (see Table 35), which includes the velocity and lateral position of the opponent vehicle as well as the surface friction. The sampling approach ensures that the input probability distributions are maintained and that the resulting simulation outputs are considered representative in this respect. By varying those three inputs, the ego vehicle is exposed to a large amount of different situations that combined will lead to conclusions about the collision risk and near-miss risk for the system under test.

6.8 Demonstration experiment

This chapter deals with the simulation experiment to demonstrate the applicability and versatility of the simulation and evaluation framework. A particular hypothesis is tested, which is explained in Section 6.8.1. This is followed by a description of the junction scenery (Section 6.8.2) and the collision scenario (Section 6.8.3) chosen for the demonstration. Section 6.8.4 gives a summary of all varying parameters, before the results for those particular variations are presented. Note that the real-world junction used for the demonstration is located in the UK and therefore, the simulations were done for left-hand driving and the speed limits are given in miles per hour.

6.8.1 Hypothesis to be tested

The advantage of the proposed simulation framework is that its modular architecture allows a number of different test objectives, e.g. to compare different vehicle sensor systems, evaluate varying intersection designs or assess the impacts of different driving behaviour models. For the demonstration, it was chosen to focus on the comparison of two different in-vehicle sensor systems with respect to their crash avoidance capabilities. In particular, a right-turn manoeuvre by an automated vehicle at a non-signalised T-junction is evaluated, where another vehicle is crossing from the right with the right of way. Against this background, the following *hypothesis* can be stated, which will be tested in the demonstration experiment:

A forward collision avoidance system is not sufficiently safe to avoid collisions with crossing traffic and therefore needs additional crossing and turning assistance.

This leads to the following *test objectives*:

1. Define a junction scenery that fits the test requirements (see Section 6.8.2).
2. Identify an appropriate collision scenario for the given test experiment (right turn, vehicle crossing) and junction scenery, based on the results of the association rule study (see Section 6.8.3).
3. Simulate the scenario with a representative combination of selected parameter variations (see Section 6.8.4).
4. Evaluate the collision and conflict probabilities for all simulation runs based on the safety indicators (see Section 6.9).

6.8.2 Scenery description and parameters

The demonstration experiment comprises a certain junction environment reconstructed from a real-world example, denoted as scenery. Instead of modelling an artificial, virtual road scenery, a real-world junction in the East Midlands, UK, was chosen to showcase the method for a practical example. This had the advantage that the environment parameters such as lane and shoulder widths, position of roadside elements, gradients, i.e. parameters that may not be directly derived from the crash clusters, were taken from an on-site inspection. Furthermore, taking a real-world scenery for the simulation experiment demonstrates the use of the framework for road operators and authorities to evaluate particular junctions in their network. This particular junction was selected after reviewing numerous potential sites around the Loughborough University campus, because 1) it is non-signalised, which is considered as a risk factor for ADS when testing crossing and turning assistance systems, 2) the sight distance is limited due to obstructing bushes, trees and a fence, 3) the junction angle is not orthogonal and both paths have one lane in each direction with relatively narrow widths, and 4) it is a hilly environment with slopes and chances of glare by the afternoon sun.

This approach is different from traditional crash simulation studies, where real-world road locations are reconstructed case-by-case from a crash sample of the underlying accident database and the effectiveness or crash avoidance rates of safety interventions is studied (Brunner et al., 2003; Canu et al., 2016; Cliff and Moser, 2001; Helmer, 2014; Sander, 2017; Sander and Lubbe, 2018). Commonly, those studies use a crash reconstruction tool such as the software PC-Crash. In this thesis, the virtual junction is not included in the OTS samples, because the crash samples are not evaluated case by case. Instead, generalised abstract scenarios are obtained from clustering and further specified by association rules. More concrete, less abstract scenarios are then produced by parametric variation using the LHS method. In other words, a large number of artificial critical scenarios are evaluated instead of accident scenarios that really happened. In this way, more parametric variation can be implied than by using a limited crash population.

Figure 93 shows four pictures of the environment, which is a rural T-junction (without any traffic islands or dualling features) with 40 mph speed limit, with a minor road terminated by a major road. There is no shoulder, neither on the major nor on the minor road. A pedestrian footpath can be seen between the major road and the houses as well as on the nearside of the minor road. The existing road markings are slightly worn but still well visible. The yield instructions are indicated by a give-way sign and a road marking before the stop line.

Table 27 lists all simulation parameters necessary to build up the virtual scenery. The first column gives the name of the parameter and the second column tells whether the parameter is kept static or variable. According to the chosen scenario, A and B are allocated to a certain path of the scenery (see Section 6.8.3). The third column gives the actual parameter values used for the simulations. It can be seen that only one parameter is set to variable (shaded in light blue), namely the surface friction for wet conditions. This is a criticality factor that affects the modelling of the scenery.

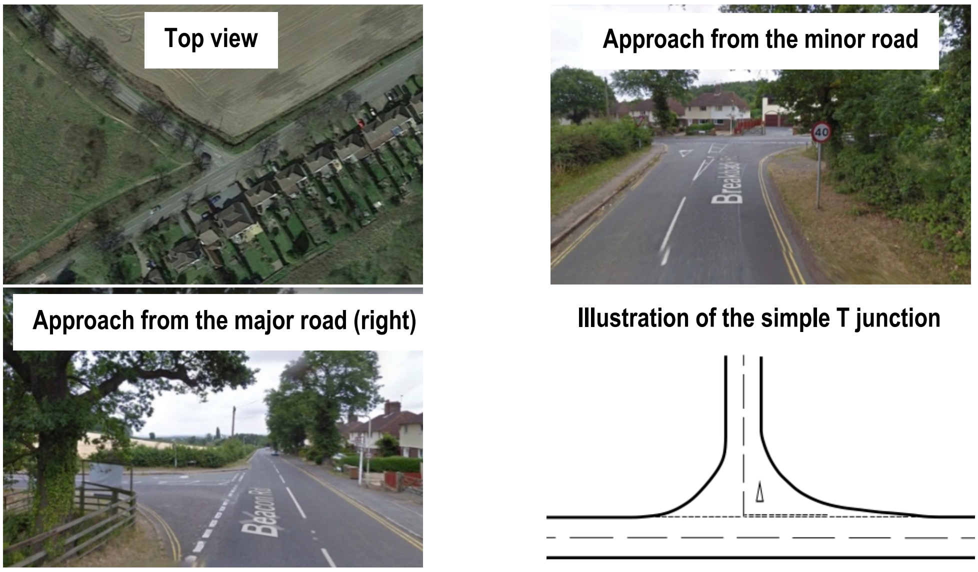


Figure 93: Example junction selected for the demonstration experiment (Source: Google Street View)

Not all of the parameter values could be derived from the accident dataset, because they were not available in the final selection of variables. This is the number of lanes, the lane width, the existence of a shoulder, the junction angle and information about existing road markings. Therefore, those values have to be assumed. All other parameters were taken from the description of a selected scenario as result of study 2, which is explained in the next section. The virtual representation of the selected scenery in CarMaker is depicted in Figure 94.

Table 27: Simulation parameter setup related to the scenery

Road Scenery Parameters	Static/Var.	Values
Area	static	Rural
Carriageway type in path of A and B	static	Single carriageway in both paths
Number of lanes in path A and B*	static	1 lane in both paths
Lane width in path of A*	static	2.7 m
Lane width in path of B*	static	2.7 m
Roadside in path of A*	static	Nearside: Pedestr. sidewalk, Offside: No shoulder
Roadside in path of B*	static	Nearside: No shoulder, Offside: Pedestr. sidewalk
Road markings available at junction*	static	Yes
Speed limit at the path of A	static	40 mph
Speed limit at the path of B	static	40 mph
Junction shape from A's path	static	T: minor road terminated by a major road
Junction angle*	static	80 degrees
Junction type	static	Simple T-junction without island or dualling
Traffic control for path of A	static	Give-way sign
Horiz. geometry	static	Straight
Light conditions	static	Daylight
Surface friction coefficient	variable	Wet conditions, uniform distribution from 0.3 to 0.9

** Parameter value that is not available from the selected accident data set and has therefore to be assumed.*

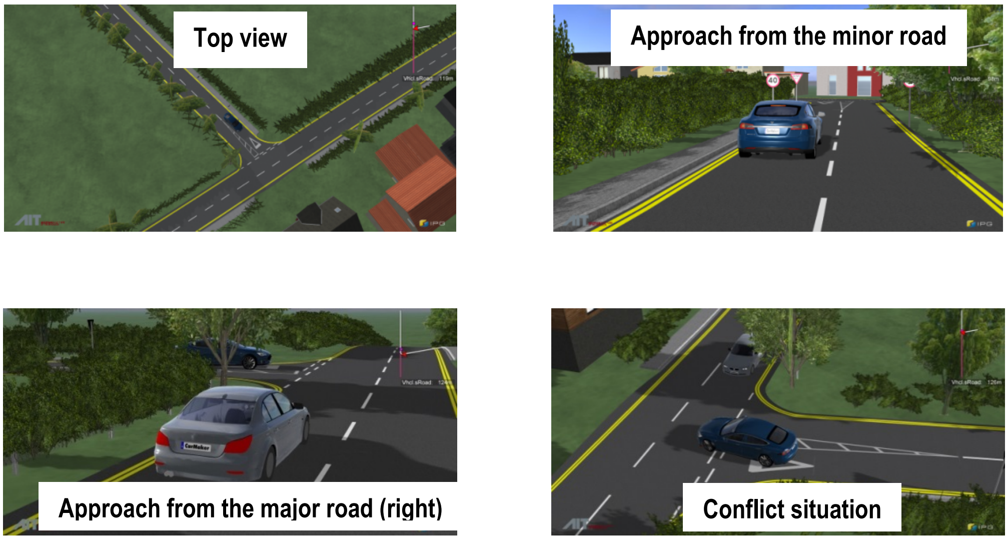
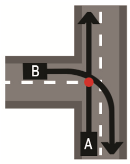


Figure 94: Virtual reconstruction of the selected junction scenery in CarMaker

6.8.3 Scenario description and parameters

The collision scenario selected for the abovementioned scenery is T-10.2 (see Section 5.4.3.1), where car B is simulated as automated vehicle and A as human-operated car:

> *"Car A is going straight on a major road and hits another car B, which is emerging from a minor road on the left with the intention to turn right. A is travelling on a single carriageway in a rural area with a speed limit of 40 mph or 50 mph without active or static yield instructions, and it is caused by B failing to give way. The surface is wet and A suffers serious injury."*



As explained in Section 6.2, the scenario is defined by parameters for the ego car and the opponent road user. In this case, the opponent is a car going straight and the ego car is the automated vehicle turning right. Hence, the collision is of type J, called "Crossing-Vehicle Turning" (see Appendix B) and the collision is caused by B failing to give way. Table 28 lists all parameters necessary to reconstruct the scenario in CarMaker. The variable parameters related to the ego car are the type of collision avoidance system, the sensor range according to weather conditions, the sensors' accuracy of object detection capability. The variable parameters related to the opponent are the lateral position in the lane and the velocity when approaching the junction.

Table 28: Simulation parameter setup related to the scenario

General Scenario Parameters	Static or variable	Values
Collision type	static	J – Crossing (Vehicle Turning)
Main causation factor	static	Fail to give way from ego car
Ego Car Parameters		
Car model*	static	see Table 29
Car sensor models*	static	see Table 30
Manoeuvre	static	Turning right into the major road
Driving lane*	static	1

Lateral position in lane*	static	Centre of the lane
Target velocity*	static	17.77 m/s (according to speed limit of 40 mph)
Driving behaviour*	static	see Table 31
Pre-crash reaction*	static	Full braking
Collision avoidance system*	variable	FCA, Crossing & Turning Assistant
Object detection probability*	variable	95% (rather satisfying) / 75% (poor)
Opponent Parameters		
Opponent model	static	Passenger car (BMW 5)
Manoeuvre	static	Going straight on the major road
Driving lane*	static	1
Lateral position in lane*	variable	Normal distribution around lane centre
Approaching velocity*	variable	Normal distribution around speed limit
Driving behaviour*	variable	see Table 32
Pre-crash reaction*	variable	Full braking
Collision avoidance system*	static	None

** Parameter value that is not available from the selected accident data set and has therefore to be assumed.*

6.8.3.1 Ego Car Parameters

The ego car is simulated as automated vehicle, which requires a number of model parameter assumptions, such as vehicle dimensions, weight, engine, transmission, suspension or sensors. The electric vehicle Tesla Model S was taken as basis for the car model (see Figure 95). The most important model parameters selected for this experiment are given in Table 29.

Figure 95: Tesla Model S and its self-driving functionality (Source: tesla.com)

Table 29: Ego car model parameter setup

Vehicle model parameters	Value
Vehicle model	Tesla S 75D
Transmission	Automatic
Power train	Electric
Engine power	190 kW
Maximum engine torque	345.0 Nm at 3750.0 rpm
Wheel base	2,959 mm
Unloaded weight	2,108 kg
Vehicle length	4,976 mm
Vehicle width and height	W 1,963 mm, H 1,435 mm
Track width	Front: 1,661 mm, Rear: 1,699 mm
Rear overhang	1,009 mm
Driving axle	Rear driven
Turning radius	11.3 m
Tire size	245 / 35 / R21

The Tesla vehicle uses a fusion of different sensors, which includes a combination of radar, cameras and ultrasonics. In fact, there are two different philosophies on which sensors to use for automated driving. While Tesla believes in camera vision and image processing methods, other manufacturers such as Google use LIDAR instead. Both approaches have their strengths and weaknesses. In this demonstration experiment, a combination of three video cameras is assumed (1x forward and 2x sideways) and the radar is disregarded. Ultrasonic sensors are also disregarded, because they are mainly relevant for low-speed manoeuvres such as parking. In CarMaker, the cameras are modelled as object sensors, which detect objects and other road users within their range. Those object sensors are the primary source for the recognition of other road users and the crash avoidance model. Table 30 lists the sensor model parameters set up in CarMaker. The main forward camera is mounted near the rear-view mirror oriented towards the direction of travel, having a horizontal field of view of 100 degrees. The sideward-facing cameras are mounted over the front doors with an orientation orthogonal to the direction of travel and a field of view of 100 degrees (see Figure 96).

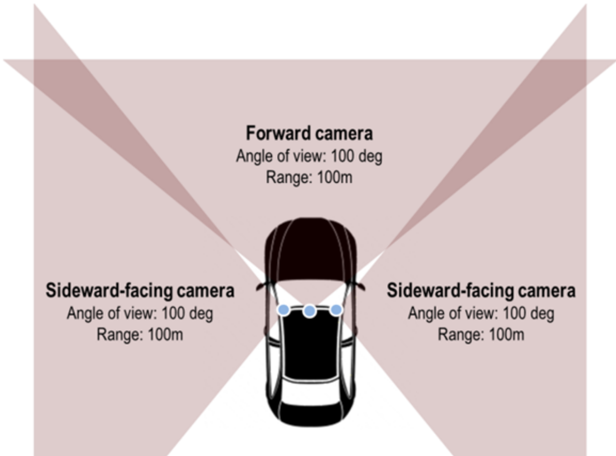


Figure 96: Illustration of the ego car's sensing system

Table 30: Ego car's sensor model parameters setup

Sensor model parameters	Value
Number and type of environment sensors	3 cameras
Main Forward Camera: Range	100 m
Main Forward Camera: Orientation	Direction of travel
Main Forward Camera: Angle (horizontal)	100 degrees
Main Forward Camera: Angle (vertical)	20 degrees
Main Forward Camera: Mounting position	Near the rearview mirror behind the windscreen
Sideward-facing Cameras: Range	100 m
Sideward-facing Cameras: Orientation	Orthogonal to direction of travel
Sideward-facing Cameras: Angle (horizontal)	100 degrees
Sideward-facing Cameras: Angle (vertical)	20 degrees
Sideward-facing Cameras: Mounting position	Vehicle sides over the front doors

Modeling automated driving behaviour is a complex task and requires assumptions on how an automated vehicle behaves in comparison to a human driver. While e.g. the reaction time can be assumed to be minimised in automated vehicles, other factors such as deceleration, acceleration and steering behaviour needs to be assumed. Nowadays, fully automated vehicle prototypes such as the Google self-driving car behave rather defensively than aggressively. Most passengers would feel uncomfortable in an automated vehicle that accelerates and brakes heavily or that does not keep an appropriate distance to the vehicle in front, which must be taken into account by the vehicle manufacturers. For this demonstration experiment, the automated driving behaviour is set to rather defensive, e.g. the maximum acceleration and deceleration is relatively low. Table 31 lists all relevant driving behaviour parameters and their values. Note that the collision avoidance models described in Section 6.6.8 ignore the maximum deceleration value, since they use much higher decelerations for emergency braking. The *reaction time* was set to a very low value, since human reactions are replaced by system reactions for steering and braking. This is also the reason why the time to change pedals was decreased to 0.1 second, which would normally equal the time to move the foot from the throttle to the brake pedal.

Table 31: Ego car's driving behaviour parameters setup

Driving behaviour parameter	Value	Comments
Max. longitudinal acceleration	2.0 m/s²	CarMaker's default value for "defensive" driving
Max longitudinal deceleration	-2.0 m/s²	Not relevant for collision avoidance manoeuvres
Max. lateral acceleration	3.0 m/s²	CarMaker's default value for "defensive" driving
Reaction time	0.01 s	Longitudinal and lateral reaction time (braking/steering)
Time to change pedals	0.1 s	An ADS is assumed to do this rather quickly
Corner cutting coefficient	0.25	CarMaker's default value for "defensive" driving
Traction control	On	Reduction of throttle if wheelspin occurs
Target speed	40 mph	Corresponds to the speed limit
Tolerated speed deviation	1.0 mph	Deviation from target speed
Tolerated lateral deviation	0.05 m	Lateral deviation from target trajectory
Max. steering wheel angle	630 deg	CarMaker's default value
Max. steering wheel velocity	1,000 deg/s	between defensive and normal
Max. steering wheel acceleration	6,000 deg/s²	between defensive and normal

In CarMaker, the way to take a curve can be parametrised by the *corner cutting coefficient*. It defines the driving course within the lane borders and lies between 0 and 1. Zero means that the vehicle precisely drives in the middle of the driving lane, while one means that the driver model uses the whole lane width to calculate its desired trajectory. A value of 0.25 was chosen, because an automated vehicle is supposed to turn precisely, while a lower value may affect passenger comfort due to more abrupt steering.

CarMaker's driver model includes an internal traction controller, which reduces the throttle when wheel spin occurs. The *traction control* was activated, because it is assumed that automated vehicles will avoid spinning wheels, as modern cars do. The *target speed* was set to the local speed limit of 40 mph, i.e. the driver model aims to hold this speed within a tolerated deviation of 1 mph. The *lateral deviation* was set to a very low value of 0.05 m to allow minimum deviation from the calculated trajectory.The maximum *steering wheel angle* was kept at the default value of 630 degrees, which means that the driver stops turning the steering wheel when this angle is reached. Additionally, the maximum *steering wheel velocity and acceleration* were set to lie between the default values for defensive and normal driving.

6.8.3.2 Opponent Parameters

The opponent vehicle in this simulation experiment is modelled as a traffic object (see Section 6.6.6). Hence, the opponent vehicle model is simplified and does not include physical dynamics models and sensor models. However, traffic objects still react to other traffic objects or the ego car, e.g. to avoid collisions, to overtake or to follow behind. Vice versa, the ego car is able to detect the traffic object with its sensors. Table 32 gives the main vehicle parameters as well as the opponent's basic driving behaviour settings. The 3D visualisation of the opponent vehicle is depicted in Figure 97.

Table 32: Opponent car model parameter setup

Vehicle model parameters	Value
Vehicle model	BMW 5
Unloaded weight	1,600 kg
Vehicle length	4,800 mm
Vehicle width	1,800 mm
Vehicle height	1,250 mm
Front / Rear overhang	800 / 1000 mm
Max. longitudinal acceleration	2.0 m/s²
Max longitudinal deceleration	-10.0 m/s²
Max. lateral acceleration	4.0 m/s²
Initial speed	40 mph

Figure 97: Opponent car model

6.8.4 Summary of parameter variations

Given the test objectives, the main parameter to vary for the comparison is the type of collision avoidance system of the ego car, which has two states (see Table 33). The safety performance of the forward collision avoidance system (AEB) is compared to the more advanced crossing and turning assistant combining forward and sideward-facing sensors (see collision avoidance systems explained in Section 6.6.8). It is expected that the advanced system results in lower collision and conflict probabilities. However, this hypothesis is tested and quantified by the demonstration experiment.

Table 33: Varying parameters with two states each to compare against

Parameter name	Possible states
Collision avoidance system	[Forward-facing sensors, Forward&sideward-facing sensors]
Object detection probability by vehicle sensor	[95%, 75%]

Since the collision avoidance models are ideal systems, i.e. they do not have any faults, a fault must be injected that leads to a system failure, denoted as *failure or fault injection* (Elgharbawy et al., 2016; Pintard, 2015; Pintard et al., 2013; Ziade et al., 2004). Table 34 gives examples of failure modes of a collision avoidance system that can lead to a non-detection or late and false detection of other road users or obstacles. In the simulation framework, a failure is represented by two different object detection probabilities, where 95% defines a rather satisfying value and 75% a rather poor value. The *object detection probability* is defined as the likelihood that another road user is detected by the vehicle sensors within their range. Vice versa, the failure probability is 5% and 15%, respectively. This assumption is simplified, since in reality the detection probability depends on several influencing parameters such as light and shade areas, where the image processing algorithms may have problems to identify objects and measure their distance, reflections or sun glare, which affects the overall image quality of the camera, or poor weather conditions such as heavy rain, snow or fog, which negatively influences the object recognition and distance measurement.

Table 34: Examples of collision avoidance system failure modes, causes and effects

System	Failure mode	Possible effects	Possible causes
Sensing system	Sensor is defect and does not record any data.	No sensor data transferred to perception/control unit.	Overheated, power supply error, cable disconnected etc
	No communication to perception unit.	No sensor data transferred to perception/control unit.	Faulty connectors, cable disconnected etc.
	Raw sensor data is noisy or erroneous.	Erroneous data transferred to perception/control unit.	Dirty sensor, interference, cable defect etc.
Perception and control unit	Image or pattern recognition algorithm fails to detect.	No, late or false detection of road users or obstacles.	Poor light/weather conditions, reflections etc.
	No communication to braking or steering systems.	Vehicle control does not react to detected objects.	Faulty connectors, cable disconnected etc.
	Wrong information is transferred.	No, late or false detection of road users or obstacles.	Cyber attack and misuse, interference, cable defect etc.
Braking system	Fail to brake.	Collision or near-miss with other road user or obstacle.	Overheating, brake valve fault etc.
	Insufficient braking	Collision or near-miss with other road user or obstacle.	Overheating, blocked wheels etc.

In the presented case, the detection state is either 0 or 1 for each run, randomly distributed, meaning that from 100 runs the opponent vehicle will not be detected in 5 runs or 25 runs, depending on the probability variation. This simplified approach was chosen, because the CarMaker software version used had limitations regarding physical sensor models when used with MATLAB/Simulink. This is why a failure injection was implemented. For future studies and future releases of CarMaker, it is planned to replace the ideal sensors with high-fidelity sensor models that can accurately replicate real-world behaviour and weaknesses of sensors.

Since both variables have two possible states, the simulations must be performed for four combinations. In addition to the varying collision avoidance system and object detection probability, there are the Monte Carlo parameters (see Table 35), which are sampled from a predefined probability distribution, as explained in Section 6.7. These include the opponent's lateral position within the lane and its velocity when approaching the junction as well as the surface friction coefficient, which influences the braking distance and vehicle stability. While the first two parameters are sampled from a normal distribution, the latter one is based on a uniform distribution in the range from 0.3 to 0.9.

Table 35: Varying parameters for Monte Carlo sampling

Parameter name	Range	Distribution	Mean	Std. dev.
Opponent's lateral position in lane	[-1;1]	Gaussian	0 m (centre)	0.2 m
Opponent's approaching velocity	[0;17.77]	Gaussian	17 m/s	4 m/s
Surface friction coefficient	[0.3;0.9]	Uniform	n/a	n/a

The *lateral position* in the lane describes the deviation of the opponent's driving trajectory from the lane centre. Note the lane centre is defined as 0 m, with the lane boundaries defined as +/- half of the lane width. Since the opponent is going straight, the trajectory and hence the lateral position in the lane remains constant throughout the simulation run, if there is no vehicle instability due to a braking manoeuvre. The mean equals the lane centre and the standard deviation has been set to 0.2 m to allow enough variation of the vehicle within the lane and to reduce the likelihood of travelling outside the lane, which has been set to 2.7 m. The probability density function (PDF) and cumulative density function (CDF) for the lateral position are given in Figure 98a.

The *approaching velocity* of the opponent vehicle is the second variable for the LHS. The speed limit at the selected junction scenery in the UK is 40 mph, which equals 17.77 m/s. Thus, the mean velocity has been set to 17 m/s, which is slightly below the speed limit. The standard deviation has been set to 4 m/s, which equals 9 mph, according to speed measurement data that were provided by the local authorities for the selection junction. The PDF and CDF for the approaching velocity are given in Figure 98b.

The surface *friction coefficient* is uniformly distributed between a range of 0.3 and 0.9, which was found to be an appropriate range for wet surface conditions (Gruber and Maurer, 2004; Velske et al., 2009, p. 263), as the scenario requires. Note that the friction coefficient remains constant for the entire road surface within the simulation run. The PDF and CDF for the friction coefficient are given in Figure 98c.

The output dataset (from the LHS) for comparison has been set to 4,000 samples for each of the four variable combinations, respectively. This results in 16,000 combinations, i.e. simulation scenarios, which comprise around 142 hours of driving for the ego car. Considering that each parameter combination requires a reference run to obtain the theoretical, ideal vehicle trajectories (as explained in Section 6.4.3), the number of simulation result files increases to 20,000.

It must be noted that not all of the simulation runs necessarily lead to a collision, because (1) the cars could miss each other due to the varying approaching speed and lateral position of the opponent, and (2) the vehicle sensors avoid the collision successfully. However, to decrease the probability of undisturbed situations, which are not relevant for the safety evaluation, the initial longitudinal position of the opponent car was set in a way that both cars would collide with each other, if the opponent drove exactly with the mean of its approaching speed distribution, if the opponent was located exactly at the mean of the lateral position distribution and if the road friction was exactly the mean of its distribution. Note that the ego car always starts at the same longitudinal position, approaches with the same speed and on the same lateral position. Through the variation of the parameters for the opponent car, and also through varying road surface friction, either the point of collision changes or there is no collision at all. In the latter case, the conflict indicators help to evaluate whether it is a near-miss or an undisturbed situation.

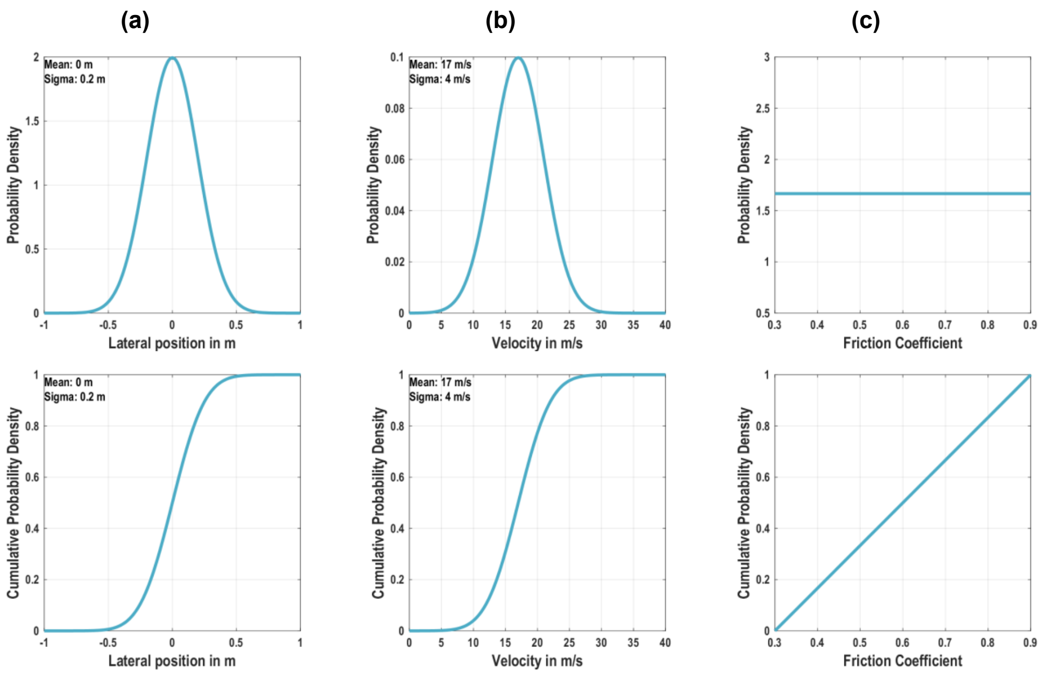


Figure 98: PDFs and CDFs of the three Monte Carlo variables (a) lateral position in lane, (b) approaching velocity and (c) friction coefficient

Figure 99 visualises the results of the Latin Hypercube sampling for the three variables according to their probability distributions. The dots represent the samples in the respective two-dimensional plane for better readability. For example, the yellow dots show the y-z plane (velocity against friction) and the sample distribution. The red dots represent the samples in the x-z plane (velocity against lateral position) and the x-y plane distribution (lateral position against friction) is shown by the blue dots.

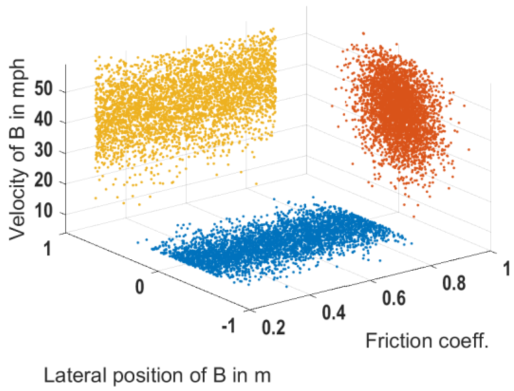


Figure 99: Results of the Latin Hybercube sampling for the three variables

6.9 Results of the safety performance evaluation

This section summarises the demonstration experiment by presenting the resulting performance indicator figures. The first three sections provide guidance through the evaluation of individual simulation runs, distinguished into an example run resulting in a collision, one resulting in a conflict and one undisturbed scenario. Following up on this, Section 6.9.4 presents the overall results of the vehicle sensor system comparison, including all 16,000 simulation runs.

6.9.1 Example evaluation of a collision

As mentioned in Section 6.5, there are different indicators used depending on whether there was a collision or a conflict. The following example simulation run was picked from the collisions that resulted when using the forward-facing sensor system with 75 percent detection probability. It is shown to demonstrate the evaluation procedure based on the collision indicators.

Figure 100 depicts a time-distance diagram of both the ego car A (blue line) and the opponent car B (ochre line) with a summary of the calculated indicators in the textbox on the left. As in all runs of this experiment, A intends to stop before the intersection, intends to detect a safe gap to cross the lane and then turns right. This explains the flattening of the curve before the time of collision. In this case, A has entered the junction without recognising B correctly and hence fails to give way. Note that due to the setting of the starting conditions, the distance from the start position to the intersection differ from A to B. Therefore, both trajectories were aligned in the plot based on the distance to the collision point, i.e. the trajectories intersect where the collision point and collision time intersect. The plot further illustrates the arrival at the conflict zone (see Figure 71). It can be seen that A enters the conflict zone approximately two seconds before B. As soon as B enters the conflict zone it collides with A, because A is still located within the conflict zone. B can therefore be called the bullet vehicle. As explained, CarMaker does not simulate impacts and deformations. From the time of the collision, the ongoing trajectory curves can thus be disregarded, because the cars continued travelling as if there was no collision.

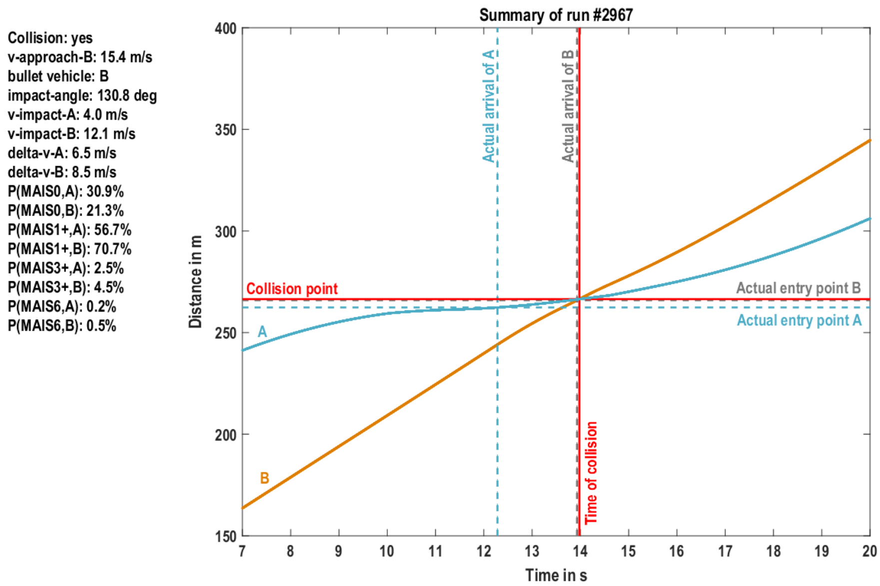


Figure 100: Summary plot of a collision scenario

The velocity of B when approaching the intersection was 15.4 m/s (34.7 mph), which is below the speed limit of 40 mph. At the time of collision, A was in the turning phase with a speed of 4 m/s. The impact speed of B reduced to 12.1 m/s due to an emergency braking manoeuvre. The corresponding changes in velocity Δ_v were estimated with 6.5 m/s for A and 8.5 m/s for B, respectively. The reason why Δ_v for A is higher than its actual impact velocity is the way it is calculated, namely based on Newton's momentum conservation principle, as Eq. (18) shows. The impact is assumed inelastic and does not include vehicle rotation or post-impact rebound.

The collision severity is estimated by the injury risk probabilities for the different MAIS levels, but only MAIS0, MAIS 1+, MAIS 3+ and MAIS 6 are given in the figure. Due to the relatively low Δ_v values, the serious or fatal injury risk is low. However, there is a 56.7 percent and 70.7 percent risk for the occupants of car A and car B, respectively, to suffer slight injuries. The probability of remaining uninjured is 30.9 percent for the occupants of car A and 21.3 percent for those of car B.

In order to calculate the timestamps of entering and exiting the conflict zone, the trajectories of both cars must be analysed in detail. Figure 101 (left) gives a screenshot of the collision situation from the CarMaker video, while Figure 101 (right) plots the trajectories within the global coordinate system of the simulation environment. Note that for car A there are two trajectories depicted, namely the front axle centre plus/minus half of the vehicle width (blue lines) as well as the rear axle centre plus/minus half of the vehicle width (grey lines). The ochre lines represent car B's trajectory. Since A was turning, its tractrix curve must be taken into account to calculate the size of the conflict hexagon. The green dots indicate the hexagon entry points of both vehicles and the resulting hexagon is highlighted in dark blue. The red dots represent the cars' position at the time of the collision. The resulting hexagon is highlighted in dark blue.

In CarMaker, the origin of the ego vehicle frame is located at the rear of the vehicle. To calculate the hexagon entry point precisely, the frontmost point of the ego car A is taken as new reference point in the global coordinate system. This is done in CarMaker by placing a so-called inertial sensor on the vehicle front and to output its coordinates. The opponent car model is simplified and there is no option to place an inertial sensor. Hence, for car B the reference point remains at the rear of the vehicle. This means that the distance to the hexagon entry point had to be corrected by the vehicle length. For the vehicle exit points, the time at which the vehicles pass with their rear end is relevant. Consequently, the origin at both cars' rear end was used for that. Given the two different origins of the vehicle frame, it can be explained why the position at the time of collision (red dots in Figure 101 right) is located at the front of car A, but at the rear of car B.

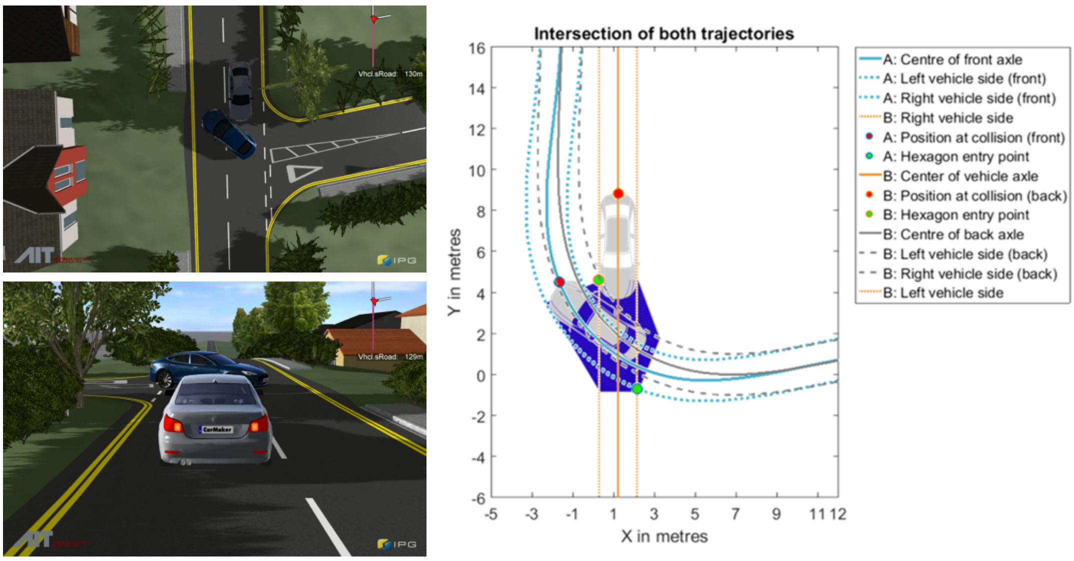


Figure 101: Screenshot of collision (left) and trajectory intersection plot (right) with B driving southwards and A turning right coming from the west

So far, Figure 100 has shown the distances covered over time and Figure 101 gave an idea on the vehicle trajectories and conflict hexagon calculation. The velocity, longitudinal acceleration and the jerk over time are depicted in Figure 102. It can be seen that the collision happened after 14 seconds indicated by the red vertical line. The blue line shows a deceleration of A before the collision, which is caused by a stop before the intersection. The collision thus occurred after the acceleration of A. On the right side, the velocity and acceleration curves of B show an emergency braking conducted right before the crash. It can be seen that there are two negative peaks in the jerk curve of B, although only the second peak exceeds -8 m/s³, which indicates a critical braking manoeuvre. Again, the curves after t_{coll} can be disregarded, because the drive would practically end there.

The fourth and last plot of this example evaluation is the Time-To-Collision curve shown in Figure 103. The TTC and TTD of A is coloured in blue, while the curves of B are shown in ochre. Three timestamps are plotted by vertical lines, namely the time of collision and the time when A and B entered the conflict zone. TTC and TTD are continuous values computed at each timestamp and decrease as the vehicles approach the conflict zone. The increase of TTC and TTD of A after 10 seconds can be explained by the vehicle stop and therefore decreased velocity. Since the time is calculated by the distance divided by velocity, the curves would become infinite if the vehicle stands still. As car A continues driving shortly after the stop, the TTC and TTD values drop again. Not surprisingly, the TTC values become zero (and theoretically negative after that) at the time when the conflict zone is reached.

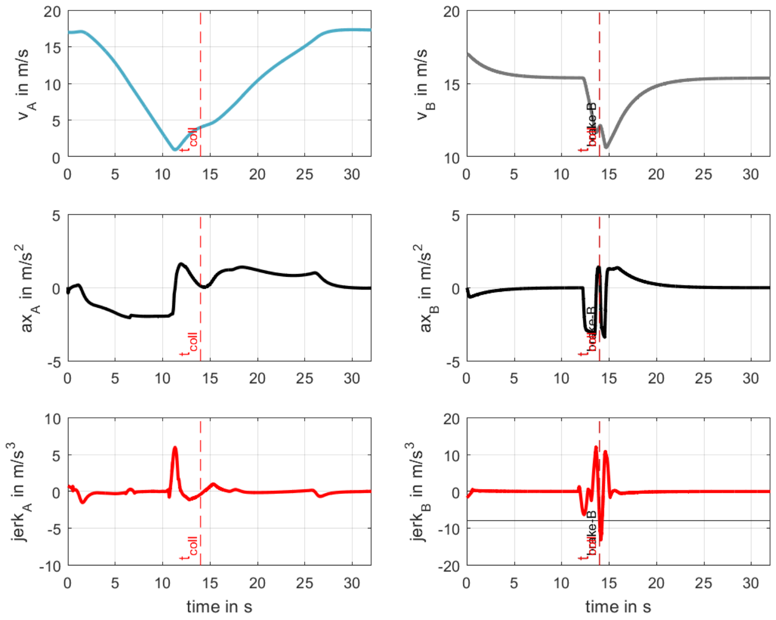


Figure 102: Velocity, longitudinal deceleration and jerk over time for vehicle A (left) and vehicle (B)

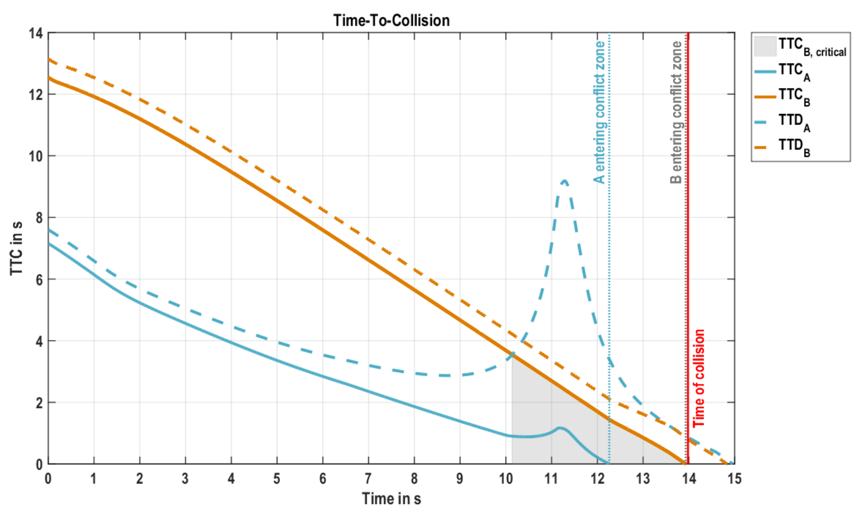


Figure 103: TTC and TTD over time for both vehicles in a collision scenario

Applying the formula used by Van der Horst (1990) in Eq. (30) reveals the timespan at which a critical TTC is reached. As long as car A has not left the conflict zone, i.e. as long as $TTD_A > TTC_B$, car B is on theoretical collision course, assuming that both vehicles continue driving at the same speed.

Summing up the indicator results for this example case, it can be observed that this particular simulation sample caused a collision, where B hit A on its offside, while A was turning right. Car A failed to give way due to inaccurate detection of B and crossed the intersection, although car B was approaching with a speed of 34.7 mph. This resulted in a 56.7 percent risk of slight injury for A.

6.9.2 Example evaluation of a conflict

In comparison to the collision scenario in the previous section, this section presents the results obtained for a near-miss event. The following example simulation run was picked from the conflicts that resulted when using the forward-facing sensor system with 75 percent detection probability. It is shown to demonstrate the evaluation procedure based on the conflict indicators.

Figure 104 depicts a time-distance diagram of both the ego car A (blue line) and the opponent car B (ochre line) with a summary of the calculated indicators in the textbox on the left. As in all runs of this experiment, A intends to stop before the intersection, intends to detect a safe gap to cross the lane and then turns right. Similar to the collision case, A has entered the junction without recognising B correctly and hence fails to give way. Both trajectories were aligned in the plot based on the distance to the conflict zone entry point. It can be seen that A enters the conflict zone approximately three seconds before B. When B arrives at the conflict zone, A has just left it, which resulted in a PET of 0.5 seconds. Since the threshold for a safe gap is a PET of one second, this simulation was classified as conflict event. Furthermore, a TTA of 2.1 seconds was measured for car B, which means that an emergency braking manoeuvre was conducted at a TTC of 2.1 seconds.

Figure 105 (left) gives a screenshot of the conflict situation from the CarMaker video, while Figure 105 (right) plots the trajectories within the global coordinate system of the simulation environment. Again, for car A there are two trajectories depicted, namely the front axle centre plus/minus half of the vehicle width (blue lines) as well as the rear axle centre plus/minus half of the vehicle width (grey lines). The ochre lines represent car B's trajectory. The green dots indicate the hexagon entry points of both vehicles and the resulting hexagon is highlighted in dark blue. Not surprisingly, the resulting conflict hexagon has a similar shape to the collision scenario presented above, because the trajectories changed only slightly. The main difference to the collision scenario is the time when the vehicles enter and exit the conflict zone. In this case, it can be seen that car B could avoid a collision by braking and the low PET shows how close the vehicles have passed this junction.

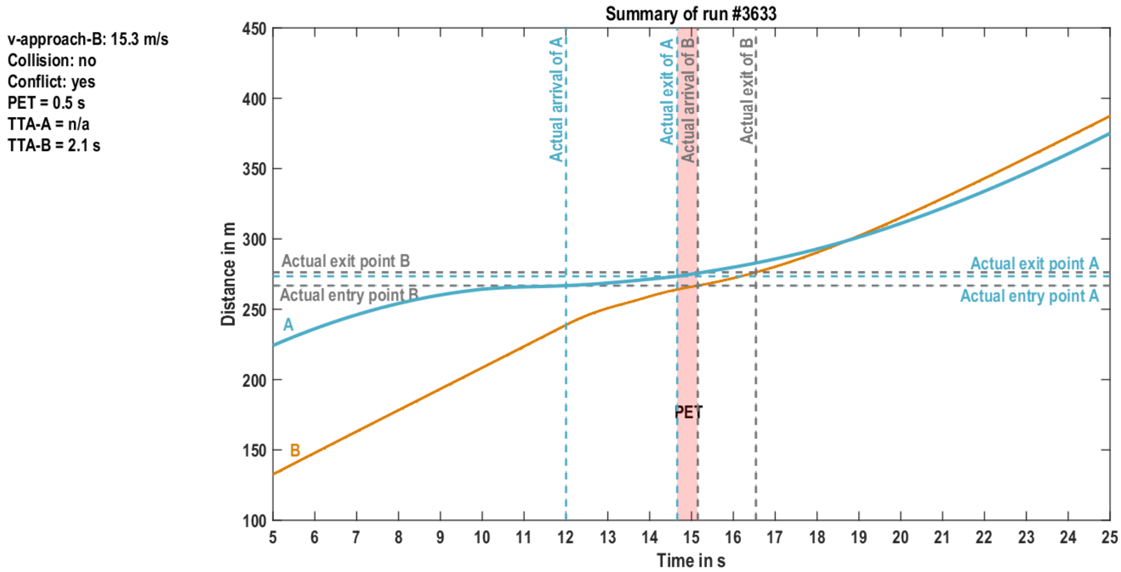


Figure 104: Summary plot of a conflict scenario

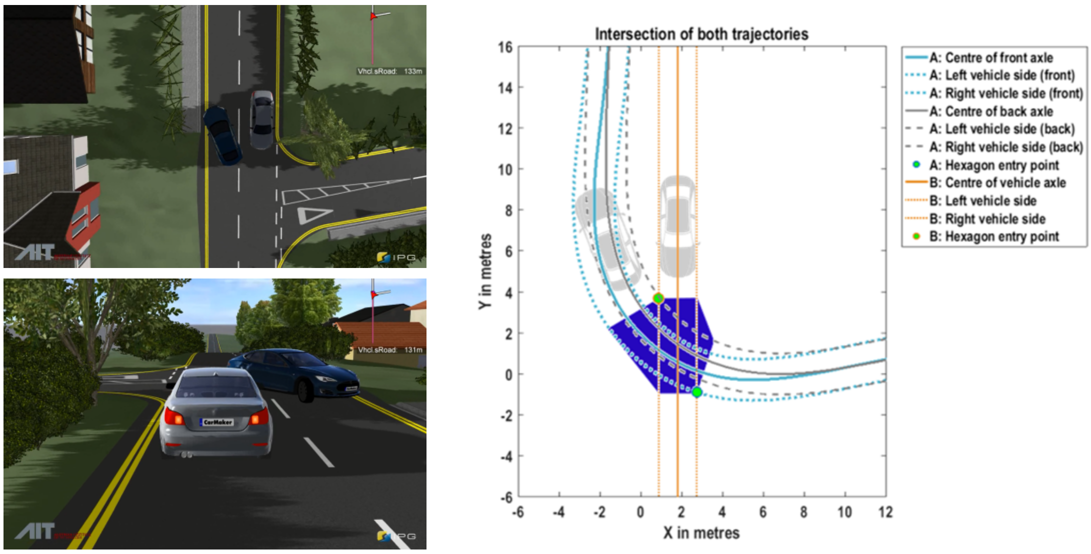


Figure 105: Screenshots of a conflict in the simulation environment (left) and trajectory intersection plot (right) with B driving southwards and A turning right coming from the west

The velocity, longitudinal acceleration and the jerk over time are depicted in Figure 106. On the right side of the figure, the velocity and acceleration curves of B show an emergency braking conducted at approximately $t_{brake-B} = 12s$. The negative jerk value exceeds -8 m/s³, which indicates a critical braking manoeuvre. It can further be seen that A did not brake at all, except when approaching the intersection to stop regularly.

The curve in Figure 107 shows the TTC and TTD over time. Two timestamps are plotted by vertical lines, namely the time when A and B entered the conflict zone. Equivalently to the collision scenario, the increase of TTC and TTD of A after 10 seconds can be explained by the vehicle stop before the junction. As car A continues driving shortly after the stop, the TTC and TTD values drop again.

Applying the formula used by Van der Horst (1990) in Eq. (30) reveals the timespan at which a critical TTC is reached. As long as car A has not yet left the conflict zone, i.e. as long as $TTD_A > TTC_B$, car B is on a theoretical collision course, which is indicated by the grey area in the plot. During this timespan, in fact slightly after A enters the conflict zone, B performed a harsh braking manoeuvre, which resulted in a TTA of 2.1 seconds. From second 12.8 on, $TTD_A < TTC_B$, which means that car A would have left the zone already, if B continued driving with the same speed.

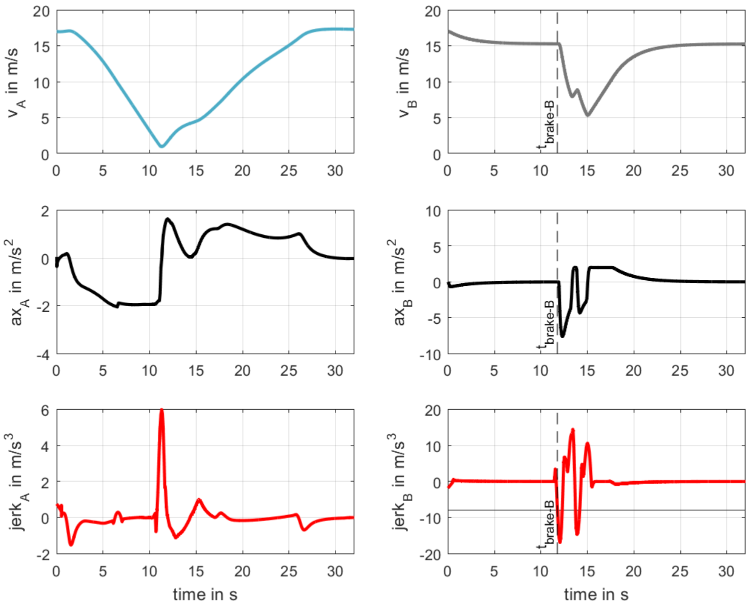


Figure 106: Velocity, longitudinal deceleration and jerk over time for vehicle A (left) and vehicle (B)

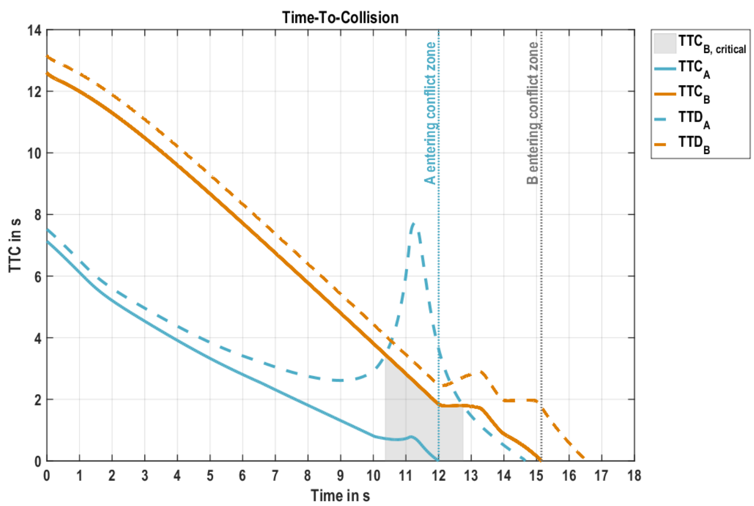


Figure 107: TTC and TTD over time for both vehicles in a conflict scenario

This example scenario can be summed up as follows: Car A failed to give way due to inaccurate detection of B and crossed the intersection, although car B was approaching with a speed of 34.4 mph. This resulted in a near-miss event with a PET of 0.5 seconds and a TTA of 2.1 seconds. If B had not pushed the brakes, a collision would have happened most likely.

6.9.3 Example evaluation of an undisturbed situation

The third example simulation run was picked from the undisturbed situations that resulted when using the forward-facing sensor system with 75 percent detection probability. Neither a conflict indicator threshold was exceeded nor a collision was detected by CarMaker's collision sensor. In this experiment, this happens when A correctly detects the position and velocity of B to find a safe gap to turn, or when the velocity of B is too high or too low to cross the intersection the same time as A does.

The summary plot in Figure 108 shows a *PET* of 1.8 seconds, which is above the critical threshold of one second. Apparently, B enters and exits the conflict zone before A arrives, which can be explained by the higher speed of B (52 mph in comparison to 34 mph in the previous cases). It demonstrates that the variation of B's approaching speed results in a variety of situations, which the ego car A is exposed to.

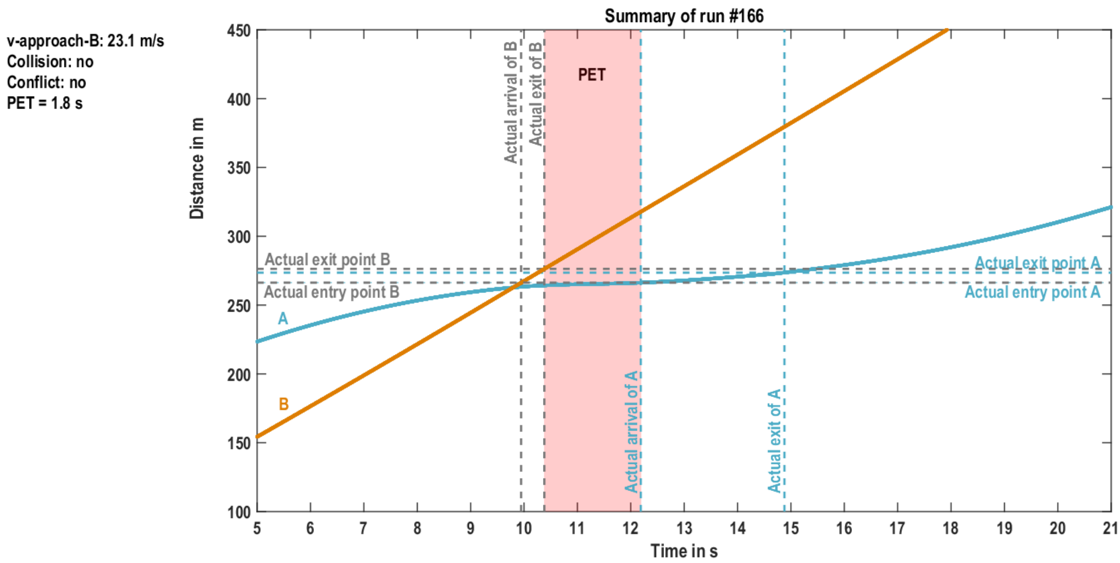


Figure 108: Summary plot of an undisturbed scenario

Figure 109 (left) gives two screenshots of the undisturbed situation from the CarMaker video, while Figure 109 (right) plots the trajectories within the global coordinate system of the simulation environment. Again, the resulting conflict hexagon has a similar shape to the scenarios presented above, because the trajectories changed only slightly. The main difference to the previous scenarios is the time, when the vehicles enter and exit the conflict zone. In this case, it can be seen that car B could safely pass the conflict zone before car A arrives at the intersection.

Figure 110 shows the velocity, longitudinal acceleration and the jerk over time. Neither A nor B performed a braking manoeuvre, expect when smoothly decelerating towards the junction. There is no critical event indicated by those curves. The increase of velocity and hence acceleration of B is caused by the starting conditions of the simulations. In each run, B has a starting velocity of 17 m/s and then reduces or increases this speed to the target approaching speed.

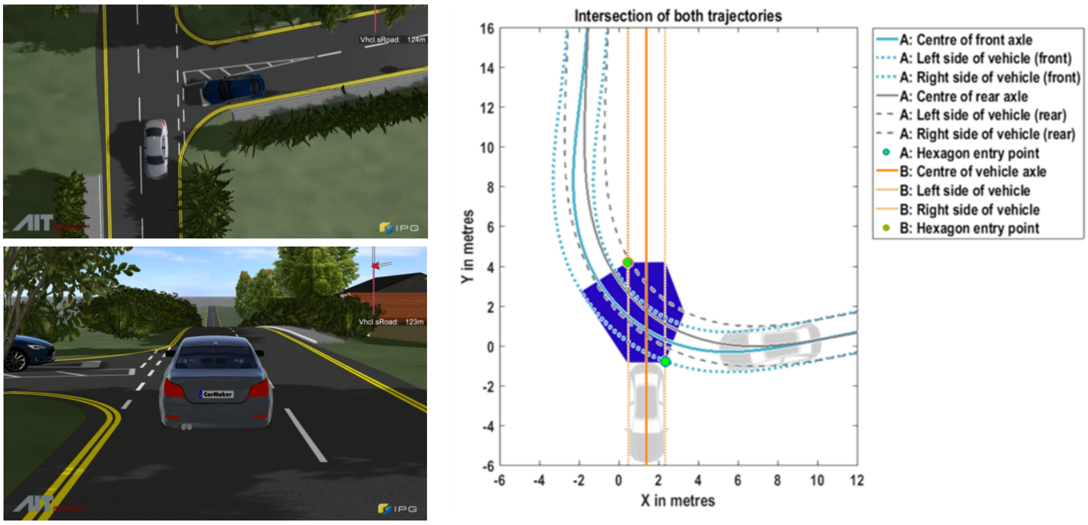


Figure 109: Screenshots of an undisturbed situation (left) and trajectory intersection plot (right) with B driving southwards and A turning right coming from the west

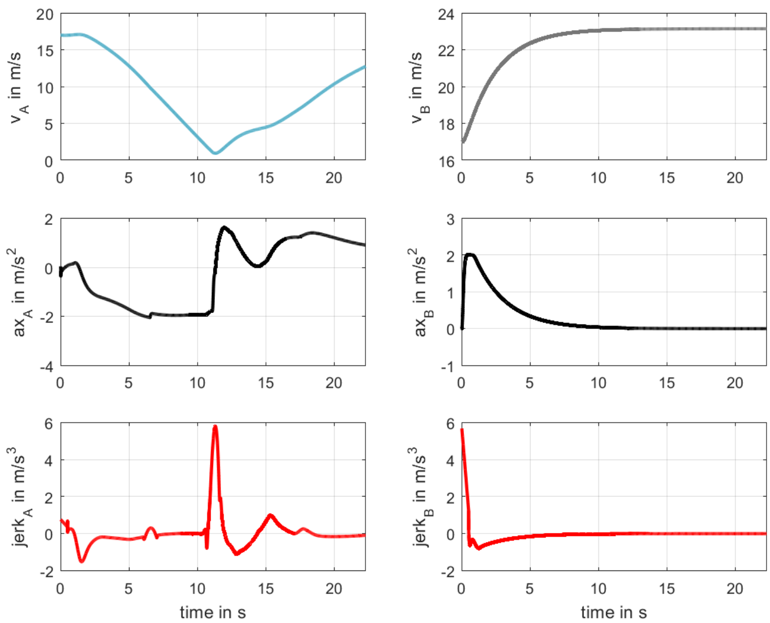


Figure 110: Velocity, longitudinal deceleration and jerk over time for vehicle A (left) and vehicle (B)

The TTC curves in Figure 111 confirm the finding derived from the previous plots. The illustration of the timespan of critical TTC shows that B (ochre curves) is on a theoretical collision course (without knowing and considering that A stops before the junction) from second 6.7 to second 8.2, but this changes as soon as $TTC_B < TTC_A$. From second 8.2 to second 9, car A is on a projected collision course, because $TTD_B > TTC_A$. In other words, during this timespan car B would still be in the conflict zone, if A continued driving with the same speed. From second 9 on, $TTD_B < TTC_A$, which means that car B would have left the zone already, if A continued driving with the same speed.

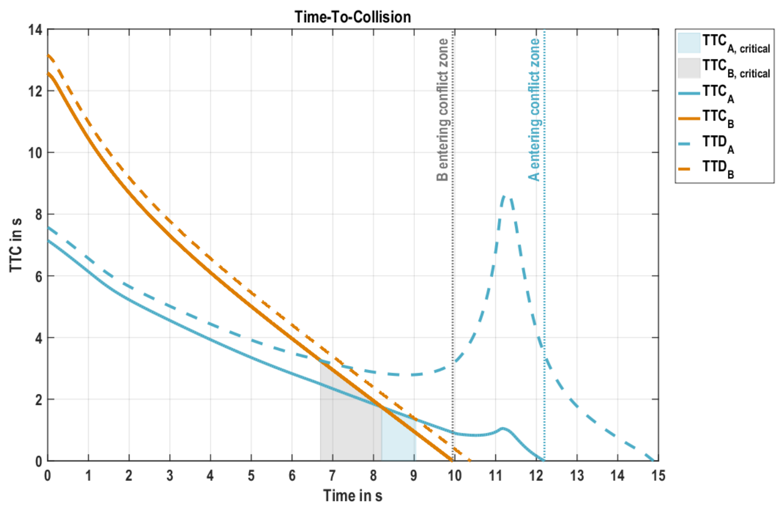


Figure 111: TTC and TTD over time for both vehicles in an undisturbed scenario

This example scenario can be summarised as follows: There was no conflict or collision, because car A correctly detected B and identified a safe gap. The PET is 1.8 seconds and therefore not critical and there were no emergency braking manoeuvres.

6.9.4 Overall comparison of vehicle sensor systems

As explained in Section 6.8.4, the varying parameters for the analysis are the type of vehicle sensors and the detection probability. This leads to four different results, each of which based on 4,000 simulation runs. Figure 112 shows a general overview of the frequency of collisions, conflicts and undisturbed situations, for the forward-facing sensor (left) in comparison to the additional sideward-facing sensor (right) and for 75 percent (top) and 95 percent detection probability (bottom). The figures thus present the overall results of the safety performance evaluation.

For the worst-case combination, i.e. 75 percent detection probability with forward-facing sensors only, the experiment resulted in 1,214 collisions, 379 conflicts and 2,407 undisturbed situations. It must be mentioned that even with zero percent detection probability, the experiment would have resulted in undisturbed situations, since due to the distribution of approaching speeds there might be runs where the cars do not come close within the conflict zone.

The additional sideward-facing sensors resulted in much lower numbers of collisions (305 for 75 percent and 49 for 95 percent detection probability). Equivalently, the number of conflicts was reduced by more than a half from 379 to 169 for the sideward-facing sensors with 75 percent detection probability. For 95 percent detection, the number decreased to more than a third from 370 to 92 conflicts.

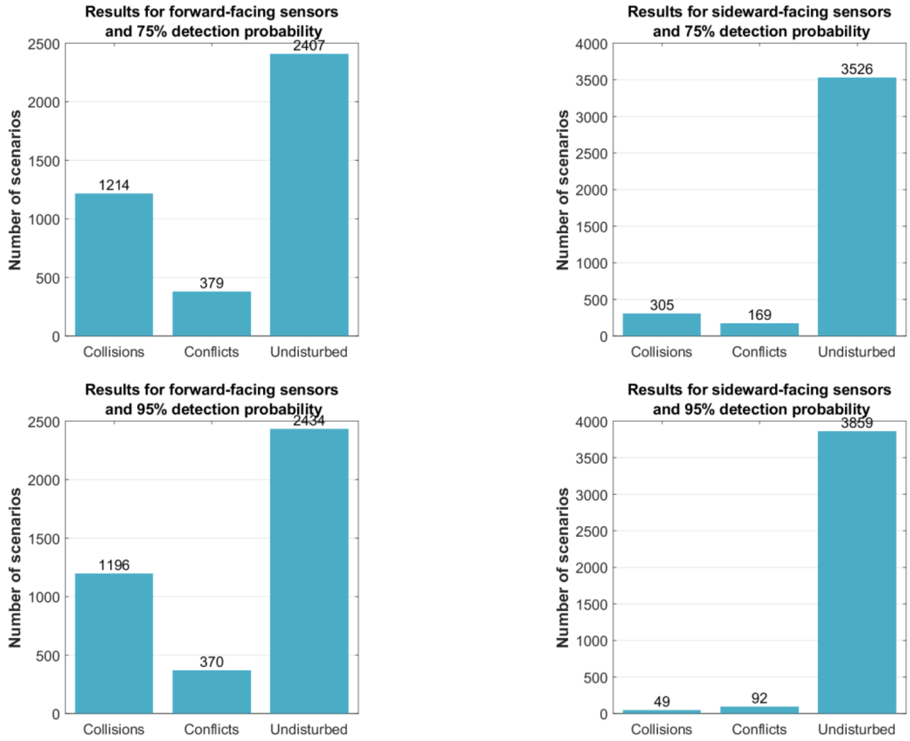


Figure 112: Frequency of collisions, conflicts and undisturbed scenarios for different detection probabilities of forward-facing sensors (left) and additional sideward-facing sensors (right)

In addition to the visual presentation above, the probabilities for collision, conflict and undisturbed situation are listed in Table 36. It can be seen that in the best case, i.e. sideward-facing sensors with 95 percent detection probability, 96.5 percent of all simulations resulted in no collision or conflict. This implies that not all of the 5% corner cases with detection failure led to a collision or conflict. The reason why the percentage of undisturbed situations is higher than the detection probability is due to the nature of the parameter variation for the opponent's approaching speed. In some runs, the vehicles were just not close enough to exceed a conflict indicator threshold.

Table 36: Probabilities for a collision, conflict and undisturbed situation for all parameter combinations

Parameter combinations	Collision	Conflict	Undisturbed
Forward-facing sensors, 75% detection probability	30.3%	9.5%	60.2%
Forward-facing sensors, 95% detection probability	29.9%	9.3%	60.9%
Sideward-facing sensors, 75% detection probability	7.6%	4.2%	88.2%
Sideward-facing sensors, 95% detection probability	1.2%	2.3%	96.5%

6.9.5 Analysis of collision indicators

Having presented the overall outcome of the parameter variations, this section will now move on to show the detailed results for different collision indicators such as injury probabilities, impact speed or impact angle. Note that for the sake of brevity, forward-facing sensor system (FCA) is henceforth referred to as *FS* and the additional sideward-facing sensor system (crossing and turning assistant) as *F+SS*. Furthermore, the ego car is represented by *A* and the opponent car by *B*.

The charts in Figure 113 show the distributions of MAIS 1+ (slight injury) injury probabilities for all detected collisions in the four variations. In general, the percentages lie between 40 and 90, which indicates a wide range of possible slight injury outcome. The mean value of all combinations is between 66 and 69 percent, and the standard deviation lies between 8.5 and 9.0 percent. There is no observable difference between μ and σ of the four combinations, but the distribution indicates that the FS simulations do have a steeper slope towards the lower probability values. However, the sample size of F+SS (95%) is considerably lower than for the other distributions.

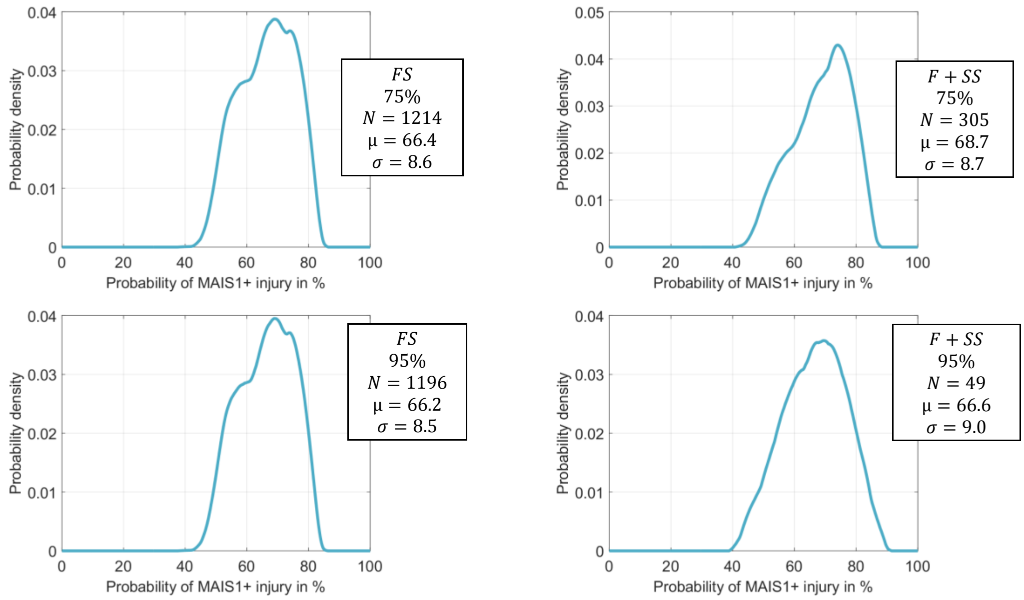


Figure 113: Probability distributions for MAIS 1+ injury probability, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

The charts for the distributions of the MAIS3+ probability are shown in Figure 114. The mean values of all four combinations are between 4.0 and 4.5 percent with a standard deviation of 1.4-1.6 percent. This means that the injury probability strongly decreases from slight to serious. The main reason for this can be found in the relatively low impact speeds of the ego car and the resulting Δ_v. More information is given in Figure 116.

As expected, the probability for a fatal injury (MAIS6) is also low, with mean values between 0.4 and 0.5 percent (see Figure 115). Note that the horizontal scale was adapted due to the low values in comparison to MAIS1+ and MAIS3+. As for the previous injury levels, this estimation is based on a simplified model for calculating Δ_v. Accordingly, only inelastic collisions are assumed and rotations or secondary impacts of the vehicles are not considered. Also, the model for the injury estimation assumed belted occupants. There is no notable difference between the four combinations and the small sample size of 49 for F+SS (95%) seems to represent the overall distributions quite accurately.

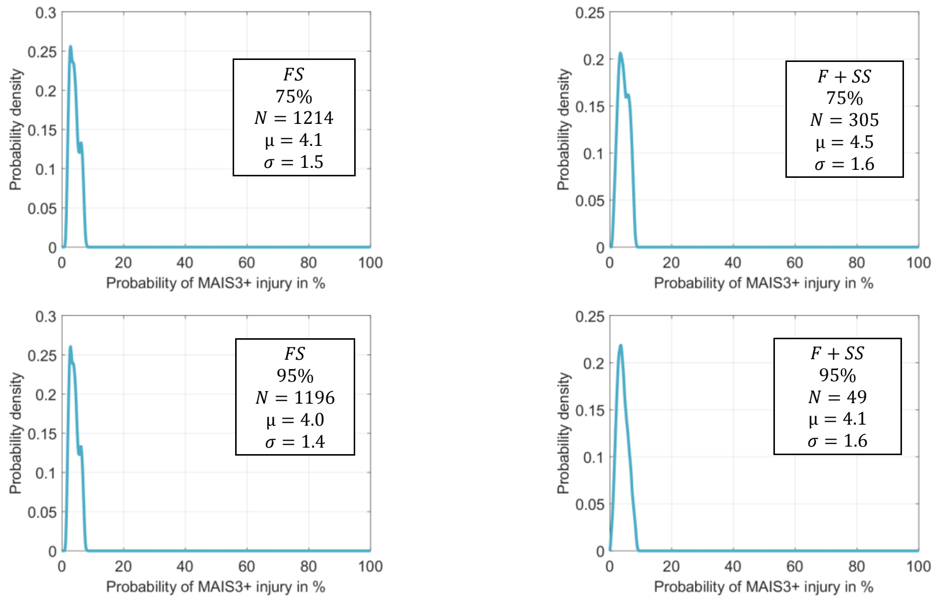


Figure 114: Probability distributions for MAIS 3+ injury probability in percent, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

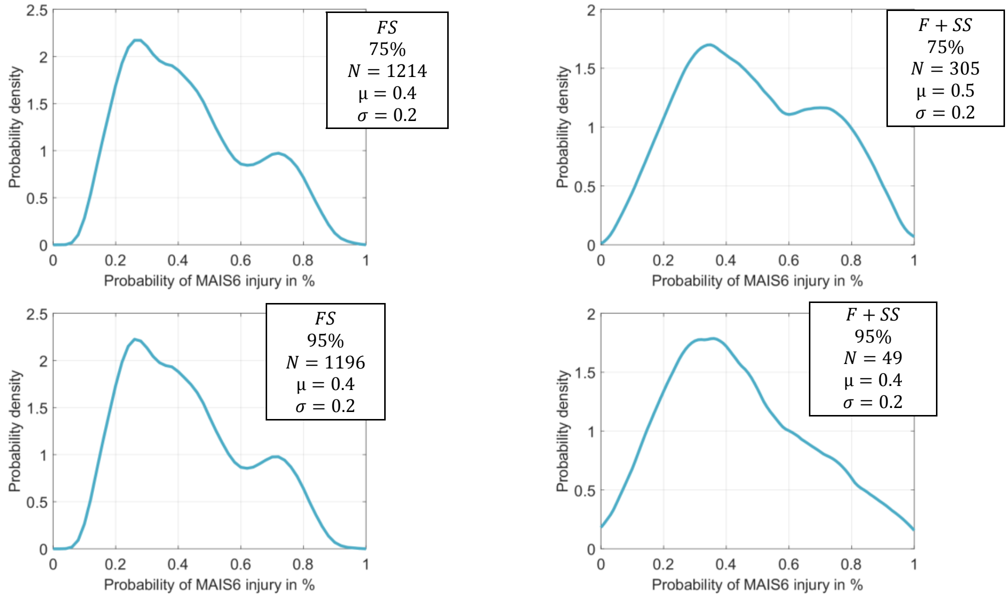


Figure 115: Probability distributions for MAIS 6 injury probability in percent, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

Due to the predefined manoeuvre for the ego car, which includes a stop before the intersection, the impact speed of the ego car has a low mean value between 6.8 and 7.3 mph (see probability density functions in Figure 116). The standard deviation lies around 2 mph. The deviations in the ego car's impact speed are caused by the different times of collisions. At some collisions, the ego car has just entered the conflict zone, i.e. it has just started to accelerate from the standstill and the impact angle is close to 90 degrees. In other collision cases, the ego car has almost completed the turn, but is still hit by the opponent car, which hence results in higher impact speeds (and lower impact angles) of the ego car.

In comparison to the ego car A, the opponent B has higher impact speeds, because it cruises downhill with a constant approaching speed defined by the given probability distribution for the Monte Carlo sampling. The mean values range from 27.3 to 29.4 mph with a standard deviation between 7.5 to 7.8 mph (see Figure 117). Similar to the previous results, the sensor systems or the detection probability do not have a notable influence on the distributions, but on the overall frequency of collisions.

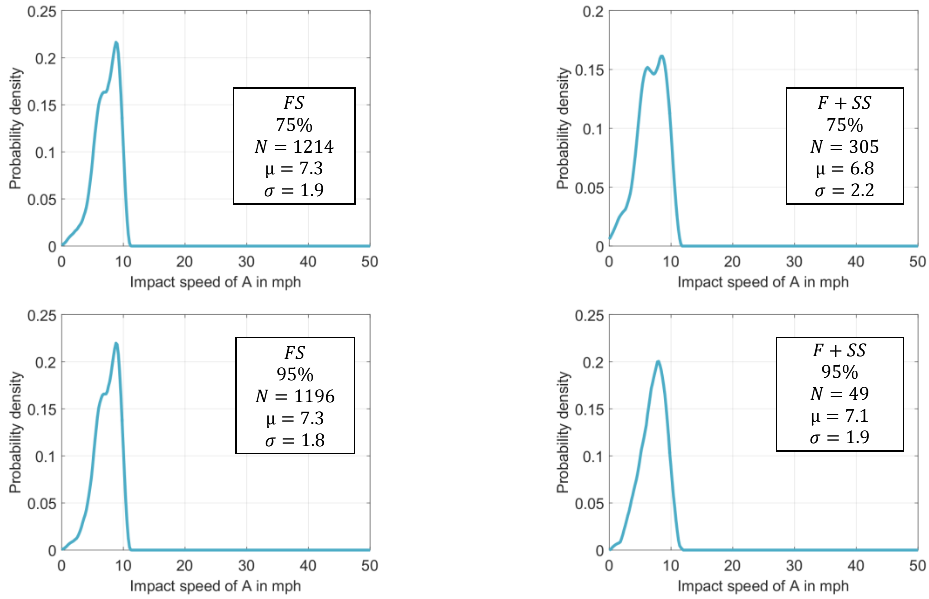


Figure 116: Probability distributions for the impact speed of the ego car, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

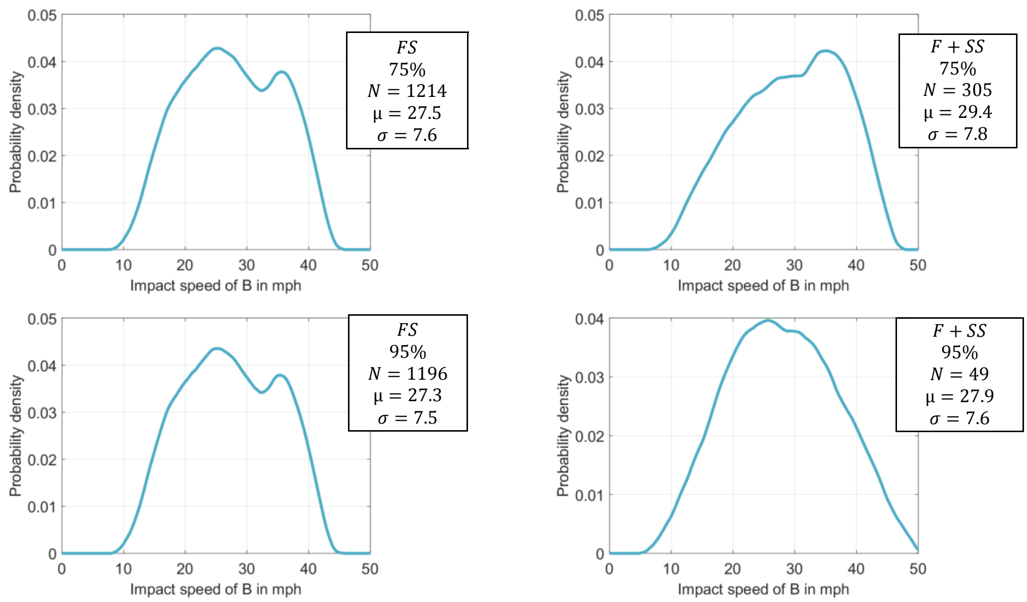


Figure 117: Probability distributions for the impact speed of the opponent car, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

The Δ_v of A in comparison to B is presented in Figure 118. The probability distributions show that Δ_v of the ego car A is considerably lower than the Δ_v of the opponent B. The mean values of B range from 17.8 to 18.7 mph, while those for A are around 14 mph. One of the factors that influence the calculation of Δ_v is the angle at which the vehicles collide. Figure 119 presents the probability distributions obtained for the impact angle. The mean values throughout all combinations lie between 111 and 115 degrees, with a standard deviation between 19.0 and 19.2 degrees. This indicates that the cars mostly collided at higher angles, i.e. when the right-turning ego car is within the conflict zone or about to leave, being hit by the opponent from the right. Lower impact angles would suggest that car A was about to enter the zone, when hit by car B.

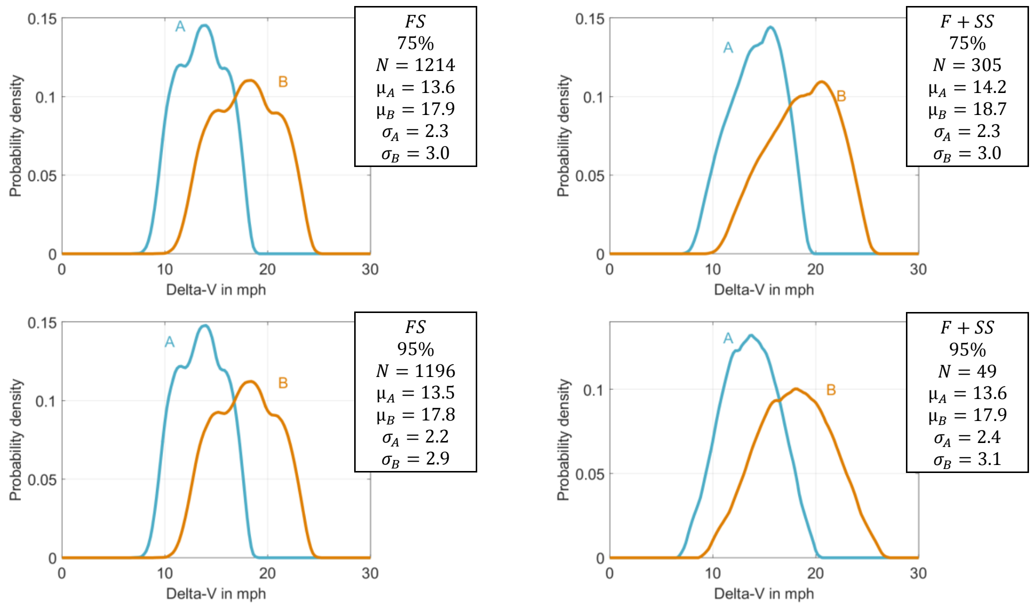


Figure 118: Probability distributions for delta-v of A and B, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

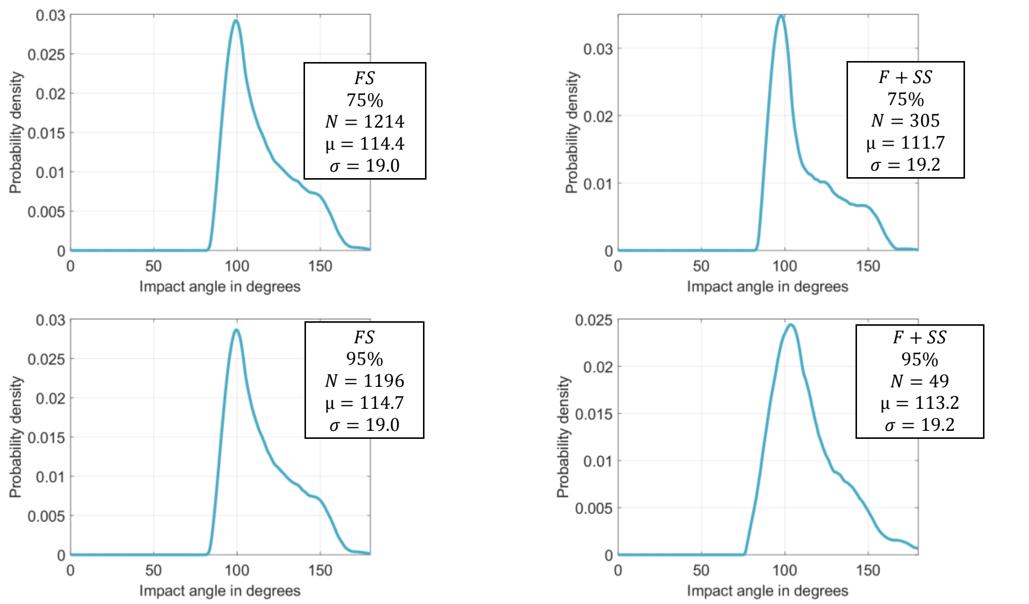


Figure 119: Probability distributions for the impact angle, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

An interesting aspect when analysing the collisions is the information about which car was the bullet vehicle, because this helps to estimate injury severities and to define the collision type. As shown in Figure 120, in approximately 60 percent of all simulation runs, the opponent vehicle was the bullet and the ego car was the target. Only in the F+SS case with 75 percent detection probability, the ego car was equally often the bullet or the target. In most cases of this particular scenario, if the ego car A was the target, then B would hit the offside of A. On the other hand, if B was the target, then A would most likely hit the nearside of B.

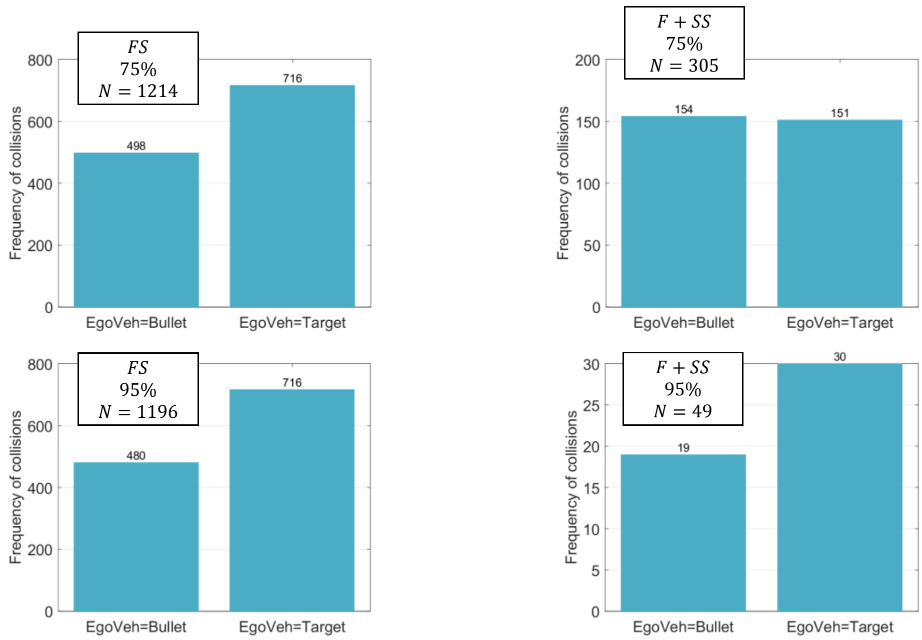


Figure 120: Frequency of collisions for the ego car as bullet or target vehicle, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

6.9.6 Influence of Monte Carlo parameters on injury severity

The three Monte Carlo parameters surface friction, lateral position and approaching speed of the opponent car were further analysed used MATLAB's curve fitting toolbox. Figure 121 depicts the fitted curves for the relationship between friction, approaching speed and slight injury probability. The regression method 'lowess' (locally weighted scatterplot smoothing, Cleveland, 1981) was used to fit a surface to the simulation data that led to a collision. The smoothing process is called local, because each smoothed value is determined by neighbouring data points. A linear polynomial is fitted using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. It is a non-parametric method and it does not produce a regression function in terms of a mathematical formula, which can be easily transferred to others. However, the curve can be easily reproduced when the data and the two model parameters are available, namely the span and the type of polynomial model.

The span defines the percentage of the total number of points used for the regression weight function. Reducing the span makes the surface follow the data more closely.

For the given dataset, a span of 10 percent was used for the FS (75%), FS (95%) and F+SS (75%). Due to the lower sample size, the span for F+SS (95%) was increased to 25%. For all cases, the polynomial model was set to linear.

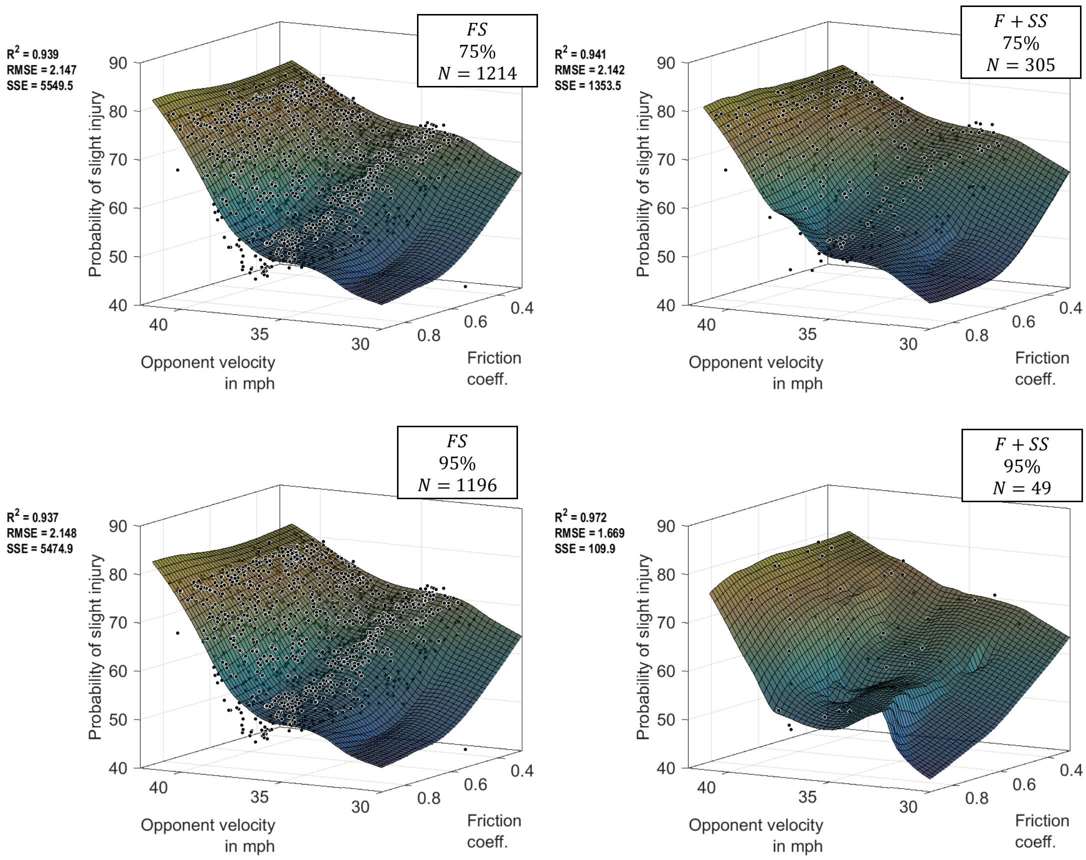


Figure 121: Fitted probability curves for slight injury by varying friction and opponent velocity, for collisions that occurred with forward-facing sensors (left) and additional sideward-facing sensors (right)

The fitted surfaces show that a higher opponent velocity leads to a higher probability to suffer slight injury, which is not surprising. Also, there is a slight increase of injury probability with a lower friction coefficient, especially at a lower velocity around 35 mph. At speeds higher than 40 mph, this increase is negligible. The coefficient of determination R^2 ranges from 0.937 to 0.972, which is indicates a good approximation of the curve. The root mean square error $RMSE$ represents the sample standard deviation of the differences between the regression curve and the actual data points. Also, the values from 1.67 to 2.15 percent are satisfactory. The SSE (Sum of Squares of Error) measures the total deviation of the observed values from the fit. It ranges from 109.9 to 5,549.5. However, for the given regression application, the SSE value is not as useful as it would be when comparing different regression models.

The curves obtained for the serious and fatal injury probability showed similar results and are therefore not presented here by additional plots. A regression analysis including the lateral position resulted in an unsatisfying goodness of fit, because its relation to injury probability appears to be arbitrary and not explanatory.

6.9.7 Analysis of conflict indicators

The previous sections presented the safety performance indicators for simulations that resulted in a collision. This section is devoted to the conflict indicators, which were the decision criteria to differentiate between an undisturbed and a critical situation. As explained by the flowchart in Figure 63, a conflict is first determined by the jerk value, which indicates a critical braking manoeuvre. In case of an increased jerk threshold, the speed driven right before braking is related to the time-to-accident in order to determine if it is a conflict or not (see Figure 72).

Figure 122 presents this TTA-conflict speed threshold function, which was applied to the simulation outputs. Note that the N given in the plots does not correspond to the number of conflicts. The number of blue points corresponds to the number of situations, where the ego car performed a harsh braking manoeuvre to avoid a collision. Hence, due to the random distribution of detection probability within each of the four combinations, the actual number of braking events (N) vary. The red line represents the threshold, with all points on the left considered as critical conflicts (see Figure 72). It is not to be confused with a regression line.

The mean value of TTA differs between the four parameter combinations and ranges from 0.53 (FS, 75%) to 0.87 seconds (F+SS, 95%). The ego car with an additional sideward-facing sensor resulted in a mean TTA value that is almost twice as high as the mean value for the forward-facing sensors. The smaller the TTA, the more critical is the situation. Consequently, the FS simulations resulted in more critical braking situations than the F+SS simulations.

In fact, the distributions of samples in Figure 122 show two clusters, which are swapped between FS and F+SS. The lower cluster around the conflict speed of 2.5 mph is larger in the FS simulations, while the upper cluster around 5 mph is larger in the F+SS case. This observation reveals that there are two main types of braking situations for the ego car. First, a braking manoeuvre performed at around 2.5 mph, which corresponds to the acceleration phase right after the standstill. Second, a braking manoeuvre performed at around 5 mph, which means that the car braked when the turning manoeuvre was almost completed. In the first type, car A enters the junction but immediately detects B. In the second case, car A detects B at a later stage of the turn and the AEB controller performs the braking.

The conflict speed vs. TTA curve for the opponent vehicle B shows a different picture (see Figure 123). Note that the x- and y-scales were adapted for better readability. There is only one main cluster observable from the plots, which ranges from a TTA of approximately 2 to 3.5 seconds with mean values around 2.6 seconds for all four combinations. The plots also reveal that the overall number of braking manoeuvres reduce from 371 for FS (75%) to 25 for F+SS (95%). In other words, the additional sideward-facing sensor decreases the number of safety-critical events for B.

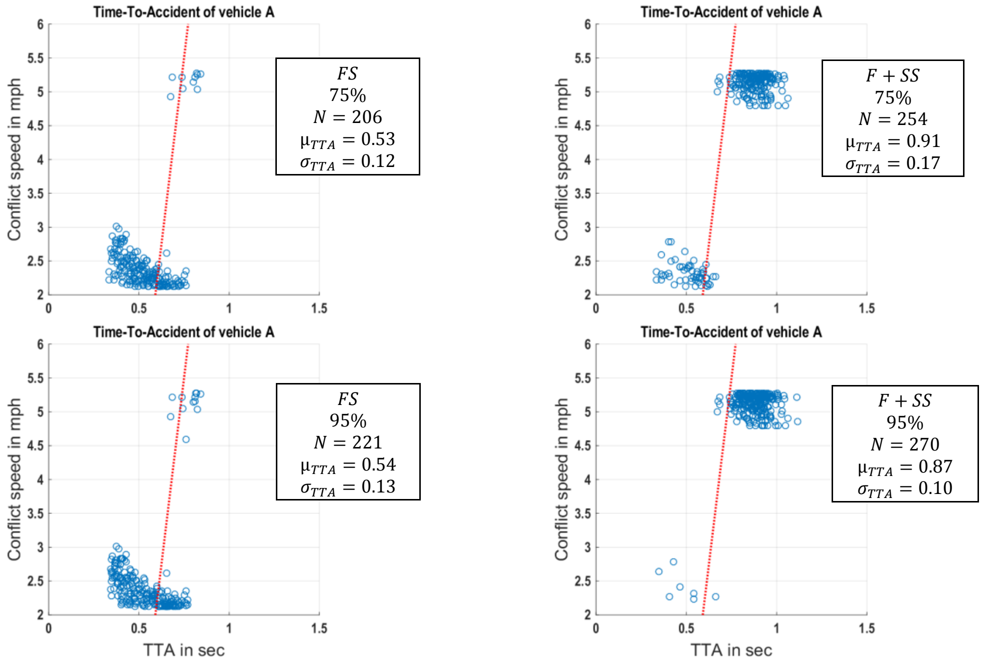


Figure 122: Conflict speed vs. *TTA* of the ego car, with forward-facing sensors (left) and additional sideward-facing sensors (right) and the red line representing the threshold function for critical events

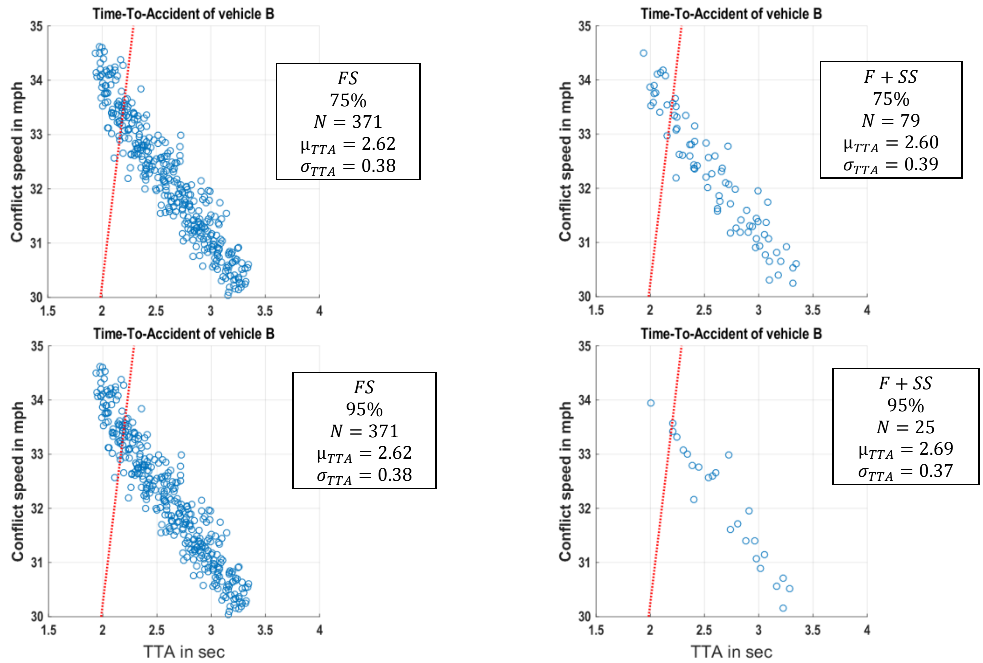


Figure 123: Conflict speed vs. *TTA* of the opponent car, with forward-facing sensors (left) and additional sideward-facing sensors (right) and the red line representing the threshold function for critical events

If a simulation run did not show a critical braking manoeuvre, the *PET* is taken as additional conflict indicator. If the *PET* falls below one second, the situation is defined as conflict. Figure 124 presents the distribution of the *PET* for the four combinations. Note that the plots show the values for all detected conflicts, i.e. also those where the *TTA* vs. conflict speed threshold was exceeded. This explains why the plots contain *PET* values higher than 1 second.

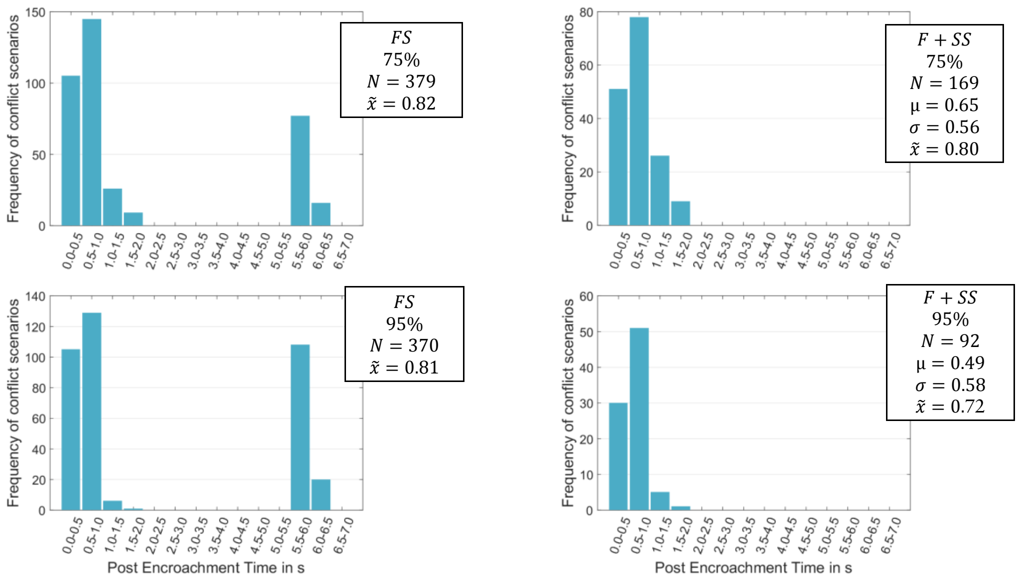


Figure 124: Frequency of conflicts by PET, for simulations with forward-facing sensors (left) and additional sideward-facing sensors (right)

In fact, the FS simulations reveal two clusters of conflicts. First, situations with the PET ranging from 0 to 2 seconds, i.e. with a close encounter of the cars. Second, conflicts with a PET between 5 and 6.5 seconds, where there was enough time between one vehicle entering and the other one leaving the conflict zone, but where one of the vehicles performed a critical braking manoeuvre. Otherwise, it would not have been classified as conflict. Due to this bipolar distribution, the median $\tilde{x}$ is given instead of the mean and standard deviation.

The second cluster of conflicts does no more exist in the F+SS simulations, as the plots on the right side clearly show. The cause for this is the improved capability of the ego car to detect the approaching opponent car by the sideward-looking sensor, which results in the ego car waiting for an appropriate gap. In the FS simulations, the ego car happens to overlook the approaching car and may perform an emergency braking as soon as the forward-facing sensor detects the opponent.

Another time-related conflict indicator is the minimum TTC, which is independent of braking manoeuvres. The minimum TTC gives the time required for the two cars to collide if they were to continue their speed and path. A threshold of 1.5 seconds was defined in Section 6.5.3.1, under which the situation would be defined as critical. It must be mentioned that the minimum TTC was not used in the evaluation algorithm to determine a conflict. However, it is still an interesting measure to analyse the simulation results.

Figure 125 depicts the frequencies of minimum TTC combined for both cars in the four combinations. The number of samples corresponds to the number of non-collisions, i.e. both conflicts and undisturbed situations. The mean values for the FS simulations are 2.08 and 2.06 seconds, respectively. For the F+SS simulations, those values increase to 2.17 and 2.18 seconds, which indicates fewer critical situations in general. For both FS and F+SS, there is a high number of runs with a minimum TTC

below 1.5 seconds. They are represented in the TTA plot in Figure 122, because at such a low TTC value, the braking controller normally activates an emergency braking.

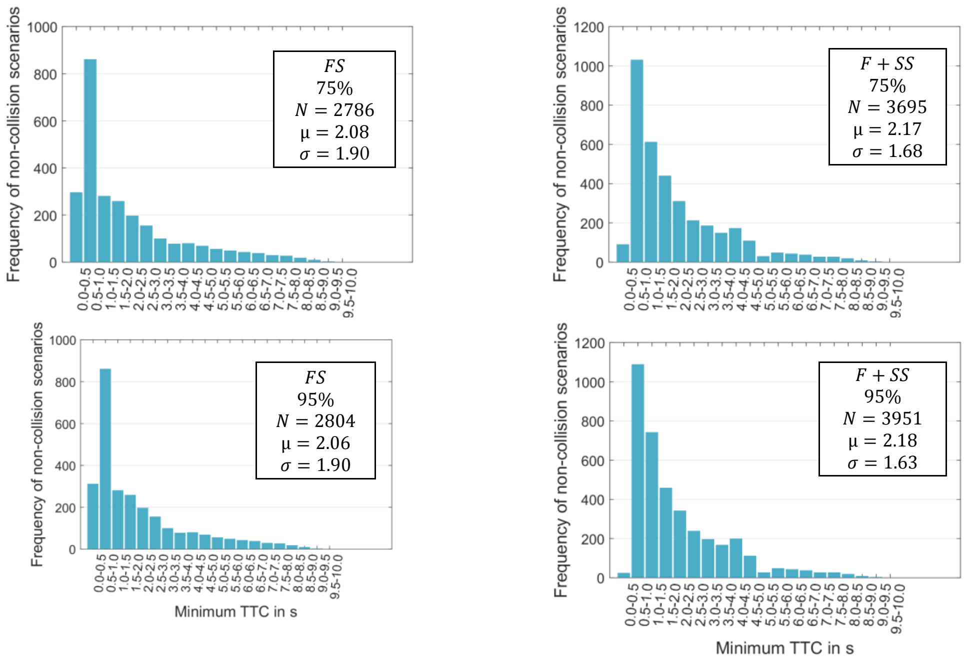


Figure 125: Frequency of non-collision scenarios by minimum TTC, for simulations with forward-facing sensors (left) and additional sideward-facing sensors (right)

In summary, the conflict evaluation complements the collision indicator analysis by revealing situations that might have led to a crash if there was no evasion action taken by one of the vehicles. The analysis has shown the versatility of the simulation framework to understand the safety problems in the particular scenario. The evaluation method is applicable to other scenarios as well and depending on the test objectives, it can be enhanced with additional safety indicators or more advanced impact severity estimations.

6.10 Conclusions

This chapter presented a simulation and evaluation framework for assessing the safety performance of automated vehicles at junctions. The third study of this thesis followed up on the scenario identification of study 2 by transferring the information from the accident data to a sub-microscopic simulation environment based on the CarMaker software combined with MATLAB/Simulink models. Basically, the simulation framework consists of a road environment model, vehicle and sensor models as well as driving control and behaviour models. The output of the simulations is analysed by an evaluation script that computes a set of quantitative safety performance indicators. Therefore, a novel approach denoted as conflict hexagon was developed.

One challenge to create representative simulations from accident data is the lack of information due to incomplete crash records. Not all of the information is available that is needed to model the static and dynamic content. The proposed methodology fills this gap by enhancing the collision scenarios with representative variations of real-world conditions, sampled by a Monte Carlo approach. Those variables included the approaching speed and lateral position of the opponent vehicle in the lane as well as the road surface friction, which has an influence on the braking distance. The ego car represents the automated car and is equipped with in-vehicle sensors to detect objects within its sensing range.

The applicability of the simulation and evaluation framework was demonstrated by an experiment based on a selected collision scenario from study 2. In particular, a right-turn manoeuvre by an automated car at a non-signalised T-junction was evaluated, where another car is crossing from the right with the right of way. The study focused on the comparison of two different in-vehicle sensor systems (basic FCA vs. crossing and turning assistance system) with respect to their crash avoidance capabilities (75% vs. 95% detection probability). For each of the four combinations, a number of 4,000 simulations were performed, all of which have an equal sampling distribution of the Monte Carlo variables for a proper comparison.

The safety performance evaluation resulted in 1,214 collisions and 379 conflicts for the worst-case setting compared to 49 collisions and 92 conflicts for the best case. For 95 percent detection probability, the number of collisions and conflicts were reduced to a tenth, when adding a crossing and turning assistant (based on sideward-facing cameras) to a basic forward collision avoidance system (based on forward-facing cameras only). This obviously demonstrated the safety performance increase in this particular junction scenario. By using additional sideward-facing sensors with 95 percent detection probability, 96.5 percent of all simulations resulted in no collision or conflict.

A detailed analysis of the resulted collisions revealed that the injury probability strongly decreases from slight (mean value around 67 percent) to serious (mean value around 4 percent). The main reason for this can be found in the relatively low impact speeds of the ego car and the resulting low delta of velocity, which is used to estimate the injury severity. Apart from the fact that the frequencies strongly declined, it was found that the different sensor types and detection probabilities do not necessarily influence the distribution of injury probability. Instead, the varying Monte Carlo parameters do. For example, a fitted probability curve showed that a higher opponent velocity leads to a higher probability to suffer injury and that there is a slight increase of injury probability with a lower friction coefficient.

The conflict indicator analysis also confirmed the hypothesis that a forward collision avoidance system is not sufficient to avoid collisions with crossing traffic and therefore needs additional side collision assistance. Simulations with forward collision avoidance only resulted in more critical braking situations than with an additional crossing and turning assistant. In general, the additional sideward-facing sensor system decreases the number of safety-critical events for both vehicles.

In summary, the presented framework eases the modelling and simulation of critical scenarios, which leads to more efficient workflows when virtually testing automated driving. Limitations include the simplified, inelastic computations of impacts and the fact that only primary collisions are considered. Furthermore, the computation efficiency is subject to future work to increase the number of simulation runs and parameter combinations. However, due to its modular architecture, the framework can be adapted to other collision scenarios and may be enhanced with customised models. The indicators in this study could be used as pass/fail criteria for future vehicle tests as part of the overall homologation and certification process.

7 Summary, discussion and conclusions

This chapter brings all three studies together and synthesises the results and the contributions to the state of the art. It further discusses the drawbacks of the chosen methods and states, which results could not be achieved and why. The chapter also points out the study results that should be used with care, as they might not be applicable for all cases. Ultimately, recommendations for future research are made.

7.1 Focus of the thesis

The aim of this dissertation was to develop a modular framework that simulates automated driving at intersections and evaluates the safety in mixed traffic situations with conventional vehicles. The scope is limited to road intersections involving all types of three- and four-legged at-grade junctions excluding roundabouts. The research was structured as follows:

- *Study 1*: The initial study was conducted to set the research scope according to relevant research gaps as well as challenges and problems for automated traffic at intersections. These were identified by reviewing literature, analysing macroscopic accident data and conducting an online expert survey.
- *Study 2*: The follow-up work clustered and evaluated safety-critical scenarios at road junctions from historical accident data, which can pose a particular safety problem involving automated cars.
- *Study 3*: A simulation and evaluation framework was developed and demonstrated, which allows examining the safety performance of automated driving systems within those scenarios.

The thesis included five research questions, which were addressed as follows.

RQ1. Which technical hurdles and challenges do ADS currently have to overcome?

Many prototypes of automated driving systems have been developed and demonstrated in recent years, but the safety risks coming with a mixed vehicle population, namely traffic with both driverless and driver-operated vehicles are still subject to research. Research gaps and challenges were identified by reviewing literature in various areas and by surveying experts with an online questionnaire. After that, a preliminary analysis of junction accident data helped to understand the main safety problems and accident circumstances at junctions.

RQ2. What are the state-of-the-art technologies to enable automated driving in junction environments?

Automated vehicles employ various technologies to drive safely and efficiently through road traffic. There have been numerous developments to facilitate a safe interaction of automated vehicles and non-automated road users at intersections, which can be either vehicle-based or infrastructure-based. A comprehensive review of literature was conducted to categorise intersection assistance systems and to discuss their strengths and weaknesses.

RQ3. What are the current collision scenarios at three- and four-legged at-grade junctions and how can they be clustered from historical accident data?

Given the challenges of ADS, there is still a need for comprehensive testing, either in virtual environments or on real-world test tracks. The challenge is to find the benchmark driving situations to be evaluated at junctions. By applying a novel clustering technique on historical in-depth accident data from the UK, a set of key crash scenarios at junctions was identified and described.

RQ4. How can those collision scenarios be represented and enhanced for sub-microscopic simulation to evaluate the safety performance of intersection assistance systems?

For testing assisted and automated driving functions, virtual testing is being used by car and sensor manufacturers, as it can decrease costs in the development cycle. In sub-microscopic traffic simulations, single vehicles are represented by detailed physical models and their interaction with the road environment and other road users can be investigated. This thesis presented a framework to transfer the derived collision scenarios to a sub-microscopic simulation environment, where the safety performance of automated driving functions was evaluated for a particular demonstration scenario. The study focussed on car-to-car scenarios, where the ego car is highly automated (SAE Level 4/5) and the opponent car is driven by a human.

RQ5. What general recommendations can be made for the safety performance indicators to be considered in virtual testing of ADS at junctions?

There is little research on the best combination of indicators to evaluate the safety performance in virtual vehicle tests at junctions. Existing approaches to estimate collision and conflict risk must be adapted to the particular junction simulation scenarios. As part of the developed framework, this thesis provides recommendations for metrics to quantify safety performance, based on a review of existing literature and tests conducted in the demonstration experiment.

7.2 Novelty and contribution to the state-of-the-art

As the research questions imply, this thesis investigates how collision scenarios can be derived from accidents in driver-operated traffic and applied to virtual, simulation-based testing, where an automated driving system replaces the driver. At the moment, there are limited European regulations on validating the reliability of highly automated road vehicles. Standardized procedures for evaluating automated driving systems are highly relevant in order to guarantee high safety in a varying environment. To this end, this research will contribute to the development and approval of novel vehicle environment recognition systems (e.g. visual or LIDAR-based systems) or infrastructure-based intersection assistance systems by providing the following outputs:

Output #1: Evaluation scenarios for testing ADS at intersections	
Novelty	A novel method to derive pre-crash scenarios from historical car accident data is presented, taking into account different intersection characteristics as well as interplay with non-automated vehicles. It employs k-medoids to cluster historical junction crash data into distinct partitions and then applies the association rules algorithm to each cluster to specify the driving scenarios in more detail.
Possible applications	A predefined set of critical scenarios allows cost-efficient testing, both in virtual and real-world test environments. The scenarios obtained will help to reduce the possible number of model parameter variations, such as vehicle trajectories, velocities as well as road and junction parameters. This leads to a faster development cycle for the automotive and supply industry as well as for manufacturers of infrastructure-based intersection assistance systems. Furthermore, the work leaves an impact on research and development in the fields of road safety and infrastructure.

Output #2: A sub-microscopic simulation framework to virtually reconstruct and evaluate collision scenarios	
Novelty	The framework uses realistic models for vehicles and their sensors as well as accurate road environment models and applies the Monte Carlo method to sample a representative set of parameter combinations, including information from the accident data as well as assumptions on critical factors for ADS.
Possible applications	The framework eases the modelling and simulation of critical scenarios, which leads to more efficient workflows when virtually testing automated driving. Due to its modular architecture, the framework can be adapted to the individual needs of future users and may be enhanced with customised models. The results of this thesis can therefore be used by the automotive industry to further test and develop vehicle sensors and automated driving systems, either by MIL, SIL or HIL. For example, the Simulink model blocks can be replaced by real hardware components or individual parameter combinations may be tested. Due to the principle of parametric variation and scenario generation, ADS can be exposed to a large number of critical situations and their safety performance can be optimised towards better detection probability and crash avoidance capability.

Output #3: Safety performance indicators as metrics to quantify impact	
Novelty	A combination of indicators (temporal and spatial proximity metrics) was used to quantify the impact of in-vehicle intersection assistance systems on

collision and near-miss risk. Those indicators are a direct output of the simulation framework. A novel approach denoted as conflict hexagon was developed to compute the numerical indicators.

Possible applications One step towards uniform criteria for approving automated vehicles for public roads is the definition of pass/fail criteria for tests. This thesis provides the baseline for such criteria by delivering safety performance indicators, which are then used to calculate the overall collision and conflict probabilities.

Furthermore, the indicators applied in the study provide the basis for a-posteriori conflict studies. This is not only relevant for virtual test simulations, but also for the analysis of real-world video observations or naturalistic driving data. The indicators can be further used to estimate the societal benefits from introducing ADS, such as reduced accident costs. Considering the current progression of automated transport, the results of this research are highly relevant, both for virtual as well as real-world testing environments.

7.3 Key results and findings

This section summarises the most important findings and results for each of the research questions.

Findings related to RQ1: Which technical hurdles and challenges do ADS currently have to overcome?

- According to the web survey conducted, the main challenges found for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather conditions. Road surface characteristics, road alignment and lighting were rated as minor influencing factors. Given the survey results, it can be concluded that intersections with poor visibility due to obstructions or bad weather, or unclear lane markings and traffic signs pose particular problems.
- As the literature review concludes, there is a clear need to improve in-vehicle environment detection and perception systems (cameras, LIDAR etc.) in adverse weather conditions. Sensor fusion techniques are being used to compensate for the weaknesses of single sensors, but they still need further testing to function reliably in complex and challenging traffic situations. In particular, predicting other road users' trajectory and movements will be a main part of future research.
- Testing procedures for ADS must be harmonised to prepare for a uniform certification and homologation process. Currently, there are numerous projects dealing with virtual or real-world testing, but uniform solutions are preferable. SAE is currently preparing the standard J3092, which will contain dynamic test

procedures for ADS on a test track. However, there are currently no standardised procedures for virtually testing ADS, especially not for intersections.

- Another main challenge as essential part of the testing procedure is to identify the testing scenarios that represent an application-specific typical or critical combination of actions or events in a traffic situation. In current approaches, these so-called benchmark scenarios can be found by analysing real-world driving data, accident data or risk analysis methods.
- Regulations such as the Vienna Convention on Road Traffic and UN regulation 79 have been amended. Test licenses have been granted in several countries for specific test sites. However, there are still legal hurdles to allow the operation of automated vehicles on public roads.

Findings related to RQ2: What are the state-of-the-art technologies to enable automated driving in junction environments?

- Intersection collision avoidance and mitigation systems can be distinguished into vehicle-based, infrastructure-only and cooperative I2V/V2I systems. Despite the many projects and research studies in the field, there is still no commercially available intersection assistance system for automated vehicles on the market.
- Forward Collision Avoidance systems are the most advanced driving assistance systems available to avoid or mitigate collisions at junctions. Current solutions are primarily tailored to rear-end and pedestrian collisions. Detecting and avoiding angle or turning collisions must be a key feature of ADS.
- An intersection collision detection system must be able to adapt to different intersection types. It has been found that most studies do not validate their models on a large variety of intersection layouts and characteristics.
- Various collision detection algorithms have been proposed, but the coverage of these algorithms is limited to only a few scenarios. Validation and verification of the systems must take into account the most critical combinations of collision parameters.
- Instead of solely focussing on collisions, it is also recommended to consider near-misses as threat measure. Near-misses do not lead to a collision, but may result in discomfort for vehicle passengers and should therefore be avoided by ADS.

Findings related to RQ3: What are the current collision scenarios at three- and four-legged at-grade junctions and how can they be clustered from historical accident data?

- In the 1,056 junction accidents investigated, more than three-quarters of the involved persons were car driver occupants, of which 38 percent were injured. The road user group showing the highest injury rate are pedestrians. In general, car occupants are among the safest road users, together with HGV occupants.

- An initial descriptive analysis of the causation factors showed that a failure to give way, to stop and to avoid are the top three precipitating factors for junction accidents.
- The k-medoids method was found to be a valid technique for the given dataset and resulted in satisfactory partitioning quality, which was measured by the average silhouette value. The clustering exercise resulted in thirteen crash clusters for T-junctions and six crash clusters for four-legged junctions, which were further analysed by association rules.
- The scenarios obtained from the association rule method revealed seven high-injury scenarios (i.e. with serious or fatal injury) for T-junctions and five high-injury scenarios for four-legged junctions, which mainly involve a failure to give way.
- There are no rear-end collisions included in the set of high-risk scenarios. This is due to the fact that the injury outcome was found to be lower for rear-end collisions than for angle collisions.
- The high-frequency scenarios at three-legged junctions do not include any of the high-injury scenarios. They include two rear-end collisions, which are not included in the high-injury scenarios.
- There was no scenario found involving car-pedestrian or car-bicycle collisions, which can be explained by the low number of vulnerable road users among all involved persons in the given accident dataset.

Findings related to RQ4: How can those collision scenarios be represented and enhanced for sub-microscopic simulation to evaluate the safety performance of intersection assistance systems?

- This thesis presents the first study that developed and demonstrated a simulation and evaluation framework for assessing junction safety involving automated cars. It includes the creation of models for the road environment, vehicles, sensors, driver behaviour and collision avoidance systems as well as a comprehensive evaluation procedure based on safety performance indicators.
- The study has shown that not all of the data needed to create the models can be derived from the accident data. Therefore, missing information was supplemented by assumptions and was enhanced with representative variations of real-world conditions, sampled by a Monte Carlo approach.
- The applicability of the simulation and evaluation framework was demonstrated by an experiment based on a selected collision scenario from study 2. In particular, a right-turn manoeuvre by an automated car at a non-signalised T-junction was evaluated, where another car is crossing from the right with the right of way

- The number of collisions and conflicts were reduced to a tenth, when adding a sideward-facing sensor to the automated vehicle's crash avoidance system. This clearly demonstrated the safety performance increase in this particular junction scenario. By using additional sideward-facing sensors with 95 percent detection probability, 96.5 percent of all simulations resulted in no collision or conflict.
- The conflict indicator analysis of the demonstration study confirmed the hypothesis that a forward collision avoidance system is not sufficient to robustly avoid collisions with crossing traffic and therefore needs additional side collision assistance. Simulations with forward-facing collision avoidance only resulted in more critical braking situations than with additional sideward-facing sensors. In general, the additional sideward-facing sensor decreases the number of safety-critical events for both vehicles.

Findings related to RQ5: What general recommendations can be made for the safety performance indicators to be considered in virtual testing of ADS at junctions?

- For different collision scenarios, different combinations of safety performance indicators are recommended. For example, TTC calculation varies for different crash types. Therefore, the evaluation algorithm was developed for angle collisions, rear-end and head-on collisions separately.
- In general, it is recommended to analyse both the collision and the conflict risk as part of the safety performance evaluation. The traffic conflict technique is a well-recognised supplement to traditional crash analysis and has been applied in various studies related to intersection safety. However, there is still no clear evidence on the relationship between crashes and conflicts, but taking into account near-misses gives a more complete picture and allows to analyse events that might have led to a collision if there was no evasion manoeuvre.
- The collision analysis should include an estimation of the injury severity. The sub-microscopic simulation environment used in this study does not include a physical impact or deformation computation. Collision severity was therefore estimated by impact angle and speed relationships.
- To identify if there was a conflict or not, a combination of several indicators is recommended. First, the analysis of the longitudinal jerk indicates a critical braking manoeuvre, which is then used to compute the time-to-accident. The TTA is further related to the speed driven before the braking and a threshold function is used to determine if it was a critical braking event or not. If there was no braking at all, the post encroachment time is used as an indicator to identify close encounters of the involved road users. A PET smaller than 1 second indicates a conflict.
- Instead of using a single point of conflict, it is recommended to define a zone that depends on the vehicle dimensions. In the study, this zone is spanned by a conflict

hexagon, which is used to obtain the projected and actual timestamps, when the vehicles enter or exit the potential area of collision. The hexagon principle can be used for both TTC as well as PET calculations.

7.4 Discussion on generated scenarios for automated driving

The validity of a virtual test of an automated vehicle is strongly related to the underlying knowledge base (Helmer et al., 2015), which the testing scenarios and simulations models are derived from. This knowledge base can be built from a variety of sources, such as previous field operational tests, naturalistic driving studies, expert opinions, surveys or historical accident data. Ideally, testing scenarios are identified from a combination of all of those sources, although the synthesis could be challenging. A database-driven method to follow this holistic approach was developed by Pütz et al. (2017) within the project PEGASUS. This thesis focused on one aspect of this knowledge base, namely the analysis of accident data to identify critical testing scenarios. Hence, instead of spanning the number of possible scenarios at junctions by all variations in traffic situations and environment conditions, the study focussed on the identification of relevant existing, hazardous situations.

The method of clustering intersection crashes into distinct groups, including such a high number of variables as used in this thesis, is novel. However, there has been much research using other methods for a similar purpose. For example, Abdel-Aty et al. (2006) analysed numerous parameters to identify crash profiles for 45 different intersection configurations in Florida. However, this was made for different AADT values and numbers of lanes, which were not included in this study. Also, the objective of this thesis is different, because it aimed at extracting relevant combinations of junction situations for simulation, while Abdel-Aty et al. (2006) provided crash profiles that assist in identifying intersections with specific problems. Therefore, the results cannot be directly compared.

Most existing research on intersection scenarios focussed on the classification of pre-crash manoeuvres, not combined with parameters about the road environment, collision partners, points of impact, injury types, causation factors and traffic control. Compared to literature, this thesis study can be seen as more detailed in terms of crash circumstances. In the European INTERSAFE project (INTERSAFE, 2005), intersection accidents were classified according to the pre-crash driving manoeuvres (in right-hand traffic). Twenty intersection situations were identified, of which the top five were: 1) Vehicle A crossing path, with vehicle B coming from the left or right (which corresponds to the high-injury scenarios X-1.1, X-2.1 and X-6.1), 2) A turning left into the path of B coming from the left (see X-4.1), 3) A turning across the path of B coming from the opposite direction (see X-6.2, T-4.1, T-13.1), 4) A turning right into the path of B coming from the left (see T-12.3) and 5) A hitting the rear of B waiting to turn left (see the high-frequency scenarios T-1.1, T-1.2, T5.1).

The TRACE project identified six different scenarios at four-legged intersections from a statistical analysis of crashes in the European Union (Molinero Martinez et al., 2008). The scenario where A crosses the road and the trajectory of the opponent vehicle B, which is turning or going straight, is more frequent and more severe than any other. 70 percent of all intersection accidents belong to that scenario. This corresponds to the most frequent scenarios X-1.1, X-4.2, X-2.1 and X-5.1, from which X-1.1 was also found as one of the high-injury scenarios.

Of all intersection-related crashes analysed by Choi (2010), about 96 percent had critical reasons attributed to drivers, while critical reasons related to vehicle or environment were assigned in less than three percent of these crashes. Wiltschko (2004) concludes that intersection assistance systems must be particularly designed to avoid red-light violations and fail to give way. This is also confirmed by this thesis since failures to give way are a precipitating factor in most scenarios.

Kim et al. (2017) evaluated collision warning systems for intersections based on scenarios derived from naturalistic driving data. They identified sixteen vehicle-to-vehicle accident scenarios and studied the accident prevention capabilities of camera-based and radar-based collision warning in those scenarios. The study primarily focussed on vehicle movements and did therefore not differentiate between different junction types and does not include information about traffic control or speed limits. A drawback of their study is that it considers only one safety indicator called safety-remaining distance and no conflicts, and the interaction with vulnerable road users is not included. In comparison to their work, this thesis uses accident data instead of naturalistic driving data. The latter is certainly interesting for deriving scenarios from a wider perspective, and a systemic approach including several sources of information for generating scenarios is recommended for the future.

It can be seen that an examination of the results from study 2 in relation to existing research shows similarities, although a direct comparison is difficult due to varying data sources and methods used. The clusters and scenarios obtained certainly delivered useful information to proceed to study 3, the virtual assessment. But it is also worth to discuss, if the results can be generalised for other purposes and future studies. A comparison of the accident data sample with UK data as given in the CARE database resulted in significant similarities in terms of the proportions of accident cases by junction type as well as by injury level. Thus, although this study was conducted in a specific area in the UK, the results should be generalizable to other areas. However, since the data is based on left-hand traffic, the direct transferability of the obtained scenarios to other countries should be validated with additional data.

A new challenge for generating virtual testing scenarios is the identification of yet unknown, automation-specific scenarios, which might be hidden in the whole spectrum of situations (Helmer et al., 2015). A limitation of study 2 is that the scenarios identified are based on human-related crash situations and do not necessarily reflect critical situations that come with sensor failure or misinterpretation of the automated driving control. Assuming that the car operates automated, some scenarios such as rear-end crashes might be avoided to a great extent by reliable environment perception and motion planning, but there might be new hazards caused by automation.

Certainly, there may be different key testing scenarios depending on which issue is targeted. For example, targeting at maximum casualty reduction for vulnerable road users will require different testing measures than targeting at the vehicles' full functionality. The pre-crash scenarios derived from this study build the foundation for further research on testing assisted and automated vehicle technologies. The study focussed on the scenarios with serious or fatal injury outcome, which were compared to the high-frequency scenarios. Although there is no doubt about the importance of vulnerable road user safety, neither the cluster analysis and thus nor the association rule method resulted in a distinct pedestrian or cyclist scenario. Considering the frequency of certain crash types at junctions, car-pedestrian and car-cyclist collisions are discounted, which might not be true if injury frequencies were taken into account. A larger sample size or additional accident data sources may have to be added to derive testing scenarios for vulnerable road user interaction.

A final issue to mention with regards to the generated scenarios is that they might change over time as road user behaviour will probably change with increased automation. This means that some situations currently encountered in traffic will become rarer, while there might be new situations coming up due to behavioural changes. For instance, new critical scenarios may arise from misunderstandings between human road users and automated vehicles. Consequently, a sound database of scenarios should be regularly updated with new knowledge and data.

7.5 Discussion on virtual assessment of safety performance

There are strong arguments that vehicle automation will have a positive impact on overall safety. However, those arguments are not yet sufficiently proven by quantitative figures (Helmer et al., 2015). In this context, virtual testing plays a vital role in demonstrating this positive impact by using appropriate indicators. Numerous research projects have been conducted to virtually assess the safety performance of automated vehicles (Beglerovic et al., 2017; Olivares et al., 2016; Pütz et al., 2017; Rodarius et al., 2015; Roesener et al., 2017). However, those studies primarily focussed on automated driving functions on the motorway, but did not particularly address junction safety. Kim et al. (2017) evaluated collision warning systems for intersections based on scenarios derived from naturalistic driving data and accident records. However, their study focuses on accident prevention rates and considers only one safety indicator and no conflicts. This thesis is the first work, to the author's knowledge, which thoroughly generated critical situations for testing automated driving at junctions, and which developed a modular simulation and evaluation framework to allow the assessment of various safety indicators for those junction scenarios.

The virtual experiments conducted within this thesis provide a quantification of the effectiveness of different characteristics of automated intersection assistance systems. Those systems are assumed to be a core element of future automated vehicles to ensure safe and efficient operation, e.g. in urban areas. The intersection assistance systems were tested in four variations, which differ in the type of vehicle sensors (forward-facing sensor only vs forward plus sideward-facing sensor) and the quality of object detection within the vehicle's environment (75% vs 95% detection probability). A lower detection probability infers the occurrence of detection failures, which are modelled by stochastic sampling. In reality, these failures may be caused by environment phenomena such as light and shade areas, reflections, sun glare or poor weather conditions such as heavy rain, snow or fog, which negatively influence the object recognition and distance measurement.

The simulation and evaluation framework was demonstrated for one selected collision scenario, for which the safety performance was calculated as the sum of the effectiveness of the system divided by the overall amount of simulations. The effectiveness was quantified by the number of occurred collisions and near-miss events, also called conflicts. Hence, each of the four combinations in this scenario resulted in a collision and conflict probability, which were compared to derive conclusions. For a complete vehicle test, the overall safety performance would result from the sum of the effectiveness in several scenarios weighted by their respective frequency. Since it was not feasible to assess all scenarios obtained in study 2, the conducted simulation experiment does not give a generalised indication of how junction safety will change with increased automation. In fact, this was not the aim of the thesis. Instead, a basic modular framework for future evaluations was built, and one particular scenario was assessed to showcase the framework's applicability and versatility.

Due to its modular architecture, the framework is applicable to other simulation environments and can be enhanced or replaced by other modules. For example, this thesis used CarMaker as simulation environment, because it provides sophisticated vehicle, road and driver models, which can be easily adapted. Basically, this environment can be replaced by other tools as long as they can provide the same output quantities. The developed framework can also be used for other vehicle types, such as automated mini-busses, which must pass junctions. The safety performance indicators and the resulting evaluation process would remain the same.

A novelty of this thesis is the method how to evaluate the safety performance from numerical indicators. A combination of indicators and respective thresholds was implemented for the particular requirements that a junction scenery imposes. This evaluation method can be used for other studies as well, as it comprises a self-contained module with defined inputs and outputs. Parameters such as indicator thresholds (e.g. minimum time-to-collision) can be set according to the needs of the virtual test.

The application of conflict and near-miss indicators in the evaluation procedure may raise a discussion about the relationship between accidents and near-misses. For almost five decades, researchers have been discussing whether near-misses as surrogates are valid measures for safety. On the one hand, there have been numerous publications that presented a foundation for certain crash-conflict ratios. On the other hand, doubts about the validity of surrogate measures were expressed due to unconvincing evidence from observed data. The presented thesis does not provide such conflict validation methods, but includes the number of detected conflicts as overall safety performance measure. For future applications of the presented framework, the expected number of accidents based on the detected conflicts may be taken into account to derive a pass/fail criterion for the tests.

The demonstration experiment proved that additional sideward-facing sensors tremendously increase the safety performance in comparison to forward-facing sensors. One might argue that this finding is not surprising. However, the two sensor systems were additionally varied by the detection probability, and the resulting analysis on indicators gave insights into the influence of detection failure on impact speeds, impact angles as well as time-related conflict measures. Hence, the study resulted in more than just collision and conflict probabilities. If applied to numerous sceneries and scenarios, the results can lead to recommendations on road infrastructure adaptations or advanced collision avoidance methods.

The study has a number of possible limitations. For example, the way of deriving collision indicators such as the delta of velocity and the resulting injury probabilities is based on a simplified approximation following Newton's principle of momentum conservation. Since CarMaker was not designed for collision simulations, effects such as deformations, rotations or rebound of the vehicles after the impact are disregarded. If CarMaker was replaced or complemented by a tool that is able to calculate those effects, the evaluation procedure would achieve a more realistic estimation of injury severity.

Furthermore, the injury risk curves used to define the probability of different injury levels are based on US crash data recorded between 1995 and 1999. Taking into account the safety improvements due to modern vehicle technologies and safety systems in the last two decades, one might argue that the overall injury risk in relation to the driven speed might have changed. An update of those injury risk curves with more recently collected data could improve the accuracy of injury estimation, but this was kept for future work.

Another limitation related to the collisions is that only primary collisions are investigated. Secondary collisions, e.g. caused by the vehicle hitting a nearby roadside object or approaching traffic after the impact, would enhance the risk assessment by an additional factor. However, this would also increase the model complexity. Another way for enhancing the simulations would be to include steering evasion manoeuvres to the collision avoidance models in addition to emergency brakings. This might also lead to collisions caused by excessive steering and instabilities of the vehicle, such as run-off-road accidents.

In comparison to a real-world system, a simulation-based approach has the advantage that the road users' trajectories and velocities are precisely known and do not have to be predicted. Thus, this thesis did not develop a trajectory prediction algorithm, but instead, used the data from a reference simulation to calculate e.g. the time-to-collision. It is significant to mention that this implies an ideal system with a 100% accurate prediction. In practice, a range of possible trajectory variations is computed as a threat assessment procedure for collision warning systems. The presented framework would have to be enhanced if such models wanted to be evaluated, too.

The environment sensor models used in this study imitate a camera vision system for object recognition, with several cameras combined to increase the field of view. Detection failures were injected stochastically to the object detection model to imitate real-world effects such as adverse lighting or weather conditions. In reality, those detection failures are minimised by sensor fusion techniques, i.e. combining image processing methods with radar, ultrasonic and LIDAR data. This thesis did not go so far, since the CarMaker software version used had limitations regarding physical sensor models when used with MATLAB/Simulink. By default, the sensor models used are ideal sensors that detect each object with 100 percent accuracy. This is why a failure injection was implemented. For future studies and future releases of CarMaker, it is planned to replace the ideal sensors with high-fidelity sensor models that can accurately replicate real-world behaviour and weaknesses of sensors.

Regarding the validity of the simulation models, it can be seen as a limitation that the automated vehicle model as a whole was not validated with real-world measurements. It must be noted that all submodels such as tires, suspension system, brakes etc. are well-recognised reference models with validated behaviour, but the entire vehicle system including environment sensors and collision avoidance control might not entirely reflect the real-world behaviour. Future work would be necessary to create an accurate replication of an automated vehicle based on calibration measurements, which was not feasible within this thesis. However, the modular design of the developed framework allows replacing or enhancing the models as required.

Besides the variation of environment detection systems and their accuracy, each simulation run was defined by a unique combination of parameters that were sampled by the Latin Hybercube method. Those parameters were the velocity and the lateral position of the opponent vehicle as well as the road friction coefficient. Other parameters such as the junction configuration or the driving manoeuvres and behaviour of the ego car were kept static. For the objectives of the demonstration experiment, this set of varying parameters was sufficient to assess the performance of the different sensor systems. Actually, the number of variations can be increased, because the Latin Hypercube supports a highly dimensional parameter space to sample from. Interesting other variations could be the driving behaviour of the opponent or the variation of traffic control at the junction.

The limited number of parameter variations was also caused by the relatively high computational requirements. On average, a simulation run needed 12 seconds to be completed while the subsequent evaluation script had an average computation time of 8 seconds. For each of the four sensor systems, 4,000 simulations were conducted. The accuracy of the outcome depends on the number of random numbers used. For an infinite number of Monte Carlo samples, the solution would be exact. Considering that there needs to be a reference simulation for each run, as explained above, the total number of simulations was 20,000. Consequently, more than 100 hours were needed to complete all calculations. It must be mentioned that computation efficiency was not a main goal of the study. Certainly, future work will include measures to improve the performance, e.g. by parallelising simulations. Seidel (2017) used cloud computing for the automated vehicle simulations, which is also a promising way to reduce simulation times.

The study applied a sub-microscopic simulation method to assess safety performance. As explained, the main focus was to investigate the main influence of sensor systems to detect the environment. The role of the passengers in the automated car was disregarded. Driver behaviour models solely defined the automated car's behaviour and scenarios that involve a handover of control to the passenger were not discussed. Without a doubt, the issue of safe handovers and observation of the passenger's state of vigilance are important elements for a future operation of ADS. Numerous driving simulator studies were conducted to examine the best ways for human-machine interaction (Bahram et al., 2015; Beller et al., 2013; Brandenburg and Skottke, 2014; Kircher et al., 2014; Körber et al., 2015; Merat et al., 2012; Naujoks et al., 2014), including different test scenarios where a handover is required. This affects ADS for SAE level 3, where the driver needs to take over in critical situations. However, current developments aim to reach level 4 systems, where the vehicle must be able to perform a safe exit manoeuvre without involving the driver at all. Therefore, this thesis has been limited to level 4 operation without including additional handover scenarios.

There are many other issues to discuss with regards to limitations and opportunities of virtual vehicle tests, because there is a lot of research ongoing in this field. New studies are being published almost daily and by the time this thesis is published, new research will be available to review. There have been lively discussions about the role of virtual testing in the vehicle certification and approval process. Real-world driving tests may become less and less important due to their high required costs and time, and virtual tests may be the main method for type approval, as it is the case today for the electronic stability control (Lutz et al., 2017). Nevertheless, virtual testing methods still suffer less fidelity and overall acceptance in comparison to real-world tests. Certification bodies tend to trust real-world tests more than simulations. This is due to missing evidence on the validity of simulations. This is also true for the results of this thesis, which would have to be confirmed with physical measurements. The collision and conflict situations reconstructed in the virtual environments must be validated with real-world driving data or investigations of selected crashes. Certainly, this is an important issue for future research.

Overall, the methods developed in this thesis and the results obtained can be highly beneficial for future assessment of automated vehicles that will be operated in road areas with at-grade junctions. Especially for non-signalised junctions, ADS are still facing technical challenges, which require comprehensive tests supported by simulation.

7.6 Recommendations for further research

Several issues have been discussed in the previous sections that lead to possible future work. In the light of the results and findings discussed above, two main questions remain to be resolved. First, what are the new risks at junctions that come with an increased penetration of automated vehicles and how must the testing scenarios be enhanced? And second, what are the key criteria and procedures for certification and homologation of ADS?

Regarding the first question, there is a need to identify risks that are related to vehicle automation only. There might be new collision types, which are currently unknown. This is also related to the question how the derived accident scenarios can be merged with scenarios obtained from other sources, e.g. real-world driving data? In the near future, a large amount of data will be available from ongoing pilot tests of ADS. This data can deliver insights and new information about risks and can thus lead to new test conditions and parameters. Further studies are required to enhance current methods for generating validation scenarios with driving data from pilot tests. While this thesis provides a novel method for evaluating junction scenarios from crash data, a holistic, systematic approach is recommended for the next step, which takes into account various sources of data. For instance, this thesis did not result in any vulnerable road user scenarios due to the dataset used. Real-world observations of the interaction between automated vehicles and pedestrians or cyclists at intersections will reveal such scenarios. Furthermore, the scenario generation procedure should be as automated as possible to reduce workload.

The second recommendation that can be made is related to the certification and homologation process for automated vehicles. More research on approval criteria is needed to develop a harmonised process. Currently, there are no standardised procedures for approving ADS and each manufacturer has its own test methods. Certainly, certification will depend on the ADS function to be approved or the areas, in which the vehicle operates. For example, the highway pilot will have different approval requirements than an urban automated shuttlebus. As soon as an automated vehicle has to pass junctions, the approval criteria must include a certain level of safety for the relevant junction scenarios. Further studies are required to establish a uniform process for certification and homologation, which necessarily have to include procedures for junctions, too.

One step towards uniform approval criteria is the definition of pass/fail criteria for automated vehicle tests. This thesis provides the baseline for such criteria by delivering safety performance indicators, which are then used to calculate the overall collision and conflict probabilities. Future work will include the investigation of reasonable thresholds for the probabilities, which can then be used as pass/fail criteria.

A promising approach towards approval of ADS is functional decomposition, as currently applied in robotics or informatics (Amersbach and Winner, 2017). According to that, the driving task is split up into several layers, each of which being tested individually. For instance, information reception can be separated from information processing. While the reception may have failures caused by dirty sensors or errors in receiving C2X messages, the processing could be influenced by inaccuracies in object classification. The understanding of the situation, the decision on which action to take and the final action could be other layers of the decomposition. This approach is expected to reduce testing complexity and might bring answers to the question, where the cut between virtual and real-world testing must be made.

7.7 Overall conclusions

In this thesis, a new validation method has been developed for automated driving systems at road junctions. The method comprises the clustering of critical traffic scenarios at junctions as well as a simulation and evaluation framework to validate those scenarios. The applicability of the framework was demonstrated by an experiment based on a selected car-to-car collision scenario. It could be shown that the number of collisions and conflicts were reduced to a tenth when adding a crossing and turning assistant to a basic forward collision avoidance system. The safety performance indicators selected and implemented in the framework can be seen as a new reference for conducting virtual tests at junctions. The outputs of this thesis lead to a faster development cycle for the automotive and supply industry as well as for manufacturers of infrastructure-based intersection assistance systems. Considering the current progression of automated transport, the results are highly relevant, both for virtual as well as real-world testing environments and are an important step towards a novel certification and homologation process of automated vehicles.

8 References

AASHTO, 2012. A policy on geometric design of highways and streets. American Association of State Highway and Transportation Officials, Washington, D.C.

Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at non-signalized intersections using machine learning techniques. Accident Analysis & Prevention 43, 461–470.

Abdel-Aty, M., Lee, C., Wang, X., Nawathe, P., Keller, J., Kowdla, S., Prasad, H., 2006. Identification of intersections' crash profiles/patterns (Final report). University of Central Florida.

Abid, H., Chung, T.C., Lee, S., Qaisar, S., 2012. Performance Analysis of LTE Smartphones-Based Vehicle-to-Infrastructure Communication, in: 9th International Conference on Ubiquitous Intelligence and Computing. IEEE, pp. 72–78.

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. Acm sigmod record 22, 207–216.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Alessandrini, A., Campagna, A., Site, P.D., Filippi, F., Persia, L., 2015. Automated Vehicles and the Rethinking of Mobility and Cities. Transportation Research Procedia 5, 145–160.

Alexander, L., Cheng, P.-M., Donath, M., Gorjestani, A., Menon, A., Shankwitz, C., 2007. Intersection Decision Support Surveillance System: Design, Performance and Initial Driver Behavior Quantization (Research Report No. Mn/DOT 2007-30), Developing Intersection Decision Support Solutions No. 3. Minnesota Department of Transportation.

Allen, B.L., Shin, B.T., Cooper, P.J., 1978. Analysis of traffic conflicts and collisions. Transportation Research Record.

AlpLab - Die Zukunft des Fahrens, 2017. Steirische Mobilität 1.

Amditis, A., Ghosh, L., 2014. From interactIVe to AdaptIVe: Evolution of automated driving. Presented at the 10th ITS European Congress, Helsinki, Finland.

Amersbach, C., Winner, H., 2017. Functional decomposition. Presented at the 8. Tagung Fahrassistenz, Munich, Germany.

Amundsen, F.H., Hydèn, C., 1977. Proceedings from the First Workshop on Traffic Conflicts. Oslo, Norway.

An, J., Choi, B., Kim, B., Kim, E., Hwang, J., 2014. Rear-end Collision warning system using linear discriminant analysis. Presented at the 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS), pp. 218–221.

Anderberg, M.R., 1973. Cluster analysis for applications, Probability and mathematical statistics, 19. Academic Press, New York.

Andritsos, P., Tsaparas, P., Miller, R.J., Sevcik, K.C., 2004. LIMBO: Scalable clustering of categorical data, in: International Conference on Extending Database Technology (EDBT). Springer, pp. 123–146.

Antonucci, N.D., Hardy, K.K., Slack, K.L., Pfefer, R., Neuman, T.R., 2004. Guidance for implementation of the AASHTO Strategic Highway Safety Plan - Volume 12: A guide for reducing collisions at signalized intersections (NCHRP Report 500: Volume 12). Transportation Research Board, Washington, D.C.

Aoude, G.S., Luders, B.D., How, J.P., Pilutti, T.E., 2010. Sampling-based threat assessment algorithms for intersection collisions involving errant drivers. IFAC Proceedings Volumes 43, 581–586.

Archer, J., 2005. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: a study of urban and suburban intersections (Dissertation). KTH Royal Institute of Technology, Stockholm.

Aria, E., 2016. Investigation of automated vehicle effects on driver's behavior and traffic performance (Master's thesis). Linköping University, Norrköping, Sweden.

Arndt, O.K., 2003. Relationship Between Unsignalised Intersection Geometry and Accident Rates (Dissertation). Queensland University of Technology.

Austroads, 2009. Guide to Road Design Part 4: Intersections and Crossings – General (AGRD04-09).

Automated driving systems to be tested on Tokyo roads from 2017, 2016. . The Japan Times Online.

Aycard, O., Baig, Q., Bota, S., Nashashibi, F., Nedevschi, S., Pantilie, C., Parent, M., Resende, P., Vu, T.-D., 2011. Intersection safety using lidar and stereo vision sensors, in: IV'2011-IEEE Intelligent Vehicles Symposium. pp. 863–869.

Äyrämö, S., Kärkkäinen, T., 2006. Introduction to partitioning-based clustering methods with a robust example (Reports of the Department of Mathematical Information Technology No. 1/2006), Series C. Software and Computational Engineering. University of Jyväskylä, Finland.

Bagdadi, O., 2013. Estimation of the severity of safety critical events. Accident Analysis & Prevention 50, 167–174.

Bagdadi, O., Várhelyi, A., 2013. Development of a method for detecting jerks in safety critical events. Accident Analysis & Prevention 50, 83–91. https://doi.org/10.1016/j.aap.2012.03.032

Bahouth, G., Graygo, J., Digges, K., Schulman, C., Baur, P., 2014. The Benefits and Tradeoffs for Varied High-Severity Injury Risk Thresholds for Advanced Automatic Crash Notification Systems. Traffic Injury Prevention 15, 134–140.

Bahram, M., Aeberhard, M., Wollherr, D., 2015. Please take over! An analysis and strategy for a driver take over request during autonomous driving, in: 2015 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2015 IEEE Intelligent Vehicles Symposium (IV), pp. 913–919.

Baker, W.T., 1972. An evaluation of the traffic conflict technique. Highway Research Record 1–8.

Baldessari, R., Bödekker, B., Deegener, M., Festag, A., Franz, W., Kellum, C.C., Kosch, T., Kovacs, A., Lenardi, M., Menig, C., 2007. Car-2-car communication consortium-manifesto.

Baldock, M.R.J., Long, A.D., Lindsay, V.L.A., McLean, J., 2005. Rear end crashes (Report No. CASR018), CASR Report Series. Centre for Automotive Safety Research.

Basma, F., Refai, H.H., 2009. Collision Avoidance System at Intersections (Final Report No. FHWA-OK-09-06). Electrical and Computer Engineering Department, University of Oklahoma.

Basma, F., Tachwali, Y., Refai, H.H., 2011. Intersection collision avoidance system using infrastructure communication, in: Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference On. IEEE, pp. 422–427.

Bauer, K.M., Harwood, D.W., 1996. Statistical models of at-grade intersection accidents (FHWA-RD-96-125, Final Technical Report). Federal Highway Administration, US.

Becic, E., Manser, M., Creaser, J., Donath, M., 2012. Cooperative Intersection Collision Avoidance System–Stop Sign Assist: Experiments to Validate Use of an In-Vehicle Interface Design (Research Report No. CTS 12-09). Center for Transportation Studies, University of Minnesota.

Beck, D., 2015. Investigation of Key Crash Types: Rear-end Crashes in Urban and Rural Environments (Research Report No. AP-R480-15). Austroads, Sydney, Australia.

Beglerovic, H., Ravi, A., Wikström, N., Koegeler, H.-M., Leitner, A., Holzinger, J., 2017. Model-based safety validation of the automated driving function Highway Pilot, in: 8th International Munich Chassis Symposium 2017. pp. 309–329.

Beller, J., Heesen, M., Vollrath, M., 2013. Improving the driver-automation interaction: an approach using automation uncertainty. Human Factors 55, 1130–1141.

Berg, F.A., Walz, F., Bürkle, H., Epple, J., 1998. Implications of velocity change delta-v and energy equivalent speed EES for injury mechanism assessment in various collision configurations. Presented at the International IRCOBI conference on the biomechanics of impacts.

Berkhin, P., 2006. A Survey of Clustering Data Mining Techniques, in: Kogan, J., Nicholas, C., Teboulle, M. (Eds.), Grouping Multidimensional Data. Springer Berlin Heidelberg, pp. 25–71.

Bifulco, G.N., Simonelli, F., Pace, R.D., 2008. Experiments toward a human-like Adaptive Cruise Control. Presented at the 2008 IEEE Intelligent Vehicles Symposium, pp. 919–924.

Bloecher, H.L., Dickmann, J., Andres, M., 2009. Automotive active safety & comfort functions using radar, in: Ultra-Wideband, 2009. ICUWB 2009. IEEE International Conference On. IEEE, pp. 490–494.

BMVI, n.d. Digitales Testfeld Autobahn [WWW Document]. URL http://www.bmvi.de/DE/Themen/Digitales/Digitales-Testfeld-Autobahn/Digitales-Testfeld-Autobahn.html (accessed 8.14.17).

Bock, Julian, 2017. Die Datenbank als Freigabewerkzeug automatisierter Fahrfunktionen im Kreislauf relevanter Szenarien. Presented at the 8. Tagung Fahrerassistenz, Munich, Germany.

Bostrom, O., Gabler, H.C., Digges, K., Fildes, B., Sunnevang, C., 2008. Injury reduction opportunities of far side impact countermeasures. Annals of Advances in Automotive Medicine 52, 289–300.

Bours, R., Tideman, M., Lages, U., Katz, R., Spencer, M., 2014. Automated Generation of Virtual Driving Scenarios from Test Drive Data, in: Proceedings of FISITA World Congress 2014. Maastricht, The Netherlands.

Brandenburg, S., Skottke, E.M., 2014. Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving? Presented at the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 2978–2983.

Brännström, M., Sjöberg, J., Helgesson, L., Christiansson, M., 2011. A Real-time Implementation of an Intersection Collision Avoidance System, in: Preprints of the 18th IFAC World Congress. Milano, Italy.

Brookhuis, K.A., de Waard, D., 1999. Limiting speed, towards an Intelligent Speed Adapter (ISA). Transportation Research Part F: Traffic Psychology and Behaviour 2, 81–90.

Broughton, J., Thomas, P., Kirk, A., Brown, L., Yannis, G., Evgenikos, P., Papantoniou, P., Candappa, N., Christoph, M., van Duijvenvoorde, K., 2013. Traffic Safety Basic Facts 2012: Junctions.

Brown, T.L., 2005. Adjusted Minimum Time-To-Collision (TTC): A Robust Approach to Evaluating Crash Scenarios. Presented at the DSC 2005 North America, Orlando, USA.

Brunner, H., Krettek, C., Otte, D., Zwipp, H., 2003. Scientific Approach and Methodology of a New In-depth Investigation Study in Germany so called GIDAS. Presented at the The 18th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Nagoya, Japan, p. 10.

Buehler, M., Iagnemma, K., Singh, S. (Eds.), 2009. The DARPA urban challenge: autonomous vehicles in city traffic, in: Springer Tracts in Advanced Robotics. Springer, Berlin.

Burgess, T.F., 2001. Guide to the Design of Questionnaires, Information Systems Services. University of Leeds.

Burn-Callander, R., 2015. This is the Lutz pod, the UK's first driverless car [WWW Document]. URL http://www.telegraph.co.uk/finance/businessclub/technology/11403306/This-is-the-Lutz-pod-the-UKs-first-driverless-car.html (accessed 7.19.15).

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods 3, 1–27.

Candappa, N., Logan, D., Van Nes, N., Corben, B., 2015. An exploration of alternative intersection designs in the context of Safe System. Accident Analysis & Prevention 74, 314–323.

Canu, B., Perrin, C., Magnin, J., Canu, A., Serre, T., Brenac, T., 2016. Interest of in-depth investigation for studying the relation between speed and accident risk. Presented at the 7th International Conference ESAR - Expert Symposium on Accident Research, Hannover, Germany, p. 11.

Carsten, O., 2001. ISA: the Best Collision Avoidance System?, in: Proceedings of 17th Conference on the Enhanced Safety of Vehicles.

CCAV, 2017. UK Connected and Autonomous Vehicle Research and Development Projects 2017. Centre for Connected & Autonomous Vehicles, UK.

Chakraborty, I., Tsiotras, P., Diaz, R.S., 2013. Time-optimal vehicle posture control to mitigate unavoidable collisions using conventional control inputs. Presented at the 2013 American Control Conference, pp. 2165–2170.

Chakraborty, I., Tsiotras, P., Lu, J., 2011. Vehicle posture control through aggressive maneuvering for mitigation of T-bone collisions. Presented at the 2011 50th IEEE Conference on Decision and Control and European Control Conference, pp. 3264–3269.

Chalmers, 2014. Successful inauguration of Astazero [WWW Document]. URL http://www.chalmers.se/en/news/Pages/Successful-inauguration-of-Astazero.aspx (accessed 7.23.15).

Chan, C.-Y., Marco, D., Misener, J., 2004. Threat assessment of traffic moving toward a controlled intersection, in: IEEE Intelligent Vehicles Symposium, 2004. Presented at the IEEE Intelligent Vehicles Symposium, 2004, pp. 931–936.

Chawla, S., Gionis, A., 2013. k-means-: A Unified Approach to Clustering and Outlier Detection, in: Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA. pp. 189–197.

Chen, J., Deutschle, S., Fuerstenberg, K., 2007. Evaluation Methods and Results of the INTERSAFE Intersection Assistants, in: 2007 IEEE Intelligent Vehicles Symposium. pp. 142–147.

Chen, L., Li, Q., Li, M., Mao, Q., 2011. Traffic sign detection and recognition for intelligent vehicle, in: 2011 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2011 IEEE Intelligent Vehicles Symposium (IV), pp. 908–913.

Chen, S., Chen, F., 2010. Simulation-Based Assessment of Vehicle Safety Behavior under Hazardous Driving Conditions. Journal of Transportation Engineering 136, 304–315.

Chien, J.-C., Lee, J.-D., Chen, C.-M., Fan, M.-W., Chen, Y.-H., Liu, L.-C., 2013. An integrated driver warning system for driver and pedestrian safety. Applied Soft Computing 13, 4413–4427.

Chin, H.-C., Quek, S.-T., 1997. Measurement of traffic conflicts. Safety Science 26, 169–185.

Choi, E.-H., 2010. Crash Factors in Intersection-Related Crashes: An On-Scene Perspective (NHTSA Technical Report DOT HS 811 366). U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C., U.S.A.

Cleveland, W.S., 1981. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. The American Statistician 35, 54.

Cliff, W.E., Moser, A., 2001. Reconstruction of Twenty Staged Collisions with PC-Crash's Optimizer, in: SAE 2001 World Congress.

Coelingh, E., Eidehall, A., Bengtsson, M., 2010. Collision Warning with Full Auto Brake and Pedestrian Detection - a practical example of Automatic Emergency Braking. Presented at the 13th International IEEE Conference on Intelligent Transportation Systems, pp. 155–160.

Council, B.C., n.d. VENTURER consortium to test driverless cars in Bristol region [WWW Document]. URL http://www.bristol.gov.uk/press/venturer-consortium-test-driverless-cars-bristol-region (accessed 8.19.15).

Cuerden, R., Pittman, M., Dodson, E., Hill, J., 2008. The UK On the Spot accident data collection study: Phase II report (Road Safety Research Report No. 73). Department for Transport, London.

Cunningham, M., Regan, M.A., 2015. Autonomous Vehicles: Human Factors Issues and Future Research, in: Proceedings of the 2015 Australasian Road Safety Conference. Gold Coast, Australia.

Dabbour, E., Easa, S., 2014. Proposed collision warning system for right-turning vehicles at two-way stop-controlled rural intersections. Transportation Research Part C: Emerging Technologies 42, 121–131.

Dabbour, E., Easa, S., 2010. Technology-Independent Algorithm for Collision Warning System at Semi-Controlled Intersections. Canadian Journal of Transportation 3.

David, N.A., Norman, J.R., 1975. Motor vehicle accidents in relation to geometric and traffic features of highway intersections (FHWA-RD-76-128 Final Report).

Davies, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1, 224–227.

de Campos, G.R., Runarsson, A.H., Granum, F., Falcone, P., Alenljung, K., 2014. Collision avoidance at intersections: A probabilistic threat-assessment and decision-making system for safety interventions, in: Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 649–654.

De Ceunynck, T., 2017. Defining and applying surrogate safety measures and behavioural indicators through site-based observations (Dissertation). Lund University.

De Nicolao, G., Ferrara, A., Giacomini, L., 2007. Onboard Sensor-Based Collision Risk Assessment to Improve Pedestrians' Safety. IEEE Transactions on Vehicular Technology 56, 2405–2413. https://doi.org/10.1109/TVT.2007.899209

Department for Transport, G.M.H., 2006. Research database: Co-operative Crash Injury Study (CCIS) Phase 8 [WWW Document]. URL http://www.dft.gov.uk/rmd/project.asp?intProjectID=12088 (accessed 9.27.15).

Department for Transport, G.M.H., 2003. Research database: Heavy Vehicle Crash Injury Study [WWW Document]. URL http://www.dft.gov.uk/rmd/project.asp?intProjectID=11124 (accessed 9.27.15).

Digges, K., Gabler, H., Mohan, P., Alonso, B., 2005. Characteristics of the Injury Environment in Far-Side Crashes. Annual Proceedings / Association for the Advancement of Automotive Medicine 49, 185–197.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R., 2006. The 100-Car Naturalistic Driving Study, Phase II - Results of the 100-Car Field Experiment (Technical report No. DOT HS 810 593). Virginia Tech Transportation Institute, Blacksburg, Virginia.

DLR, n.d. Testfeld Niedersachsen für automatisiertes, vernetztes Fahren | DLR Verkehr [WWW Document]. URL http://verkehrsforschung.dlr.de/de/projekte/testfeld-niedersachsen-fuer-automatisiertes-vernetztes-fahren (accessed 1.21.18).

Dokic, J., Müller, B., Meyer, G., 2015. European Roadmap - Smart Systems for Automated Driving (VDI/VDE-IT Report). EPoSS.

Donath, M., Shankwitz, C., Ward, N., Creaser, J., 2007. Intersection Decision Support: An Overview (Report No. 6 - MN/RC 2007-33), Developing Intersection Decision Support Solutions. Minnesota Department of Transportation.

Eckardt, A., 2015. Germany Set to Open Up Autobahn to Self-Driving Vehicles. CNBC.

Eckert, A., Hohm, A., Lueke, S., 2013. An integrated ADAS solution for pedestrian collision avoidance, in: Proceedings of the 23rd International Conference on the Enhanced Safety of Vehicles, Seoul, Republic of Korea. pp. 13–0298.

Elavarasi, S.A., Akilandeswari, J., 2014. Survey on clustering algorithm and similarity measure for categorical data. ICTACT Journal on Soft Computing 4.

Elgharbawy, M., Schwarzhaupt, A., Scheike, G., Frey, M., Gauterin, F., 2016. A Generic Architecture of ADAS Sensor Fault Injection for Virtual Tests, in: 13th IEEE International Conference on Computer Systems and Applications.

Elvik, R., Vaa, T., 2004. The handbook of road safety measures. Elsevier, Oslo, Norway.

Emig, R., Schramm, H., 1989. Traction Control for Commercial Vehicles (SAE Technical Paper No. 890046). SAE International, Warrendale, PA.

Erbsmehl, C.T., Wagner, M., 2012. Nutzung des IPG CarMaker zur Simulation realer Verkehrsunfälle der GIDAS Datenbank. Presented at the IPG Technology Conference, Karlsruhe, Germany.

ERTRAC, 2017. Automated Driving Roadmap, Version 7.0. Brussels, Belgium.

Ester, M., Kriegel, H., S, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. AAAI Press, pp. 226–231.

Etemad, A., 2013. interactIVe - Accident avoidance by active intervention for Intelligent Vehicles. Presented at the 20th ITS World Congress, Tokyo.

European Commission, 2015. Junctions - European Commission [WWW Document]. URL http://ec.europa.eu/transport/road_safety/specialist/knowledge/road/getting_initial_safety_design_principles_right/junctions_en.htm (accessed 7.21.15).

European Commission, n.d. CARE database - Statistics – accidents data - Mobility and transport [WWW Document]. Mobility and transport. URL /transport/road_safety/specialist/statistics_en (accessed 4.10.18).

Evans, L., 1994. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. Accident Analysis & Prevention 26, 609–616.

Fancher, P., Bareket, Z., Ervin, R., 2004. Driving Assistance Provided by Adaptive Cruise Control (SAE Technical Paper No. 2004-21–0076). SAE International, Warrendale, PA.

Feng, F., Bao, S., Sayer, J.R., Flannagan, C., Manser, M., Wunderlich, R., 2017. Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. Accident Analysis & Prevention 104, 125–136.

Fernandez Llorca, D., Milanes, V., Parra Alonso, I., Gavilan, M., Garcia Daza, I., Perez, J., Sotelo, M.Á., 2011. Autonomous Pedestrian Collision Avoidance Using a Fuzzy Steering Controller. IEEE Transactions on Intelligent Transportation Systems 12, 390–401.

Fildes, B., Keall, M., Bos, N., Lie, A., Page, Y., Pastor, C., Pennisi, L., Rizzi, M., Thomas, P., Tingvall, C., 2015. Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes. Accident Analysis & Prevention 81, 24–29.

Fitch, G.M., Bowman, D.S., Llaneras, R.E., 2014. Distracted Driver Performance to Multiple Alerts in a Multiple-Conflict Scenario. Hum Factors 56, 1497–1505.

Fitzpatrick, K., Wooldridge, M.D., Blaschke, J.D., 2004. Urban Intersection Design Guide: Volume 1

(FHWA/TX-05/0-4365-P2 Vol. 1). Texas Department of Transportation.

Frye, C., 2001. International cooperation to prevent collisions at intersections. Public Roads 65.

Gabauer, D.J., Gabler, H.C., 2006. Comparison of Delta-V and Occupant Impact Velocity Crash Severity Metrics Using Event Data Recorders. Annual Proceedings / Association for the Advancement of Automotive Medicine 50, 57–71.

Gabler, H.C., Digges, K., Fildes, B.N., Sparke, L., 2005. Side impact injury risk for belted far side passenger vehicle occupants. SAE Technical Paper.

Ganti, V., Gehrke, J., Ramakrishnan, R., 1999. CACTUS—clustering categorical data using summaries, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 73–83.

Gasser, T.M., 2012. Ergebnisse der Projektgruppe Automatisierung: Rechtsfolgen zunehmender Fahrzeugautomatisierung. Berichte der Bundesanstalt für Straßenwesen, Unterreihe „Fahrzeugsicherheit“ Heft F83.

Gennarelli, T.A., Wodzin, E., 2008. Abbreviated injury scale 2005: Update 2008. Barrington, III, Association for the Advancement of Automative Medicine.

Gettman, D., Head, L., 2003. Surrogate Safety Measures From Traffic Simulation Models (Final report, FHWA-RD-03-050). Federal Highway Administration.

Gettman, D., Pu, L., Sayed, T., Shelby, S.G., 2008. Surrogate safety assessment model and validation: Final report (No. FHWA-HRT-08-051).

Gibson, D., Kleinberg, J.M., Prabhakar, R., 1998. Clustering Categorical Data: An Approach Based on Dynamics Systems, in: Proceedings of the 24th VLDB Conference. New York, NY, USA.

GIDAS - German In-Depth Accident Study [WWW Document], n.d. URL http://gidas.org (accessed 12.30.17).

Golembiewski, G.A., Chandler, B., 2011. Intersection safety - A manual for Local Rural Road Owners (FHWA-SA-11-08). Federal Highway Administration.

Goodman, L.A., 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 61, 215–231. https://doi.org/10.1093/biomet/61.2.215

Gornea, D., Popescu, D., Stamatescu, G., Fratila, R., 2014. Mono-camera robotic system for tracking moving objects. Presented at the 9th IEEE Conference on Industrial Electronics and Applications, pp. 1820–1825.

Graab, B., Donner, E., Chiellino, U., Hoppe, M., 2008. Analyse von Verkehrsunfällen hinsichtlich unterschiedlicher Fahrerpopulationen und daraus ableitbarer Ergebnisse für die Entwicklung adaptiver Fahrerassistenzsysteme, in: 3. Tagung Aktive Sicherheit Durch Fahrerassistenz.

Grapinet, M., De Souza, P., Smal, J.-C., Blosseville, J.-M., 2013. Characterization and simulation of optical sensors. Accident Analysis & Prevention 60, 344–352.

Gruber, J., Maurer, P., 2004. Zustandserfassung der Straßenoberfläche auf Bundesstraßen B. Straßenforschungsheft 542.

Guha, S., Rastogi, R., Shim, K., 1999. ROCK: A robust clustering algorithm for categorical attributes, in: Proceedings of the 15th International Conference on Data Engineering. IEEE, pp. 512–521.

Guillaume, S., Guillet, F., Philippé, J., 1998. Improving the discovery of association rules with intensity of implication, in: Żytkow, J.M., Quafafou, M. (Eds.), Principles of Data Mining and Knowledge Discovery, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 318–327.

Hahsler, M., Buchta, C., Gruen, B., Hornik, K., 2017. arules: Mining Association Rules and Frequent Itemsets (R package version 1.5-2).

Hahsler, M., Grün, B., Hornik, K., 2005. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. Journal of Statistical Software 14.

Haleem, K., Abdel-Aty, M., Mackie, K., 2010. Using a reliability process to reduce uncertainty in predicting crashes at non-signalized intersections. Accident Analysis & Prevention 42, 654–666.

Hallerbach, S., Eberle, U., Köster, F., 2017. Absicherungs- und Bewertungsmethoden für kooperative hochautomatisierte Fahrzeuge. Presented at the AAET 2017 - Automatisiertes und vernetztes Fahren, Braunschweig, Germany.

Hamming, R.W., 1950. Error detecting and error correcting codes. The Bell System Technical Journal 29, 147–160.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques, 3rd ed. Elsevier.

Hanna, J.T., Flynn, T.E., Tyler, W.L., 1976. Characteristics of intersection accidents in rural municipalities. Transportation Research Record 601.

Hansen, T., Schulz, M., Knoop, M., Konigorski, U., 2016. Trajectory Planning for Automated Lane Changes. ATZ worldwide 118, 60–65.

Happee, R., 2015. Dutch Automated Vehicle Initiative - The need for on-road assessment of the human interaction with automated driving. Presented at the Verkehr Aktuell, Munich, Germany.

Harrison, R.L., Granja, C., Leroy, C., 2010. Introduction to Monte Carlo Simulation, in: AIP Conference Proceedings. pp. 17–21.

Harwood, D.W., 1995. Median Intersection Design (NCHRP Report No. 375). Transportation Research Board.

Hassan, A.M., Mackay, M., Padmanaban, J., Davis, M.S., 1999. Comparison of Real World Side Impact Collisions which Occurred in the United Kingdom and United States. Annual Proceedings / Association for the Advancement of Automotive Medicine 43, 89–100.

Hauer, E., Garder, P., 1986. Research into the validity of the traffic conflicts technique. Accident Analysis & Prevention 18, 471–481.

Hayashi, H., Inomata, R., Fujishiro, R., Ouchi, Y., Suzuki, K., Nanami, T., 2013. Development of pre-crash safety system with pedestrian collision avoidance assist, in: Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV), 13.

Hayward, J.C., 1972. Near-miss determination through use of scale of danger. Highway Research Record 384, 24–34.

He, Z., Xu, X., Deng, S., 2002. Squeezer: An efficient algorithm for clustering categorical data. J. Comput. Sci. & Technol. 17, 611–624.

Heck, P., Bellin, J., Matoušek, M., Wonneberger, S., Sychrovskỳ, O., Šára, R., Maurer, M., 2013. Collision mitigation for crossing traffic in urban scenarios, in: Intelligent Vehicles Symposium (IV), 2013 IEEE. IEEE, pp. 559–566.

Helmer, T., 2014. Development of a Methodology for the Evaluation of Active Safety using the Example of Preventive Pedestrian Protection, Springer Theses. Springer.

Helmer, T., Wang, L., Kompass, K., Kates, R., 2015. Safety Performance Assessment of Assisted and Automated Driving by Virtual Experiments: Stochastic Microscopic Traffic Simulation as Knowledge Synthesis. Presented at the IEEE 18th International Conference on Intelligent Transportation Systems, pp. 2019–2023.

Hill, J., Thomas, P., Smith, M., Byard, N., Rillie, I., 2001. The methodology of on the spot accident investigations in the UK, in: Proceedings of 17th Conference on the Enhanced Safety of Vehicles (ESV). National Highway Traffic Safety Administration, U.S. Department of Transportation, Amsterdam.

Hillenbrand, J., Kroschel, K., 2006. A study on the performance of uncooperative collision mitigation systems at intersection-like traffic situations. Presented at the 2006 IEEE Conference on Cybernetics and Intelligent Systems, pp. 1–6.

Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemisch, F., Krueger, H.-P., Bartels, A., Beutner, A., Pagle, K., 2008. Highly automated vehicles for intelligent transport: HAVEit approach, in: Conference Proceedings of the 15th World Congress on ITS.

Holzmann, H., 2006. Einsatz der CarMaker-Produktfamilie bei der simulationsbasierten Entwicklung von Fahrdynamik-Regelsystemen. Presented at the IPG CarMaker User Conference, Karlsruhe, Germany.

Hsu, C.-C., 2006. Generalizing Self-Organizing Map for Categorical Data. IEEE Transactions on Neural Networks 17, 294–304.

Huang, Z., 1997. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining., in: SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery. pp. 1–8.

Huang, Z., Ng, M.K., 1999. A fuzzy k-modes algorithm for clustering categorical data. IEEE Transactions on Fuzzy Systems 7, 446–452.

Huh, K., Park, J., Hwang, J., Hong, D., 2008. A stereo vision-based obstacle detection system in vehicles. Optics and Lasers in Engineering 46, 168–178.

Hydèn, C., 1987. The development of a method for traffic safety evaluation: The Swedish traffic conflicts technique, Bulletin 70. Institut för Trafikteknik, LTH Lund.

Ibanez, A.G., Flores, C., Reyes, P.D., Barba, A., Reyes, A., 2011a. A Performance Study of the 802.11p Standard for Vehicular Applications, in: 2011 7th International Conference on Intelligent Environments (IE). Presented at the 2011 7th International Conference on Intelligent Environments (IE), pp. 165–170.

Ibanez, A.G., Flores, C., Reyes, P.D., Barba, A., Reyes, A., 2011b. A Performance Study of the 802.11p Standard for Vehicular Applications. Presented at the 7th International Conference on Intelligent Environments (IE), pp. 165–170.

IEEE, 2007. IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999) 1–1076.

iMobility Forum, 2013. Roadmap - Automation in Road Transport, Version 1.0.

INTERSAFE, 2005. Deliverable D40.4 - Requirements for intersection safety applications.

IPG Automotive, 2017. CarMaker Reference Manual Version 6.0.

IPG Automotive, 2016. CarMaker Reference Manual Version 5.1.

Ismail, K., Sayed, T., Saunier, N., Lim, C., 2009. Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data. Transportation Research Record: Journal of the Transportation Research Board 2140, 44–54.

ISO, 2011. ISO 26262 (all parts) - Road vehicles -- Functional safety.

Jääskeläinen, J., 2013. Automated Driving from the European Perspective. Presented at the Workshop on Automation in Road Transport, Brussels, Belgium.

Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles 37, 547–579.

Jang, J.A., Choi, K., Cho, H., 2012. A Fixed Sensor-Based Intersection Collision Warning System in Vulnerable Line-of-Sight and/or Traffic-Violation-Prone Environment. IEEE Transactions on Intelligent Transportation Systems 13, 1880–1890.

Junietz, P., Schneider, J., Winner, H., 2017. Metrik zur Bewertung der Kritikalität von Verkehrssituationen und -szenarien. Presented at the Workshop Fahrerassistenz und automatisiertes Fahren.

Jurewicz, C., Sobhani, A., Woolley, J., Dutschke, J., Corben, B., 2015a. Proposed vehicle impact speed - severe injury probability relationships for selected crash types, in: Proceedings of the 2015 Australasian Road Safety Conference. Gold Coast, Australia.

Jurewicz, C., Tofler, S., Makwasha, T., 2015b. Improving the Performance of Safe System Infrastructure (Research Report No. AP-R498-15). Austroads.

Kalra, N., Paddock, S.M., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? Transportation Research Part A: Policy and Practice 94, 182–193.

Karlsson, R., Jansson, J., Gustafsson, F., 2004. Model-based statistical tracking and decision making for collision avoidance application, in: Proceedings of the 2004 American Control Conference. pp. 3435–3440 vol.4.

Kaufman, L., Rousseeuw, P.J. (Eds.), 1990. Finding Groups in Data, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Keller, M., Hass, C., Seewald, A., Bertram, T., 2014. Driving simulator study on an emergency steering assist, in: Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference On. IEEE, pp. 3008–3013.

Kienle, M., Franz, B., Winner, H., Bengler, K., Baltzer, M., Flemisch, F., Kauer, M., Weißgerber, T., Geyer, S., Bruder, R., Hakuli, S., Meier, S., 2014. Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. IET Intelligent Transport Systems 8, 183–189.

Killi, D.V., Vedagiri, P., 2014. Proactive Evaluation of Traffic Safety at An Non-signalized Intersection Using Micro- Simulation. Journal of Traffic and Logistics Engineering 2, 140–145.

Kim, T., Jeong, H.Y., 2014. A Novel Algorithm for Crash Detection Under General Road Scenes Using Crash Probabilities and an Interactive Multiple Model Particle Filter. IEEE Transactions on Intelligent Transportation Systems 15, 2480–2490.

Kim, Y., Tak, S., Kim, J., Yeo, H., 2017. Identifying major accident scenarios in intersection and evaluation of collision warning system, in: 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, Yokohama, Japan, pp. 1–6.

Kircher, K., Larsson, A., Hultgren, J.A., 2014. Tactical Driving Behavior With Different Levels of Automation. IEEE Transactions on Intelligent Transportation Systems 15, 158–167.

Klappstein, J., Stein, F., Franke, U., 2007. Detectability of Moving Objects Using Correspondences over Two and Three Frames, in: Pattern Recognition, Lecture Notes in Computer Science. Presented at the Joint Pattern Recognition Symposium, Springer, Berlin, Heidelberg, pp. 112–121.

Köhler, S., Schreiner, B., Ronalter, S., Doll, K., Brunsmann, U., Zindler, K., 2013. Autonomous evasive maneuvers triggered by infrastructure-based detection of pedestrian intentions, in: Intelligent Vehicles Symposium (IV), 2013 IEEE. IEEE, pp. 519–526.

Körber, M., Weißgerber, T., Kalb, L., Blaschke, C., Farid, M., 2015. Prediction of take-over time in highly automated driving by two psychometric tests. DYNA 82, 195–201.

Kullgren, A., 2008. Dose-response models and EDR data for assessment of injury risk and effectiveness of safety systems. Presented at the IRCOBI conference, Bern, Switzerland.

Kumar, S., Toshniwal, D., 2015. A data mining framework to analyze road accident data. Journal of Big Data 2.

Lai, F., Carsten, O., Tate, F., 2012. How much benefit does Intelligent Speed Adaptation deliver: An analysis of its potential contribution to safety and environment. Accident Analysis & Prevention 48, 63–72.

Lamotte, O., Galland, S., Contet, J.-M., Gechter, F., 2010. Submicroscopic and Physics Simulation of Autonomous and Intelligent Vehicles in Virtual Reality. IEEE, pp. 28–33.

Lategahn, H., Geiger, A., Kitt, B., 2011. Visual SLAM for autonomous ground vehicles, in: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 1732–1737.

Lategahn, H., Stiller, C., 2013. Hochgenaue Eigenlokalisierung mit einer Monokamera - Kameralokalisierung. Presented at the 6. Tagung Fahrerassistenz, Munich, Germany.

Laureshyn, A., De Ceunynck, T., Karlsson, C., Svensson, Å., Daniels, S., 2017. In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. Accident Analysis & Prevention 98, 46–56.

Laureshyn, A., Svensson, Å., Hydén, C., 2010. Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. Accident Analysis & Prevention 42, 1637–1646.

Layfield, R.E., Summersgill, I., Hall, R.D., Chatterjee, K., 1996. Accidents at urban priority crossroads and staggered junctions (TRL Technical Report No. 185). Transport Research Laboratory (TRL), Crowthorne, UK.

Le, L., Festag, A., Baldessari, R., Zhang, W., 2009. Communication-Based Intersection Safety: Motivation, Challenges and State-of-the-Art. Presented at the 6th International Workshop on Intelligent Transportation (WIT), Hamburg, Germany.

Lee, J., Park, B., 2012. Development and Evaluation of a Cooperative Vehicle Intersection Control Algorithm Under the Connected Vehicles Environment. IEEE Transactions on Intelligent Transportation Systems 13, 81–90. https://doi.org/10.1109/TITS.2011.2178836

Lee, P., Hobbs, C., 2013. Understanding ISO 26262 ASILs. Electronic Design.

Lee, S.E., Knipling, R.R., DeHart, M.C., Perez, M.A., Holbrook, G.T., Brown, S.B., Stone, S.R., Olson, R.L., 2004. Vehicle-based countermeasures for signal and stop sign violations (Report DOT HS 809 716).

Lemmer, K., 2017. PEGASUS - Effectively ensuring automated driving. Presented at the 19. VDA Technischer Kongress 2017, Berlin, Germany.

Leonard, J., How, J., Teller, S., Berger, M., Campbell, S., Fiore, G., Fletcher, L., Frazzoli, E., Huang, A., Karaman, S., others, 2009. A perception-driven autonomous urban vehicle. The DARPA Urban Challenge 163–230.

Li, W., Tarko, A., 2011. Effect of Arterial Signal Coordination on Safety. Transportation Research

Record: Journal of the Transportation Research Board 2237, 51–59.

Li, X., Shaobin, W., Li, F., 2014. Fuzzy based collision avoidance control strategy considering crisis index in low speed urban area. Presented at the 2014 IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific), pp. 1–6.

Liebemann, E.K., Meder, K., Schuh, J., Nenninger, G., 2004. Safety and Performance Enhancement: The Bosch Electronic Stability Control (ESP) (SAE Technical Paper No. 2004-21–0060). SAE International, Warrendale, PA.

Lindner, F., Kressel, U., Kaelberer, S., 2004. Robust recognition of traffic signals. Presented at the IEEE Intelligent Vehicles Symposium, 2004, pp. 49–53.

Lloyd, S.P., 1982. Least Squares Quantization in PCM. IEEE Transactions on Information Theory 28, 129–137.

Lord, D., 1996. Analysis of Pedestrian Conflicts with Left-Turning Traffic. Transportation Research Record 1538, 61–67.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44, 291–305.

Lourenco, F., Lobo, V., Bacao, F., 2004. Binary-based similarity measures for categorical data and their application in Self-Organizing Maps. JOCLAD 88, 1–18.

Lutin, J.M., Kornhauser, A.L., Lerner-Lam, E., others, 2013. The revolutionary development of self-driving vehicles and implications for the transportation engineering profession. ITE Journal 83.

Lutz, A., Schick, B., Holzmann, H., Kochem, M., Meyer-Tuve, H., Lange, O., Mao, Y., Tosolin, G., 2017. Simulation methods supporting homologation of Electronic Stability Control in vehicle variants. Vehicle System Dynamics 55, 1432–1497.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. The Regents of the University of California.

Mages, M., Hopstock, M., Klanner, F., 2015. Kreuzungsassistenz, in: Winner, H., Hakuli, S., Wolf, G. (Eds.), Handbuch Fahrerassistenzsysteme. Vieweg+Teubner, pp. 975–993.

Mages, M.A., 2008. Top-Down-Funktionsentwicklung eines Einbiege- und Kreuzenassistenten (Dissertation). Technische Universität Darmstadt.

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

Maile, M.A., Neale, V.L., Ahmed-Zaid, F., Basnyake, C., Caminiti, L., Doerzaph, Z.R., Kass, S., Kiefer, R., Losh, M., Lundberg, J., Masselink, D., McGlohon, E., Mudalige, P., Pall, C., Peredo, M., Perez, M.A., Popovic, Z., Stinnett, J., Sudweeks, J., VanSickle, S., Kiger, S., 2008. Cooperative Intersection Collision Avoidance System Limited to Stop Sign and Traffic Signal Violations (CICAS-V), Phase I Final Report.

Maisch, W., Jonner, W.-D., Mergenthaler, R., Sigl, A., 1993. ABS5 and ASR5: The New ABS/ASR Family to Optimize Directional Stability and Traction (SAE Technical Paper No. 930505). SAE International, Warrendale, PA.

Mansfield, H., Bunting, A., Martens, M., Van der Horst, R., 2008. Analysis of the On the Spot (OTS) Road Accident Database (Road Safety Research Report No. 80). Department for Transport, London, UK.

Masuda, S., 2017. Software Testing Design Techniques Used in Automated Vehicle Simulations, in: 2017 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW). pp. 300–303.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21, 239–245.

Merat, N., Jamson, A.H., Lai, F.C.H., Carsten, O., 2012. Highly Automated Driving, Secondary Task Performance, and Driver State. Human Factors: The Journal of the Human Factors and Ergonomics Society 54, 762–771.

Metropolis, N., Ulam, S., 1949. The Monte Carlo Method. Journal of the American Statistical Association 44, 335–341.

Meyer, E., 2016. Understanding Injury Mechanisms from Real World Accidents (Master's thesis). Chalmers University of Technology, Gothenburg, Sweden.

Milanes, V., Villagra, J., Godoy, J., Simo, J., Perez, J., Onieva, E., 2012. An Intelligent V2I-Based Traffic Management System. IEEE Transactions on Intelligent Transportation Systems 13, 49–58.

Miller, R., Huang, Q., 2002. An adaptive peer-to-peer collision warning system, in: IEEE 55th Vehicular Technology Conference. pp. 317–321.

Minderhoud, M.M., Bovy, P.H., 2001. Extended time-to-collision measures for road traffic safety assessment. Accident Analysis & Prevention 33, 89–97.

Mir, Z.H., Filali, F., 2014. LTE and IEEE 802.11p for vehicular networking: a performance evaluation. EURASIP Journal on Wireless Communications and Networking 2014.

Mirabadi, A., Sharifian, S., 2010. Application of association rules in Iranian Railways (RAI) accident data analysis. Safety Science 48, 1427–1435.

Misener, J., Chan, C.-Y., Dickey, S., Kim, Z.W., Kuhn, T., Lian, T., Nelson, D., Sharafsaleh, A., Shladover, S., Spring, J., Vanderwerf, J., Zhang, W.-B., Zhang, L., Zhou, K., 2010. Cooperative Intersection Collision Avoidance System (CICAS): Signalized Left Turn Assist and Traffic Signal Adaptation (Final Report for Task Order 6608 No. UCB-ITS-PRR-2010-20), California PATH Research Report. University of California, Berkeley, Richmond, CA.

Molinero Martinez, A., Evdorides, H., Naing, C.L., Kirk, A., Tecl, J., Barrios, J.M., Simon, M.C., Phan, V., Hermitte, T., 2008. Accident causation and pre-accidental driving situations. Part 2. In-depth accident causation analysis (TRACE deliverable D2.2).

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

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accident Analysis & Prevention 49, 58–72.

Najm, W.G., Smith, J.D., Smith, D.L., 2001. Analysis of Crossing Path Crashes (DOT-VNTSC-NHTSA-01-03). U.S. Department of Transportation, National Highway Traffic Safety Administration.

National Instruments, n.d. What is the ISO 26262 Functional Safety Standard? [WWW Document]. URL http://www.ni.com/white-paper/13647/en/#toc2 (accessed 8.17.15).

Naujoks, F., Mai, C., Neukum, A., 2014. The effect of urgency of take-over requests during highly automated driving under distraction conditions. Advances in Human Aspects of Transportation 7, 431.

Neubauer, C., Matthews, G., Langheim, L., Saxby, D., 2012. Fatigue and Voluntary Utilization of Automation in Simulated Driving. Human Factors: The Journal of the Human Factors and Ergonomics Society 54, 734–746.

NHTSA, 2013. Preliminary Statement of Policy Concerning Automated Vehicles.

NHTSA, 2001. Tire Pressure Monitoring System (FMVSS No. 138).

Ni, C., Gordon, T.J., Zhang, D., Gao, Y., 2016. Test and validation of novel Lane-Departure Prevention System. Presented at the 2016 International Conference for Students on Applied Engineering (ICSAE), pp. 441–446.

Nienhuser, D., Gumpp, T., Zollner, J.M., Dillmann, R., 2008. Recognition and attribution of variable message signs and lanes. Presented at the 2008 IEEE Intelligent Vehicles Symposium, pp. 55–60.

Nitsche, P., Mocanu, I., Reinthaler, M., 2014. Requirements on Tomorrow's Road Infrastructure for Highly Automated Driving, in: The 3rd International Conference on Connected Vehicles & Expo (ICCVE 2014). Vienna, Austria.

Nitsche, P., Thomas, P., Stuetz, R., Welsh, R., 2017. Pre-crash scenarios at road junctions: A clustering method for car crash data. Accident Analysis & Prevention 107, 137–151.

Normalization - PHP-ML - Machine Learning library for PHP [WWW Document], n.d. URL https://php-ml.readthedocs.io/en/latest/machine-learning/preprocessing/normalization/ (accessed 2.8.18).

Obeng, K., 2007. Some determinants of possible injuries in crashes at signalized intersections. Journal of Safety Research 38, 103–112.

Okuda, R., Kajiwara, Y., Terashima, K., 2014. A survey of technical trend of ADAS and autonomous driving, in: Proceedings of Technical Program - 2014 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), pp. 1–4.

Olivares, S.P., Rebernik, N., Eichberger, A., Stadlober, E., 2016. Virtual Stochastic Testing of Advanced Driver Assistance Systems, in: Advanced Microsystems for Automotive Applications 2015, Lecture Notes in Mobility. Springer, Cham, pp. 25–35.

Olstam, J., 2009. Simulation of Surrounding Vehicles in Driving Simulators (Dissertation). Linköping University, Norrköping, Sweden.

On-road Automated Vehicle Standards Committee, 2014. SAE J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.

Oxford Dictionaries, n.d. autonomous - definition of autonomous in English | Oxford Dictionaries [WWW Document]. Oxford Dictionaries | English. URL https://en.oxforddictionaries.com/definition/autonomous (accessed 8.28.17).

Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Safety Science 47, 145–154.

Parker, M.R., Zegeer, C.V., 1989. Traffic Conflict Techniques for Safety and Operations: Observers manual (Report No. FHWA-IP-88-027). U.S. Department of Transportation, Federal Highway Administration.

Partrouche, D., 2006. Intelligent Speed Adaptation in curves for autonomous vehicles (Master's thesis). University of Joseph Fourier,Grenoble, Grenoble.

Pearson, K., 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5 50, 157–175.

Perkins, S.R., Harris, J.L., 1968. Traffic conflict characteristics-accident potential at intersections. Highway Research Record 225.

Perkins, S.R., Harris, J.L., 1967. Criteria for traffic conflicts characteristics (Report GMR 632). General Motors Corporation.

Persaud, B., Nguyen, T., 1998. Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. Transportation Research Record: Journal of the Transportation Research Board 1635, 113–120.

Pfeffer, R., Haselhoff, M., 2016. Video Injection Methods in a Real-world Vehicle for Increasing Test Efficiency. ATZ worldwide 118, 44–49.

Pickering, D., Hall, R.D., 1985. Accidents at rural T-junction, in: Planning & Transport Res & Comp, Sum Ann Mtg.

Pierowicz, J., Jacoy, E., Lloyd, M.L., Bittner, A., Pirson, B., 2000. Intersection Collision Avoidance Using ITS Countermeasures (DOT HS 809 171), Intersection Collision Avoidance System Performance Guidelines. U.S. Department of Transportation, National Highway Traffic Safety Administration.

Pintard, L., 2015. From safety analysis to experimental validation by fault injection-case of automotive embedded systems. INP Toulouse.

Pintard, L., Fabre, J.-C., Kanoun, K., Leeman, M., Roy, M., 2013. Fault Injection in the Automotive Standard ISO 26262: An Initial Approach, in: Vieira, M., Cunha, J.C. (Eds.), Dependable Computing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 126–133.

Pirdavani, A., Brijs, T., Bellemans, T., Wets, G., 2010. A Simulation-Based Traffic Safety Evaluation of Signalized and Un-Signalized Intersections, in: Proceedings of the 15th International Conference Road Safety on Four Continents. pp. 1229–1239.

Plavsic, M., 2010. Analysis and Modeling of Driver Behavior for Assistance Systems at Road Intersections. TU München, München.

Polders, E., Daniels, S., Hermans, E., Brijs, T., Wets, G., 2015. Crash Patterns at Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board 2514, 105–116.

Polychronopoulos, A., Tsogas, M., Amditis, A., Scheunert, U., Andreone, L., Tango, F., 2004. Dynamic situation and threat assessment for collision warning systems: the EUCLIDE approach. Presented at the IEEE Intelligent Vehicles Symposium, 2004, pp. 636–641.

Pütz, A., Zlocki, A., Bock, J., Eckstein, L., 2017. System validation of highly automated vehicles with a database of relevant traffic scenarios, in: Proceedings of the 12th ITS European Congress. Strasbourg, France.

Rezanková, H., 2009. Cluster analysis and categorical data. Statistika 46, 216–232.

Riegger, L., Carlander, M., Lidander, N., Murgovski, N., Sjöberg, J., 2016. Centralized MPC for autonomous intersection crossing. Presented at the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1372–1377.

Roberts, J., 2015. Driverless taxis to become a major form of transport "in 10 years." Horizon Magazine.

Robinson, T., Chan, E., Coelingh, E., 2010. Operating platoons on public motorways: An introduction to the SARTRE platooning programme, in: 17th World Congress on Intelligent Transport Systems. p. 12.

Rodarius, C., Duflis, J., Fahrenkrog, Felix, Roesener, C., Varhelyi, Andras, Fernandez, R., Wang, L., Seininger, P., Willemsen, Dehlia, van Rooij, L., 2015. AdaptIVe Deliverable D7.1 - Test and Evaluation Plan.

Roesener, C., Sauerbier, J., Zlocki, A., Fahrenkrog, F., Wang, L., Várhelyi, A., de Gelder, E., Dufils, J., Breunig, S., Mejuto, P., 2017. A Comprehensive Evaluation Approach for Highly Automated Driving, in: 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.

Roessler, B., 2010. Status of European project Intersafe-2 on cooperative intersection safety, in: Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference On. pp. 381–386.

Rolfsmeier, A., 2013. Virtual validation of advanced driver assistance functions. HANSER automotive 7–8.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Ruta, A., Porikli, F., Watanabe, S., Li, Y., 2011. In-vehicle camera traffic sign detection and recognition. Machine Vision and Applications 22, 359–375. https://doi.org/10.1007/s00138-009-0231-x

Sacchi, E., Sayed, T., 2016. Conflict-based Safety Performance Functions to Predict Traffic Collisions by Type. Transportation Research Record Journal of the Transportation Research Board 2583, 50–55.

Salamati, K., Schroeder, B., Rouphail, N.M., Cunningham, C., Long, R., Barlow, J., 2012. Development and Implementation of a Conflict-based Assessment of Pedestrian Safety (CAPS) to Evaluate Accessibility of Complex Intersections. Transp Res Rec 2011, 148–155.

Salim, F.D., Cai, L., Indrawan, M., Loke, S.W., 2008. Road Intersections as Pervasive Computing Environments: Towards a Multiagent Real-Time Collision Warning System, in: 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom). pp. 621–626.

Salim, F.D., Loke, S.W., Rakotonirainy, A., Srinivasan, B., Krishnaswamy, S., 2007. Collision Pattern Modeling and Real-Time Collision Detection at Road Intersections, in: Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference. Seattle, WA.

Sander, U., 2017. Opportunities and limitations for intersection collision intervention—A study of real world 'left turn across path' accidents. Accident Analysis & Prevention 99, Part A, 342–355.

Sander, U., Lubbe, N., 2018. Market penetration of intersection AEB: Characterizing avoided and residual straight crossing path accidents. Accident Analysis & Prevention 115, 178–188.

Sandin, J., 2009. An analysis of common patterns in aggregated causation charts from intersection crashes. Accident Analysis & Prevention 41, 624–632.

Sanz Villegas, M.T., 2011. CARE database.

Sato, I., Yamano, C., Yanagawa, H., 2011. Crossing obstacle detection with a vehicle-mounted camera, in: Intelligent Vehicles Symposium (IV), 2011 IEEE. IEEE, pp. 60–65.

Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M., Neubauer, C., 2013. Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. J Exp Psychol Appl 19, 287–300.

Sayed, T., Zein, S., 1999. Traffic conflict standards for intersections. Transportation Planning and Technology 22, 309–323.

Schikuta, E., 1993. Grid-clustering: A fast hierarchical clustering method for very large data sets (Technical Report). Center for Research on Parallel Computation, Rice University, Houston, Texas.

Schirokoff, A., Pilli-Sihvola, E., Sihvola, N., 2012. Assessing the Safety Impacts of Intersection Safety Systems. Procedia - Social and Behavioral Sciences, Transport Research Arena 2012 48, 1515–1524.

Schubert, R., Mattern, N., Bours, R., 2014. Simulation of Sensor Models for the Evaluation of Advanced Driver Assistance Systems. ATZelektronik worldwide 9, 26–29.

Schwarz, G., 1978. Estimating the Dimension of a Model. The Annals of Statistics 6, 461–464.

scikit, n.d. Selecting the number of clusters with silhouette analysis on KMeans clustering — scikit-learn 0.17.1 documentation [WWW Document]. URL http://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html (accessed 8.16.16).

Seidel, D., 2017. Simulationsbasierte Auslegungsmethodik für automatisierte intelligente Transportsysteme. Presented at the 8. Tagung Fahrerassistenz, Munich, Germany.

Seo, B., Park, Y., Han, S., Battaglia, S., 2014. Crashworthiness Improvement through Integrated Analysis of Active and Passive Safety Systems, in: Proceedings of FISITA World Congress 2014. Maastricht, The Netherlands.

Sharma, D., Stern, S., Brophy, J., Choi, E., 2007. An overview of NHTSA's crash reconstruction software WinSMASH, in: Proceedings of the 20th International Technical Conference on Enhanced Safety of Vehicles.

Shbeeb, L., 2000. Development of traffic conflicts technique for different environments: A comparative study of pedestrian conflicts in Sweden and Jordan. Bulletin / Lunds tekniska högskola, Institutionen för teknik och samhälle, Lunds universitet 188.

Shirazi, M.S., Morris, B.T., 2017. Looking at Intersections: A Survey of Intersection Monitoring, Behavior and Safety Analysis of Recent Studies. IEEE Transactions on Intelligent Transportation Systems 18, 4–24.

Silverstein, C., Brin, S., Motwani, R., 1998. Beyond Market Baskets: Generalizing Association Rules to Dependence Rules. Data Mining and Knowledge Discovery 2, 39–68.

Sobhani, A., Young, W., Bahrololoom, S., Sarvi, M., others, 2013. Calculating time-to-collision for analysing right turning behaviour at signalised intersections. Road & Transport Research: A Journal of Australian and New Zealand Research and Practice 22, 49.

Songchitruksa, P., Tarko, A.P., 2006. The extreme value theory approach to safety estimation. Accident Analysis & Prevention 38, 811–822.

Stein, G.P., Gdalyahu, Y., Shashua, A., 2010. Stereo-Assist: Top-down stereo for driver assistance systems, in: 2010 IEEE Intelligent Vehicles Symposium. pp. 723–730.

Steinhaus, H., 1956. Sur la division des corp materiels en parties. Bull. Acad. Polon. Sci 1, 801–804.

Stigson, H., Kullgren, A., Rosén, E., 2012. Injury Risk Functions in Frontal Impacts Using Data from Crash Pulse Recorders. Annals of Advances in Automotive Medicine / Annual Scientific Conference 56, 267–276.

Stoff, A., Liers, H., 2013. Ausweichfunktionalität für Kreuzungsszenarien zur Unfallfolgelinderung durch Optimierung der Crash-Kompatibilität. Presented at the VDI Wissensforum 2013.

Svensson, Å., 1998. A method for analysing the traffic process in a safety perspective (Dissertation). Lund University, Lund, Sweden.

Tarko, A., Davis, G., Saunier, N., Sayed, T., Washington, S., 2009. Surrogate measures of safety. White paper, ANB20 (3) Subcommittee on Surrogate Measures of Safety.

Tatar, M., 2016. Test and Validation of Advanced Driver Assistance Systems Automated Search for Critical Scenarios. ATZ elektronik worldwide Magazine 11, 54–57.

TD 42/95 - Geometric Design of Major/Minor Priority Junctions, 1995. , in: Design Manual for Roads and Bridges, Volume 6 - Road Geometry. The Highways Agency, Scottish Executive, Welsh Assembly Government, The Department for Regional Development Northern Ireland.

Templeton, B., n.d. Affordable robocars -- will it be cameras or LIDAR? | Brad Ideas [WWW Document]. URL http://ideas.4brad.com/affordable-robocars-will-it-be-cameras-or-lidar (accessed 3.20.15).

Thomas, P., Bradford, M., 1988. Side impact severity - The use of discriminant analysis to classify injury, IRCOBI (International Research Council on the Biokinetics of Impacts). Accident Research Unit, Institute for Consumer Ergonomics, Loughborough, UK.

Thomas, P., Frampton, R., 1999. Injury patterns in side collisions-a new look with reference to current test methods and injury criteria, in: 43rd Stapp Car Crash Conference Proceedings.

Thomas, P., Yannis, G., Duchamp, G., Vis, M., Jahi, H., Vallet, G., Dupont, E., Martensen, H., 2009. SafetyNet Final Activity Report. VSRC – Vehicle Safety Research Centre.

Thrun, S., Burgard, W., Fox, D., 2006. Probabilistic robotics, Intelligent robotics and autonomous agents series. MIT Press, Cambridge, Mass.

Tideman, M., van Der Voort, M.C., van Arem, B., Tillema, F., 2007. A review of lateral driver support systems, in: 2007 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 992–999.

Tolouei, R., Maher, M., Titheridge, H., 2013. Vehicle mass and injury risk in two-car crashes: A novel methodology. Accident Analysis & Prevention 50, 155–166. https://doi.org/10.1016/j.aap.2012.04.005

TRL, 2014. Greenwichs digital credentials driven home after TRL led consortium wins £8m trial to pilot futuristic automated vehicles [WWW Document]. URL http://www.trl.co.uk/news-hub/trl-press-releases/2014/december/greenwichs-digital-credentials-driven-home-after-trl-led-consortium-wins-%C2%A38m-trial-to-pilot-futuristic-automated-vehicles/ (accessed 7.19.15).

Tsugawa, S., 2005. Issues and recent trends in vehicle safety communication systems. IATSS Research 29, 7–15.

Überbacher, M., Wolze, P., Burtsche, T., 2017. Experiencing Safety Function Testing. ATZ worldwide 119, 54–57.

UK Department for Transport, 2015. The pathway to driverless cars - Summary report and action plan. London, UK.

Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., Maurer, M., 2015. Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving. Presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 982–988.

UN, 1968. Chapter XI. Transport and Communications, B. Road Traffic - 19. Convention on Road Traffic. Vienna, 8 November 1968.

Uno, H., Kageyama, Y., Yamaguchi, A., Okabe, T., 2013. Effectiveness study of crash avoidance technologies by using clustering and self organizing map, in: Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV). Seoul, Korea.

Valero-Mora, P.M., Tontsch, A., Welsh, R., Morris, A., Reed, S., Touliou, K., Margaritis, D., 2013. Is naturalistic driving research possible with highly instrumented cars? Lessons learnt in three research centres. Accident Analysis & Prevention 58, 187–194.

Van der Horst, A.R.A., 1990. A time-based analysis of road user behaviour in normal and critical encounters (Dissertation). TU Delft, Delft, the Netherlands.

van Dijke, J.P., 2011. CITYMOBIL: Advanced road transport for the urban environment. Final results, in: 18th ITS World Congress.

Van Maren, P., 1980. Correlation of Design and Control Characteristics with Accidents at Rural Multi-Lane Highway Intersections in Indiana: Interim Report (Interim report No. FHWA/IN/JHRP-77/22). Purdue University, Indiana State Highway Commission.

Vanholme, B., 2012. Highly Automated Driving on Highways based on Legal Safety (Dissertation). University of Evry-Val-d'Essonne, Frankreich.

Vanholme, B., Gruyer, D., Lusetti, B., Glaser, S., Mammar, S., 2013. Highly Automated Driving on Highways Based on Legal Safety. IEEE Transactions on Intelligent Transportation Systems 14, 333–347. https://doi.org/10.1109/TITS.2012.2225104

Velske, S., Mentlein, H., Eymann, P., 2009. Strassenbau, Strassenbautechnik, 6. neu bearb. Auflage. ed. Werner, Neuwied, Germany.

Visvikis, C., Smith, T.L., Pitcher, M., Smith, R., 2013. Study on lane departure warning and lane change assistant systems (Project Report PPR 374). Transport Research Laboratory (TRL).

Vogt, A., Bared, J., 1998. Accident Models for Two-Lane Rural Segments and Intersections. Transportation Research Record: Journal of the Transportation Research Board 1635, 18–29.

Volvo Car Group, 2013. Volvo Car Group initiates world unique Swedish pilot project with self-driving cars on public roads [WWW Document]. URL https://www.media.volvocars.com/global/en-gb/media/pressreleases/136182/volvo-car-group-initiates-world-unique-swedish-pilot-project-with-self-driving-cars-on-public-roads (accessed 7.19.15).

von Neumann-Cosel, K., Dupuis, M., Weiss, C., 2009. Virtual Test Drive - Provision of a consistent tool-set for [D, H, S, V]-in-the-loop, in: DSC 2009 Europe. Monaco.

Wachenfeld, W., Winner, H., 2015. Die Freigabe des autonomen Fahrens, in: Autonomes Fahren.

Wåhlberg, A.., 2000. The relation of acceleration force to traffic accident frequency: a pilot study. Transportation Research Part F: Traffic Psychology and Behaviour 3, 29–38.

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis & Prevention 38, 1137–1150.

Wegner, P., 1960. A technique for counting ones in a binary computer. Communications of the ACM 3, 322.

Weigl, T., 2014. Development Process For Autonomous Vehicles (Master's thesis). Technische Universität München.

Weng, J., Zhu, J.-Z., Yan, X., Liu, Z., 2016. Investigation of work zone crash casualty patterns using association rules. Accident Analysis & Prevention 92, 43–52.

Williams, M.J., 1981. Validity of the traffic conflicts technique. Accident Analysis & Prevention 13, 133–145.

Wiltschko, T., 2004. Sichere Information durch infrastrukturgestützte Fahrerassistenzsysteme zur Steigerung der Verkehrssicherheit an Straßenknotenpunkten (Dissertation). Universität Stuttgart.

Winner, H., 2015. Grundlagen von Frontkollisionsschutzsystemen, in: Handbuch Fahrerassistenzsysteme. Springer Vieweg.

Wu, K.-F., Jovanis, P.P., 2012. Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. Accident Analysis & Prevention 45, 507–516.

Yamamura, Y., Tabe, M., Kanehira, M., Murakami, T., 2001. Development of an Adaptive Cruise Control System with Stop-and-Go Capability (SAE Technical Paper No. 2001- 01–0798). SAE International, Warrendale, PA.

Zaki, M., Sayed, T., Shaaban, K., 2014. Use of Drivers' Jerk Profiles in Computer Vision–Based Traffic Safety Evaluations. Transportation Research Record Journal of the Transportation Research Board 2434, 103.

Zegeer, C.V., Deen, R.C., 1978. Traffic conflicts as a diagnostic tool in highway safety. Transportation Research Record.

Zengyou, H., Xiaofei, X., Shengchun, D., 2004. A Link Clustering Based Approach for Clustering Categorical Data, in: Proceedings of the WAIM Conference. China.

Zengyou, H., Xiaofei, X., Shengchun, D., Bin, D., 2003. K-histograms: An efficient clustering algorithm for categorical dataset (Technical Report No. Tr-2003-08). Department for Computer Science and Engineering, Harbin Institute of Technology, China.

Zheng, L., Ismail, K., Meng, X., 2014. Traffic conflict techniques for road safety analysis: open questions and some insights. Canadian Journal of Civil Engineering 41, 633–641.

Ziade, H., Ayoubi, R.A., Velazco, R., 2004. A survey on fault injection techniques. The International Arab Journal of Information Technology 1, 171–186.

Zlocki, A., Bock, J., Eckstein, L., 2017. Database of relevant traffic scenarios for highly automated vehicles. Presented at the Automated Vehicle Test & Development Symposium 2017, Stuttgart, Germany.

9 Abbreviations

Abbreviation	Description
AADT	Annual Average Daily Traffic
ABS	Anti-lock Braking System
ACC	Adaptive Cruise Control
ADAS	Advanced Driving Assistance Systems
ADS	Automated Driving Systems
AEB	Automated Emergency Braking
CDF	Cumulative Probability Density Function
CICAS	Cooperative Intersection Collision Avoidance Systems
DSRC	Dedicated Short Range Communication
DST	Deceleration to Safety Time
DTI	Distance-to-Intersection
EDR	Event Data Recorder
EES	Energy equivalent speed
ESC	Electronic Stability Control
FCA	Forward Collision Avoidance
FCW	Forward Collision Warning
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GUI	Graphical User Interface
HIL	Hardware-in-the-Loop
HMI	Human Machine Interface
I2V	Infrastructure-to-Vehicle
ICAMS	Intersection Collision Avoidance and Mitigation Systems
ICT	Information and Communication Technologies
ISA	Intelligent Speed Adaptation
ITS	Intelligent Transport Systems
LHS	Latin Hybercube Sampling
LIDAR	Light Detection and Ranging
LKA	Lane Keeping Assist
MAIS	Maximum Abbreviated Injury Scale
MIL	Model-in-the-loop
NHTSA	National Highway Traffic Safety Administration
OEM	Original Equipment Manufacturer
OpenCRG	Open Curved Regular Grid
OSM	Open Street Map
OTS	On the Spot
PDF	Probability Density Function
PET	Post Encroachment Time
RAIDS	Road Accident In-Depth Studies
RMSE	Root-Mean-Squared Error
SAE	Society of Automotive Engineers
SCA	Side Collision Avoidance
SIL	Software-in-the-loop
SLAM	Simultaneous Location and Mapping
SSAM	Surrogate Safety Assessment Methodology
TCS	Traction Control System
TLC	Time to Line Crossing
TTA	Time-to-Accident
TTC	Time-to-Collision
TTD	Time-to-Disappear
TTI	Time-to-Intersection
TTR	Time-to-React
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
VRU	Vulnerable Road User
WiFi	Wireless Fidelity

10 List of figures

11 List of tables

Appendix A: Web survey questions

Screenshots of the web survey “Tomorrow's Road Infrastructure for Automated Driving”

Tomorrow's Road Infrastructure for Automated Driving

* Erforderlich

Technologies for Highly Automated Driving - Where are we now?

Automated driving systems and self-driving cars consist of various subsystems to tackle certain situations in traffic.

Regarding technical readiness, how close to market introduction do you see the following systems? *

	Ready for market	Prototyping	Still under research	Not sure
Automated lane keeping	○	○	○	○
Automated lane changing	○	○	○	○
Automated overtaking	○	○	○	○
Automated lane merging	○	○	○	○
Fully automated steering	○	○	○	○
Fully automated braking	○	○	○	○
Dynamic speed control and adaptation	○	○	○	○
Automated safety pull-over in emergency cases	○	○	○	○
Fully automated parking	○	○	○	○
Obstacle collision avoidance (including lateral and longitudinal control)	○	○	○	○
Collision avoidance with other road users (including lateral and longitudinal control)	○	○	○	○

If not mentioned above, please state other systems, which you consider relevant for automated driving.

« Zurück | Weiter »

33 % abgeschlossen

Tomorrow's Road Infrastructure for Automated Driving

* Erforderlich

Automated lane assistance

This group of systems assists drivers by keeping or changing the lane automatically. It is also capable of performing overtaking or lane merging maneuvers through automatic steering and/or braking. The systems are primarily based on vision sensors, but also employ radar technology when it comes to autonomous lane change maneuvers.

To which extent do the following factors influence the performance of automated lane assistance systems? *

1 = highest influence, 4 = lowest influence

	1	2	3	4
Cross-border differences of road infrastructure	○	○	○	○
Complex urban road environments (pedestrian crossings, bicycle lanes, parked cars)	○	○	○	○
Temporary road work zones	○	○	○	○
Quality of lane markings (e.g. visibility, discontinuities, remnants of old markings)	○	○	○	○
Visibility of traffic signs (e.g. dirt, damage)	○	○	○	○
Discontinuous or damaged road edges or kerbs	○	○	○	○
Poor visibility due to bad weather (e.g. fog, snow)	○	○	○	○
Poor visibility due to darkness	○	○	○	○
Quality of street lights (e.g. brightness, discontinuities)	○	○	○	○
Glare, e.g. due to sunshine or other cars	○	○	○	○
Road alignment (e.g. multiple curves, low radii)	○	○	○	○
Ruts and uneven road surface	○	○	○	○
Slippery road surface	○	○	○	○
Cracks on the road surface	○	○	○	○

If not mentioned above, please state other influencing factors.

« Zurück | Weiter »

50 % abgeschlossen

Tomorrow's Road Infrastructure for Automated Driving

* Erforderlich

Collision avoidance systems

This bundle of systems helps avoid collisions with other traffic participants (e.g. cars, pedestrians, bicyclists) and obstacles on the road by means of automated steering and/or braking. Radar, Lidar and/or vision sensors are used for these systems.

To which extent do the following factors influence the performance of collision avoidance systems? (same factors as in previous question) *

1 = highest influence, 4 = lowest influence

	1	2	3	4
Cross-border differences of road infrastructure	○	○	○	○
Complex urban road environments (pedestrian crossings, bicycle lanes, parked cars)	○	○	○	○
Temporary road work zones	○	○	○	○
Quality of lane markings (e.g. visibility, discontinuities, remnants of old markings)	○	○	○	○
Visibility of traffic signs (e.g. dirt, damage)	○	○	○	○
Discontinuous or damaged road edges or kerbs	○	○	○	○
Poor visibility due to bad weather (e.g. fog, snow)	○	○	○	○
Poor visibility due to darkness	○	○	○	○
Quality of street lights (e.g. brightness, discontinuities)	○	○	○	○
Glare, e.g. due to sunshine or other cars	○	○	○	○
Road alignment (e.g. multiple curves, low radii)	○	○	○	○
Ruts and uneven road surface	○	○	○	○
Slippery road surface	○	○	○	○
Cracks on the road surface	○	○	○	○

If not mentioned above, please state other influencing factors.

« Zurück | Weiter »

66 % abgeschlossen

Tomorrow's Road Infrastructure for Automated Driving

* Erforderlich

Speed control

This group of systems adapts the vehicle's speed dynamically based on the legal speed limit or external factors such as road alignment, weather conditions, congestion, road works, etc. Cooperative vehicle to infrastructure (V2I) technology is employed, as well as vision sensors capable of identifying and processing traffic signs.

To which extent do the following factors influence the performance of speed control systems? (same factors as in previous question) *

1 = highest influence, 4 = lowest influence

	1	2	3	4
Cross-border differences of road infrastructure	○	○	○	○
Complex urban road environments (pedestrian crossings, bicycle lanes, parked cars)	○	○	○	○
Temporary road work zones	○	○	○	○
Quality of lane markings (e.g. visibility, discontinuities, remnants of old markings)	○	○	○	○
Visibility of traffic signs (e.g. dirt, damage)	○	○	○	○
Discontinuous or damaged road edges or kerbs	○	○	○	○
Poor visibility due to bad weather (e.g. fog, snow)	○	○	○	○
Poor visibility due to darkness	○	○	○	○
Quality of street lights (e.g. brightness, discontinuities)	○	○	○	○
Glare, e.g. due to sunshine or other cars	○	○	○	○
Road alignment (e.g. multiple curves, low radii)	○	○	○	○
Ruts and uneven road surface	○	○	○	○
Slippery road surface	○	○	○	○
Cracks on the road surface	○	○	○	○

If not mentioned above, please state other influencing factors.

« Zurück | Weiter »

83 % abgeschlossen

Tomorrow's Road Infrastructure for Automated Driving

* Erforderlich

General information about you

You have accomplished the tricky part!
To complete the survey, please give some information about your person and provide comments if you wish.

What is your domain of work? *

- [ ] Research and Development
- [ ] Academia
- [ ] Industry
- [ ] Service provider
- [ ] Economics
- [ ] Marketing
- [ ] Public authority
- [ ] Sonstiges:

In which company or institute do you work? (optional)
Please give the name of your company/university/institute.

If you have any other comments, please enter them in the space provided below.

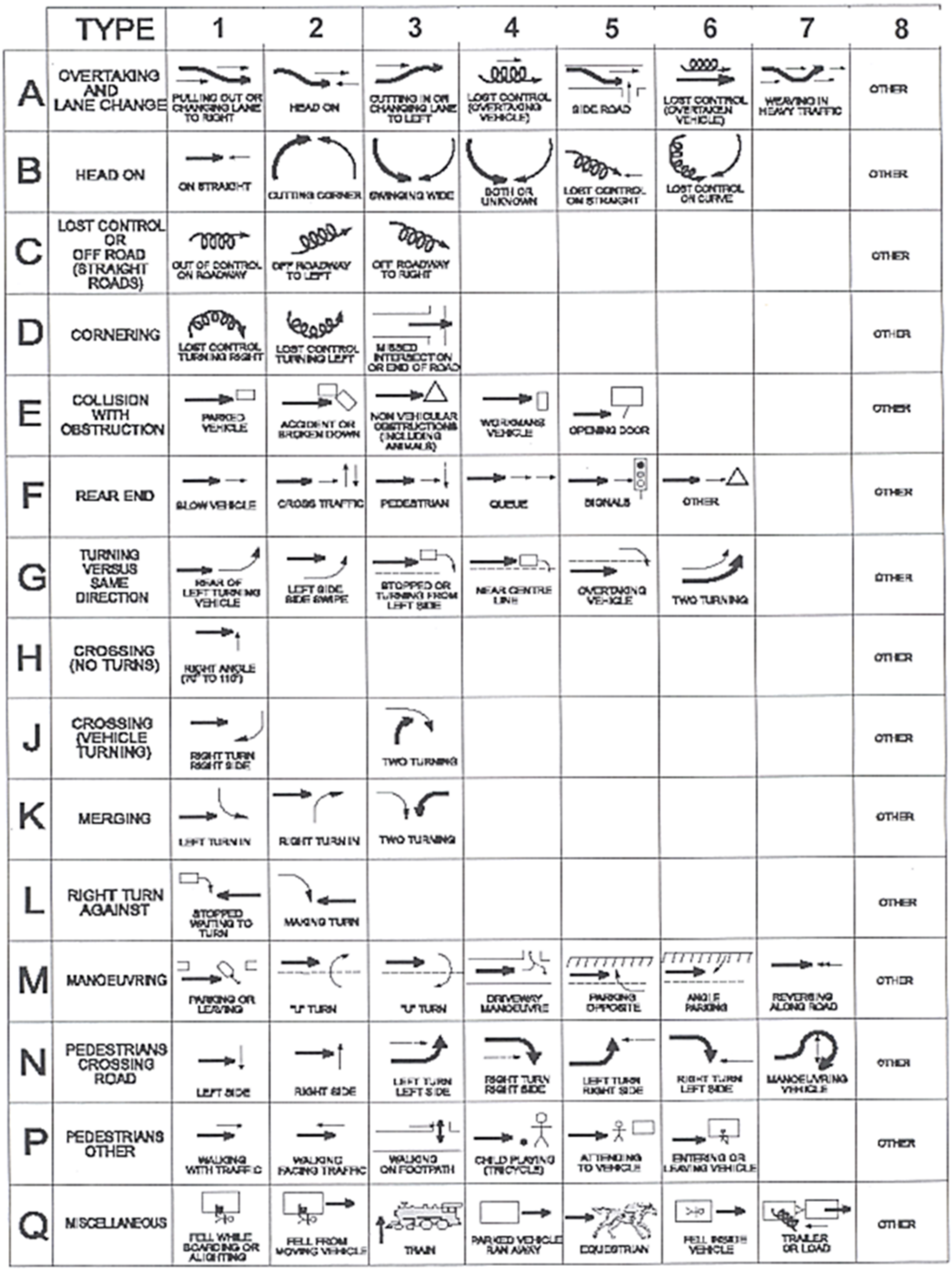
If you want to receive the survey results, please give your e-mail address.

« Zurück Senden

Geben Sie niemals Passwörter über Google Formulare weiter.

100 %: Sie haben es geschafft.

Appendix B: Collision code sheet



Collision Code Sheet

	TYPE	1	2	3	4	5	6	7	8
A	OVERTAKING AND LANE CHANGE	PULLING OUT OR CHANGING LANE TO RIGHT	HEAD ON	CUTTING IN OR CHANGING LANE TO LEFT	LOST CONTROL (OVERTAKING VEHICLE)	SIDE ROAD	LOST CONTROL (OVERTAKEN VEHICLE)	WEAVING IN HEAVY TRAFFIC	OTHER
B	HEAD ON	ON STRAIGHT	CUTTING CORNER	SWINGING WIDE	BOTH OR UNKNOWN	LOST CONTROL ON STRAIGHT	LOST CONTROL ON CURVE		OTHER
C	LOST CONTROL OR OFF ROAD (STRAIGHT ROADS)	OUT OF CONTROL ON ROADWAY	OFF ROADWAY TO LEFT	OFF ROADWAY TO RIGHT					OTHER
D	CORNERING	LOST CONTROL TURNING RIGHT	LOST CONTROL TURNING LEFT	MISSED INTERSECTION OR END OF ROAD					OTHER
E	COLLISION WITH OBSTRUCTION	PARKED VEHICLE	ACCIDENT OR BROKEN DOWN	NON VEHICULAR OBSTRUCTIONS (INCLUDING ANIMALS)	WORKMANS VEHICLE	OPENING DOOR			OTHER
F	REAR END	SLOW VEHICLE	CROSS TRAFFIC	PEDESTRIAN	QUEUE	SIGNALS	OTHER		OTHER
G	TURNING VERSUS SAME DIRECTION	REAR OF LEFT TURNING VEHICLE	LEFT SIDE SIDE SWIPE	STOPPED OR TURNING FROM LEFT SIDE	NEAR CENTRE LINE	OVERTAKING VEHICLE	TWO TURNING		OTHER
H	CROSSING (NO TURNS)	RIGHT ANGLE (70° TO 110°)							OTHER
J	CROSSING (VEHICLE TURNING)	RIGHT TURN RIGHT SIDE		TWO TURNING					OTHER
K	MERGING	LEFT TURN IN	RIGHT TURN IN	TWO TURNING					OTHER
L	RIGHT TURN AGAINST	STOPPED WAITING TO TURN	MAKING TURN						OTHER
M	MANOEUVRING	PARKING OR LEAVING	"U" TURN	"U" TURN	DRIVEWAY MANOEUVRE	PARKING OPPOSITE	ANGLE PARKING	REVERSING ALONG ROAD	OTHER
N	PEDESTRIANS CROSSING ROAD	LEFT SIDE	RIGHT SIDE	LEFT TURN LEFT SIDE	RIGHT TURN RIGHT SIDE	LEFT TURN RIGHT SIDE	RIGHT TURN LEFT SIDE	MANOEUVRING VEHICLE	OTHER
P	PEDESTRIANS OTHER	WALKING WITH TRAFFIC	WALKING FACING TRAFFIC	WALKING ON FOOTPATH	CHILD PLAYING (TRICYCLE)	ATTENDING TO VEHICLE	ENTERING OR LEAVING VEHICLE		OTHER
Q	MISCELLANEOUS	FELL WHILE BOARDING OR ALIGHTING	FELL FROM MOVING VEHICLE	TRAIN	PARKED VEHICLE RAN AWAY	EQUESTRIAN	FELL INSIDE VEHICLE	TRAILER OR LOAD	OTHER

OTS 2 : Collision Type Coding Form v1.0

Appendix C: OTS database tables

Figure 126 give an overview on the table relationships in the MS Access database, which was created for the OTS accident data analysis.

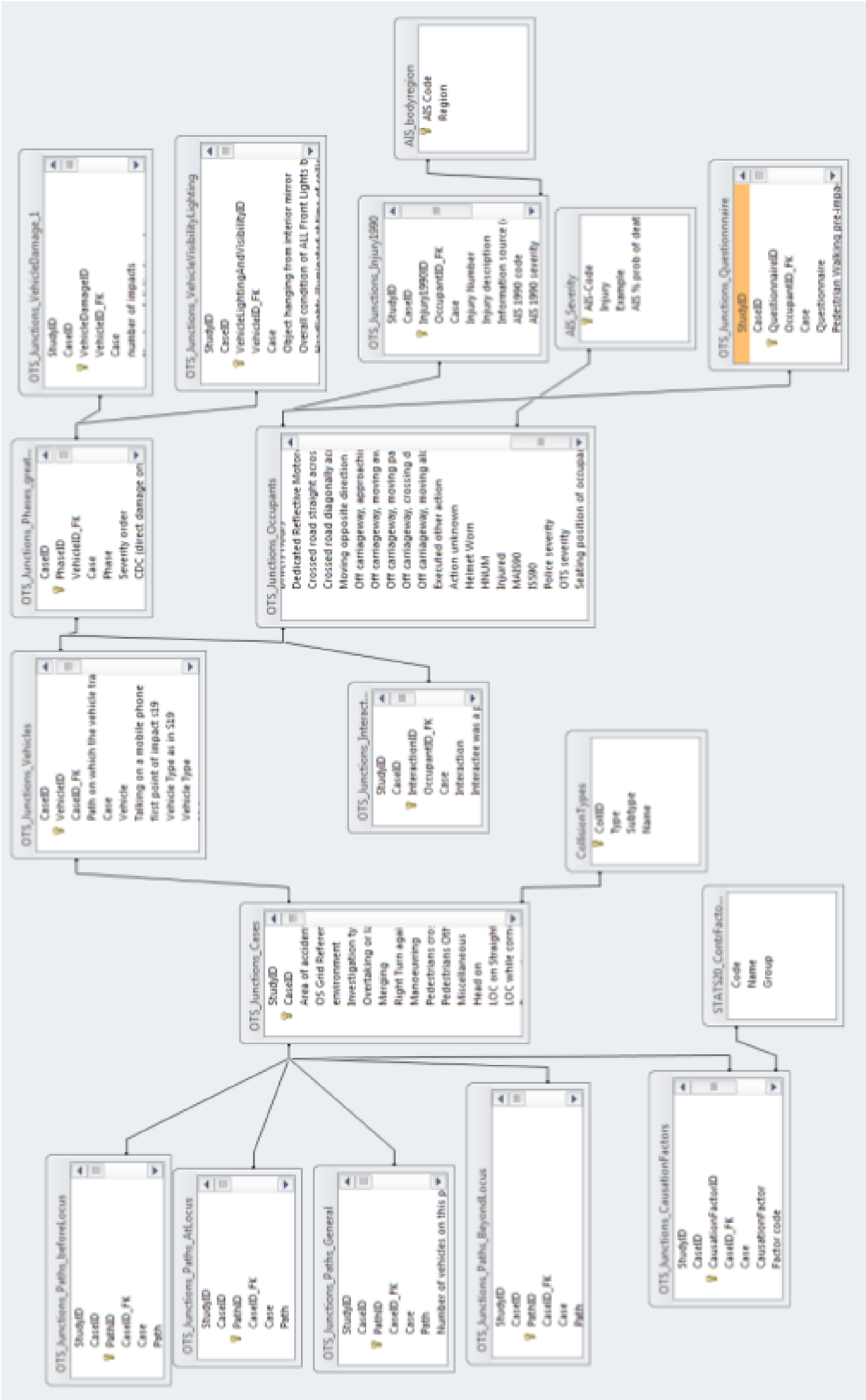
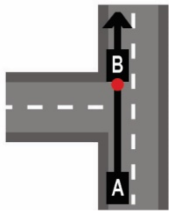


Figure 126: Relationships of all OTS tables in the database

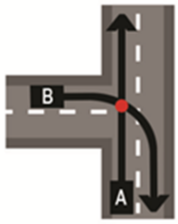
Appendix D: Overview of all scenarios at T-junctions

The association rules resulted in 24 scenarios for three-legged junctions, which are described on the following pages. The figures illustrate the scenarios in a simplified manner to better understand the descriptions in the text. The red dots in the figures are the points of impact (i.e. front, offside or nearside). Surface conditions, area (rural/urban), speed limits, vehicle types and injury levels are not shown. Note that all crashes in the data occurred on UK roads with left-hand traffic and that the road parameters are given for the path of A. Some scenarios were combined due to similar characteristics.

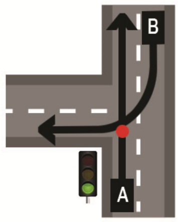
Scenario T-1.1 (F: Rear End): Car A is going straight on a major road and hits another car at the rear end. A is travelling on a rural dual carriageway with 70 mph speed limit without active or passive yield instruction, and it is caused by A failing to stop or failing to avoid or by other precipitating factors from B. The surface is dry and A remains uninjured or suffers slight injury.



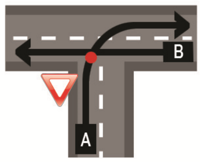
Scenario T-1.2 / 7.4 / 9.1 (J: Crossing/Vehicle Turning): Car A is going straight on a major road and hits another car B with its front, which turns right from a minor road joining from the left. A is travelling on an urban single carriageway with 30 to 50 mph speed limit without active or passive yield instruction, and it is caused by B failing to give way. The surface is dry and A suffers slight injuries.



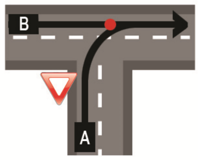
Scenario T-1.3 (L: Right Turn Against): Car A is going straight on a major road and hits another car B with its front, which is turning right into a minor road. A is travelling in dark light conditions on an urban single carriageway with 40 to 50 mph speed limit controlled by traffic lights, and it is caused by B failing to give way or manoeuvring inappropriately. A suffers slight injuries.



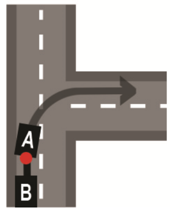
Scenario T-2.1/8.1 (J: Crossing/Vehicle Turning): Car A is turning right into a major road and hits another car with its front, which is coming from the right. A is travelling on an urban single carriageway with 30 to 50 mph speed limit controlled by a static give-way sign and it is caused by A failing to give way, leading to no or slight injury.



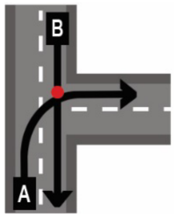
Scenario T-2.2 (K: Merging): Car A is pulling out from a minor road to turn right into a major road and is hit by another car B on its nearside, which is going straight. A is travelling on an urban or rural road controlled by a static give-way sign, and it is caused by A failing to give way.



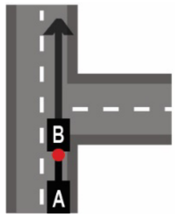
Scenario T-3.1 (F: Rear End): Car A is hit by another car or HGV B on its rear end, while waiting to turn right. A is travelling on a single carriageway with 30 to 60 mph speed limit without active or static yield instruction. It is caused by B failing to stop, leading to no or slight injury.



Scenario T-4.1 (L: Right Turn Against): Car A is turning into a minor road and is hit by a PTW B on its nearside, which is going straight in the opposing direction. A is travelling on a single carriageway with 40-50 mph speed limit without active or static yield instruction. It is caused by A failing to give way or manoeuvring inappropriately.

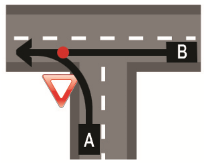


Scenario T-5.1 / 5.2 /11.2 (F: Rear End): Car A is going straight on a major road and hits car B at the rear-end, which is going straight. A is travelling on a single or dual carriageway with 30 to 60 mph speed limit without active or static yield instruction. It is caused by A failing to stop or to avoid, leading to no or slight injury.

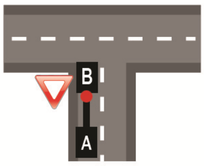


Scenario T-5.3 (M: Manoeuvring): Car A goes straight on a major road and hits an LGV or HGV B with its front, which is manoeuvring on the junction. A is travelling on a single or dual carriageway with 60 mph speed limit, and it is caused by an inappropriate manoeuvre by B. There is no clear indication on the type of traffic control.

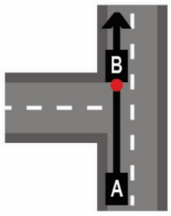
Scenario T-6.1 (M: Manoeuvring): Car A is turning left into a major road and collides with another car, cycle or PTW B, which is coming from the right. A is travelling on a rural or urban dual carriageway road. There is no clear indication on the speed limit, traffic control, point of impact (either front or offside of A) or injury severity.



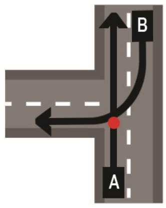
Scenario T-6.2 (F: Rear End): Car A is approaching a T-junction terminated by a major road and hits another car or PTW B on its rear end, which is waiting to turn. A is travelling on a rural single or dual carriageway road with 30 to 60 mph speed limit controlled by a static give-way sign, and it is caused by A failing to stop. The surface is dry and A remains uninjured or suffers slight injury.



Scenario T-7.1 (F: Rear End): Car A is going straight on a major road and hits another car B on its rear-end. A is travelling on a rural dual carriageway road without active or static yield instruction, and it is caused by A failing to stop or to avoid. The surface is dry and both drivers remain uninjured. There is no clear indication on the position or intended manoeuvre of B.



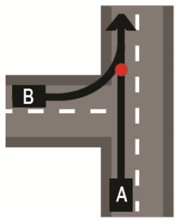
Scenario T-7.2 (L: Right Turn Against): Car A is going straight on a major road and hits another road user B with its front, who is turning right into a minor road. A is travelling on an urban road with 30 mph speed limit either without yield instruction or controlled by traffic lights, and it is caused by B failing to give way. The surface is dry and both road users remain uninjured. There is no clear indication on the opponent road user type.



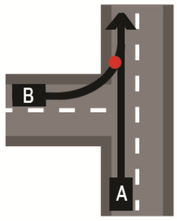
Scenario T-7.3 (H: Crossing / No Turns): Car A is going straight over a T-junction and hits another car B with its front. A is travelling on an urban road with 30 mph speed limit without active or static yield instruction, and it is caused by B failing to give way. The surface is wet and both drivers remain uninjured. There is no clear indication on the manoeuvre taken by B. Since collision type H is unlikely to happen on a T-junction, this scenario is disregarded.

Since collision type H is unlikely to happen on a T-junction, this scenario is disregarded.

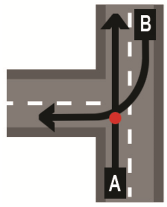
Scenario T-7.5 (M: Manoeuvring): Car A is going straight on a rural road and hits another road user B with its front, which is pulling out of a minor road from the left. This is caused by B failing to give way. The surface is dry, there are dark light conditions without street light and both drivers remain uninjured. There is no clear indication on the speed limit, carriageway type or traffic control.



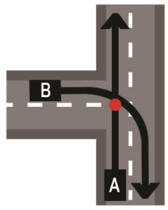
Scenario T-9.2 (M: Manoeuvring): Car A is going straight on an urban road with 30 mph speed limit and is hit by another car B on its nearside, which is pulling out of a minor road. This is caused by B failing to give way. The surface is dry. There is no clear indication on the carriageway type or traffic control.



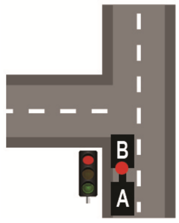
Scenario T-10.1 (L: Right Turn Against): Car A is going straight on a major road and hits another car B with its front, which is coming from the opposing direction and is turning right into a minor road. A is travelling on a single carriageway with a speed limit of 40 mph or 50 mph without active or static yield instruction, and it is caused by B failing to give way. The surface is dry and B suffers serious or fatal injury.



Scenario T-10.2 (J: Crossing/Vehicle Turning): Car A is going straight on a major road and hits another car B, which is emerging from a minor road on the left with the intention to turn right. A is travelling on a single carriageway in a rural area with a speed limit of 40 mph or 50 mph without active or static yield instructions, and it is caused by B failing to give way. The surface is wet and A suffers serious injury.



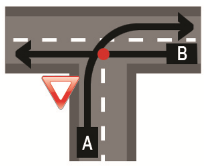
Scenario T-10.3 (F: Rear End): Car A is going straight on a major road and hits another car B on the rear end. A is travelling on a rural single carriageway road with 30 mph speed limit controlled by a traffic light, and it is caused by A failing to avoid. The surface is dry and A suffers slight injury. There is no clear indication on the position or intended manoeuvre of B.



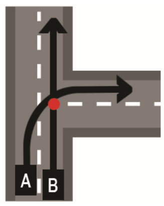
Scenario T-11.1 (M: Manoeuvring): Car A is manoeuvring (possibly making a U-turn) on a T-junction and is hit by another road user B on its offside. A is travelling on an urban and rural road, and it is caused by A manoeuvring inappropriately. There is no clear indication on the injury severity, carriageway type, surface condition, speed limit or traffic control.

Due to unclear indications on the manoeuvres, there was no figure created for this scenario.

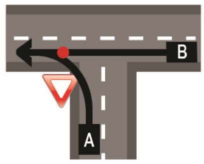
Scenario T-12.1 (J: Crossing/Vehicle Turning): Car A is turning right into a major road and is hit by a PTW B on the offside, which is going straight on the crossing path. A is travelling on a rural single carriageway controlled by a static give-way sign and it is caused by A failing to give way. The surface is wet and B suffers serious or fatal injury.



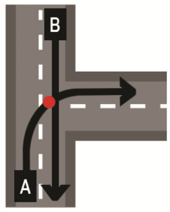
Scenario T-12.2 (G: Turning versus Same Direction): Car A is turning right into a minor road and is hit on the offside by a PTW B, which is overtaking. A is travelling on an urban single carriageway with 30 mph speed limit without active or static yield instruction, and it is caused by an inappropriate overtake from B.



Scenario T-12.3 (M: Manoeuvring): Car A is turning left into a major road and is hit by a PTW B on its offside, which is going straight on the major road from the right. A is travelling on an urban single carriageway with 30 mph speed limit controlled by give-way signs, and it is caused by A failing to give way. B suffers serious or fatal injury.



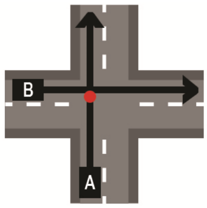
Scenario T-13.1 (L: Right Turn Against): Car A is turning right into a minor road and hits a PTW B with its front, which is going straight in the opposing direction. A is travelling on a rural single carriageway with 30 to 50 mph speed limit without active or static yield instruction, and it is caused by A failing to give way or manoeuvring inappropriately. The surface is wet and B suffers serious or fatal injury.



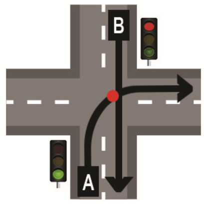
Appendix E: Overview of all scenarios at four-legged junctions

The association rules resulted in 18 scenarios for four-legged junctions, which are described on the following pages. Figures illustrate the scenarios in a simplified manner to better understand the descriptions in the text. The red dots in the figures are the points of impact (i.e. front, offside or nearside). Surface conditions, area (rural/urban), speed limits, vehicle types and injury levels are not shown. Note that all crashes in the data occurred on UK roads with left-hand traffic and that the road parameters are given for the path of A.

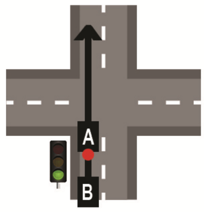
Scenario X-1.1 (H: Crossing / No Turns): Car A is going straight on a major road and hits another car B with its front, which is crossing the path from the left. A is travelling on a rural single carriageway with 60 mph speed limit without active or static yield instruction and it is caused by B failing to give way.



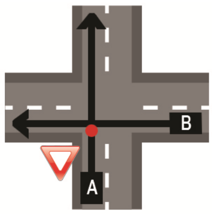
Scenario X-1.2 (L: Right Turn Against): Car A is turning right into a minor road and hits another car B, which is coming from the opposing direction. A is travelling on an urban road with 40 to 50 mph speed limit controlled by traffic lights, and it is caused by B violating the red light. Max. injury: Slight.



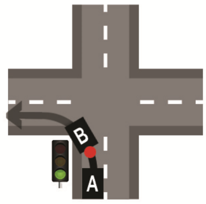
Scenario X-1.3 (F: Rear End): Car A is crossing a junction and is hit by another car B on its rear-end. A is travelling on an urban dual carriageway road with 40-50 mph speed limit controlled by traffic lights, and it is caused by B failing to avoid. Surface: Dry. Max. injury: Slight.



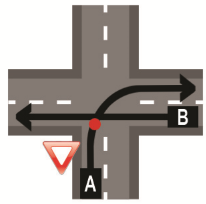
Scenario X-2.1 (H: Crossing / No Turns): Car A is crossing a four-legged junction and hits another car or PTW B with its front, which is crossing the path from the right. A is travelling on a rural single carriageway road with 40-50 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. A remains uninjured or suffers slight injury. Max. injury of B: Serious or fatal. No clear indication on the surface condition.



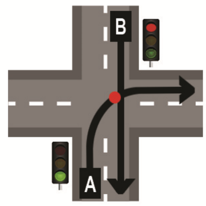
Scenario X-2.2 (F: Rear End): Car A is turning left and hits another car or LGV/HGV B at the rear end or is hit on its rear end. A is travelling on an urban or rural single carriageway road with 40-50 mph speed limit controlled by traffic lights, and it is caused by A or B failing to avoid. There is a slight left curve on the path of A. Max injury: Slight. There is no clear indication on the surface condition.



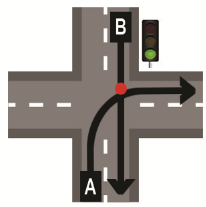
Scenario X-2.3 (J: Crossing/Vehicle Turning): Car A is turning right into a major road and hits another car B with its front, which is coming from the right. A is travelling on an urban single carriageway road with 30 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. The surface is wet and both drivers remain uninjured.



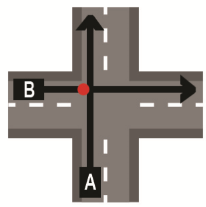
Scenario X-2.4 (L: Right Turn Against): Car A is turning right into a major road and hits another car B with its front, which is coming straight from the opposite direction. A is travelling on an urban single carriageway road with 30 mph speed limit controlled by traffic lights, and it is caused by B failing to give way. Max. injury: Slight.



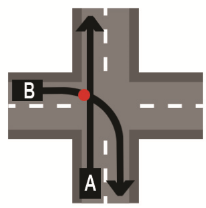
Scenario X-3.1 (L: Right Turn Against): Car A is turning right into a minor road and is hit by another car B on its nearside, which is coming from the opposite direction. A is travelling on a rural single carriageway road with 40-50 mph speed limit controlled by traffic lights, and it is caused by A failing to give way or manoeuvring inappropriately. Max. injury: Serious or fatal.



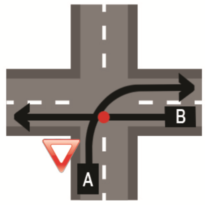
Scenario X-3.2 (H: Crossing / No Turns): Car A is going straight over a junction and is hit by another car B on its nearside, which is crossing from the left. A is travelling on a road with 30 to 50 mph speed limit without active or static yield instruction. It is caused by B failing to give way. There is no clear indication on the area (rural/urban) or injury severity.



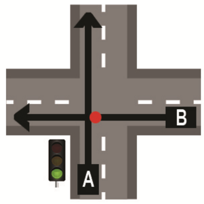
Scenario X-3.3 (J: Crossing/Vehicle Turning): Car A is going straight on a major road and is hit by another car B on its nearside. A is travelling on an urban single carriageway road with 30 mph speed limit without active or static yield instruction. It is caused by B failing to give way. The surface is wet and A suffers slight injury.



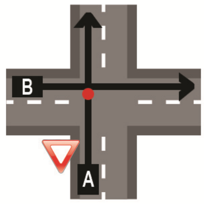
Scenario X-4.1 (J: Crossing/Vehicle Turning): Car A is turning right into a major road and is hit by a car or LGV B on the offside, which is going straight on the major road from the right. A is travelling on a rural dual carriageway road with 40-50 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. The surface is wet and A suffers serious or fatal injuries.



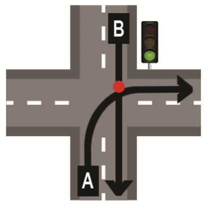
Scenario X-4.2 (H: Crossing / No Turns): Car A is crossing a major road and is hit by another car B on its offside, which is crossing from the right. A is travelling on an urban single carriageway road controlled by traffic lights, and it is caused by A failing to give way. Max. injury: Serious or fatal. There is no clear indication on the surface condition.



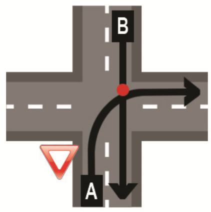
Scenario X-5.1 (H: Crossing / No Turns): Car A is crossing a major road and is hit by another car B on its nearside, which is crossing from the left. A is travelling on an urban or rural single carriageway road with 30 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. There is no clear indication on the injury severity.



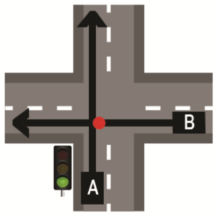
Scenario X-5.2 (L: Right Turn Against): Car A is turning right into a major road and is hit by another car B on its nearside, which is coming from the opposite direction. A is travelling on an urban road with 30 to 50 mph speed limit controlled by traffic lights, and it is caused by A failing to give way or manoeuvring inappropriately. The surface is dry and A suffers no or slight injuries. There is no clear indication on the carriageway type.



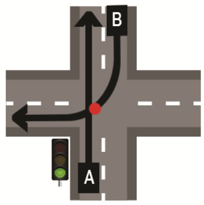
Scenario X-5.3 (M: Manoeuvring): Car A is manoeuvring on a junction and is hit by another car B on its nearside. A is travelling on a single carriageway road with 40 to 50 mph speed limit controlled by static give-way signs, and it is caused by A failing to give way. Max. injury: Slight. There is no clear indication on the manoeuvre by A or B.



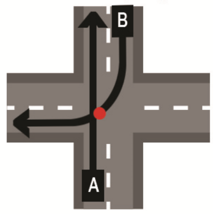
Scenario X-6.1 (H: Crossing / No Turns): Car A is going straight on a major road and is hit by car B on the offside, which comes from a minor road and crosses the path from the right. A is travelling on a single carriageway road with 30 mph speed limit controlled by traffic lights, and it is caused by B failing to give way. The surface is wet and B suffers serious or fatal injuries.



Scenario X-6.2 (L: Right Turn Against): Car A is going straight on a major road and is hit by car B on its offside, which turns right from the opposing direction. A is travelling on a road with 60 mph speed limit controlled by traffic lights, and it is caused by B losing control of the vehicle. B suffers serious or fatal injuries.



Scenario X-6.3 (M: Manoeuvring): Car A is going straight on a major road and is hit by another car B on its offside, which is manoeuvring on the junction. A is travelling on an urban single carriageway road with 30 to 50 mph speed limit without active or static yield instruction, and it is caused by B failing to give way. The surface is dry and A suffers slight injury.



Appendix F: Association rule tables and graphs for each cluster

On the following pages, the association rules obtained for each cluster (13 for T-junctions and 6 for four-legged junctions) and collision type (see Appendix B) are given. Note that only rules with up to three items are presented, sorted by the support value. Furthermore, the network graphs are depicted for each cluster, which represent the amount of associations identified between the respective antecedent node and the given consequent in the centre. The weight or thickness of each edge represents the number of associations identified between the respective antecedent node and the given consequent in the centre. In other words, nodes with thick edges indicate dominant crash attributes and thus define the scenario. For antecedent nodes that are not present in the graph, there were no associations found in the rules. Thus they can be considered negligible for the respective scenario.

Cluster T-C1: "The car hits another car or LGV with its front, while going straight on a road with a minor roads joining from the left."

Table 37: All rules (up to 3 items) obtained for cluster T-C1 with collision type F (scenario T-1.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=F & RdType=DualCgw	Surf=Dry	0.101	0.773	1.389
Coll=F & RdType=SingCgw	DrvInj=Uninjured	0.095	0.762	1.839
Coll=F & SpdLim=70mph	Surf=Dry	0.077	0.867	1.558
Coll=F & Prec=FailStopDriver	FirstImpact=Front	0.065	1.000	1.310
Coll=F & Prec=FailAvoidOther	RdType=DualCgw	0.053	0.818	3.841
Coll=F & Prec=FailStopDriver	DrvInj=Uninjured	0.053	0.818	1.975
Coll=F & Prec=FailAvoidDriver	RdType=DualCgw	0.047	0.800	3.756
Coll=F & Prec=FailAvoidDriver	DrvInj=Uninjured	0.047	0.800	1.931
Coll=F & SpdLim=60mph	Area=Rural	0.041	1.000	2.600
Coll=F & Light=DarkSL	DrvInj=Uninjured	0.041	0.875	2.113
Coll=F & SpdLim=60mph	DrvInj=Uninjured	0.036	0.857	2.069
Coll=F & Prec=FailStopOther	DrvInj=Uninjured	0.036	0.750	1.811
Coll=F & SpdLim=40=50mph	Area=Urban	0.036	1.000	1.625
Coll=F & Prec=FailStopOther	Light=DayNSL	0.036	0.750	1.440
Coll=F & SpdLim=40=50mph	HorizGeom=Straight	0.036	1.000	1.420
Coll=F & TrfCtrl=GW	Surf=Dry	0.036	0.750	1.348

Table 38: All rules (up to 3 items) obtained for cluster T-C1 with collision type J (scenario T-1.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J & TrfCtrl=None	RdType=SingCgw	0.219	0.949	1.253
Coll=J & TrfCtrl=None	Prec=FailGiveWayOther	0.183	0.795	2.099
Coll=J & Prec=FailGiveWayOther	TrfCtrl=None	0.183	1.000	1.280
Coll=J & RdType=SingCgw	Prec=FailGiveWayOther	0.178	0.769	2.031
Coll=J & Prec=FailGiveWayOther	RdType=SingCgw	0.178	0.968	1.278
Coll=J & DrvInj=Slight	Prec=FailGiveWayOther	0.136	0.767	2.024
Coll=J & SpdLim=30mph	TrfCtrl=None	0.136	1.000	1.280
Coll=J & Surf=Dry	FirstImpact=Front	0.130	0.957	1.253
Coll=J & Light=DayNSL	DrvInj=Slight	0.124	0.955	1.629
Coll=J & Light=DayNSL	RdType=SingCgw	0.124	0.955	1.260
Coll=J & SpdLim=30mph	Prec=FailGiveWayOther	0.107	0.783	2.067
Coll=J & Surf=Dry	DrvInj=Slight	0.107	0.783	1.336
Coll=J & Surf=Wet	RdType=SingCgw	0.107	0.947	1.251
Coll=J & Light=DayNSL	Prec=FailGiveWayOther	0.101	0.773	2.040
Coll=J & Area=Rural	FirstImpact=Front	0.101	1.000	1.310
Coll=J & Surf=Wet	Prec=FailGiveWayOther	0.089	0.789	2.085
Coll=J & Area=Rural	DrvInj=Slight	0.089	0.882	1.506
Coll=J & SpdLim=40=50mph	FirstImpact=Front	0.089	1.000	1.310
Coll=J & SpdLim=40=50mph	Light=DayNSL	0.071	0.800	1.536

Coll=J & SpdLim=40=50mph	DrvInj=Slight	0.071	0.800	1.366
Coll=J & DrvInj=Uninjured	Area=Urban	0.065	0.846	1.375
Coll=J & Light=DaySLUnk	FirstImpact=Front	0.065	1.000	1.310
Coll=J & Light=DarkSL	TrfCtrl=None	0.053	1.000	1.280
Coll=J & Light=DarkSL	Area=Urban	0.047	0.889	1.444
Coll=J & HorizGeom=RightSlight	Prec=FailGiveWayOther	0.041	0.778	2.054
Coll=J & Light=DarkSL	Prec=FailGiveWayOther	0.041	0.778	2.054
Coll=J & Light=DarkSL	SpdLim=30mph	0.041	0.778	1.546
Coll=J & HorizGeom=RightSlight	DrvInj=Slight	0.041	0.778	1.328
Coll=J & Prec=PoorMnvrOther	TrfCtrl=None	0.041	1.000	1.280
Coll=J & Prec=PoorMnvrOther	Surf=Dry	0.036	0.857	1.541

Table 39: All rules (up to 3 items) obtained for cluster T-C1 with collision type L (scenario T-1.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L	DrvInj=Slight	0.189	0.762	1.301
Coll=L & Manvr=GoingAheadOther	DrvInj=Slight	0.183	0.775	1.323
Coll=L & TrfCtrl=None	RdType=SingCgw	0.178	0.968	1.278
Coll=L & HorizGeom=Straight	DrvInj=Slight	0.142	0.774	1.322
Coll=L & Area=Rural	RdType=SingCgw	0.118	0.952	1.257
Coll=L & Prec=FailGiveWayOther	DrvInj=Slight	0.112	0.826	1.410
Coll=L & Light=DayNSL	RdType=SingCgw	0.112	0.950	1.254
Coll=L & Area=Rural	DrvInj=Slight	0.101	0.810	1.382
Coll=L & Light=DayNSL	DrvInj=Slight	0.095	0.800	1.366
Coll=L & Light=DayNSL	Area=Rural	0.089	0.750	1.950
Coll=L & Surf=Wet	DrvInj=Slight	0.089	0.882	1.506
Coll=L & SpdLim=30mph	DrvInj=Slight	0.083	0.778	1.328
Coll=L & Prec=PoorMnvrOther	RdType=SingCgw	0.065	1.000	1.320
Coll=L & DrvInj=Uninjured	RdType=SingCgw	0.059	1.000	1.320
Coll=L & TrfCtrl=Light	DrvInj=Slight	0.053	0.900	1.536
Coll=L & Prec=PoorMnvrOther	Surf=Dry	0.053	0.818	1.471
Coll=L & TrfCtrl=Light	HorizGeom=Straight	0.053	0.900	1.278
Coll=L & HorizGeom=RightSlight	RdType=SingCgw	0.047	1.000	1.320
Coll=L & Light=DarkSL	HorizGeom=Straight	0.047	1.000	1.420
Coll=L & Light=DarkSL	FirstImpact=Front	0.047	1.000	1.310
Coll=L & HorizGeom=RightSlight	DrvInj=Slight	0.041	0.875	1.494
Coll=L & Light=DarkSL	Area=Urban	0.041	0.875	1.422
Coll=L & HorizGeom=RightSlight	Prec=FailGiveWayOther	0.036	0.750	1.980
Coll=L & Light=DarkSL	Surf=Wet	0.036	0.750	1.760

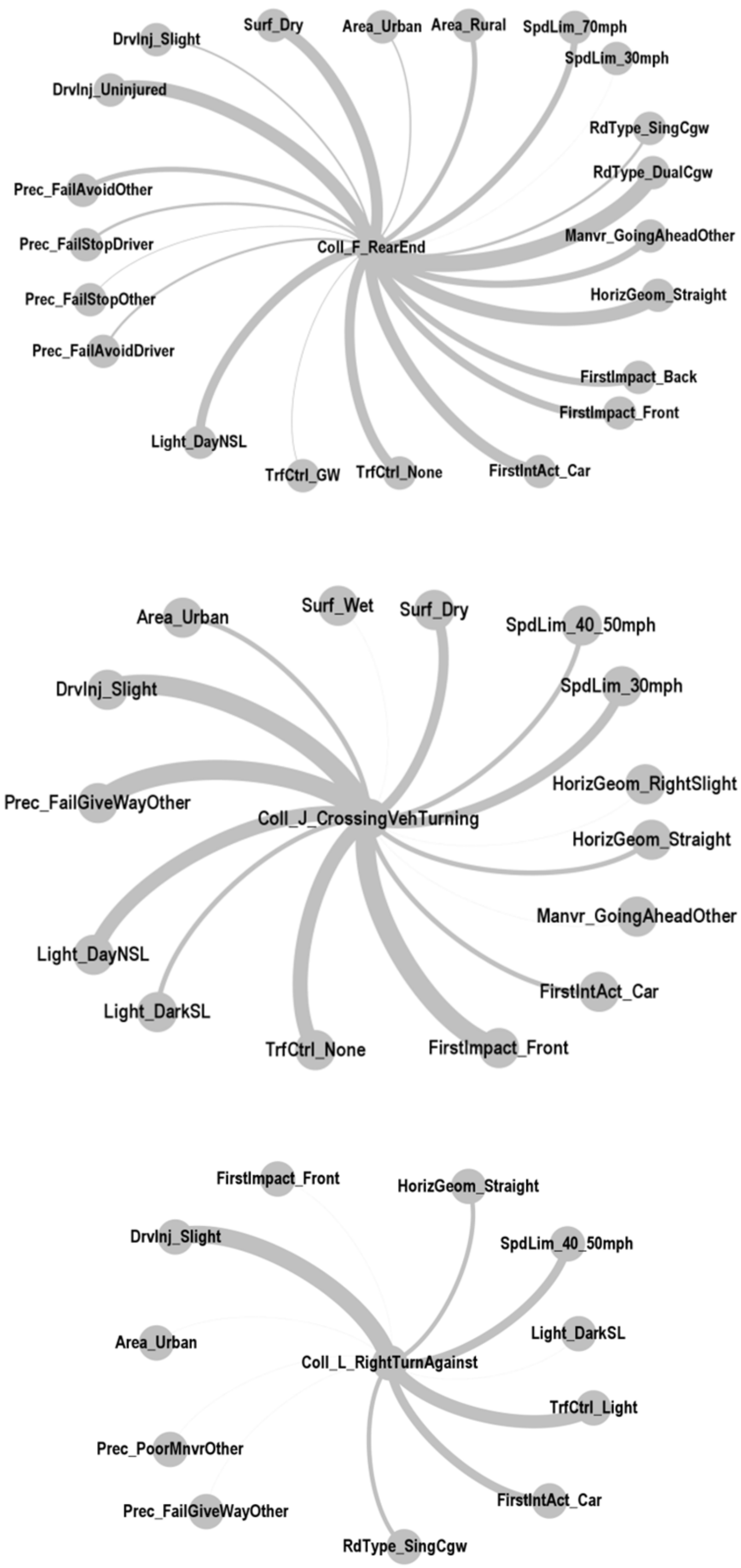


Figure 127: Weighted, directed graphs obtained from all association rules for cluster T-C1

Cluster T-C2: "The car hits another car or PTW with its front, while turning right into a major road."

Table 40: All rules (up to 3 items) obtained for cluster T-C2 with collision type J (scenario T-2.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J & Area=Urban	FirstImpact=Front	0.473	0.972	1.262
Coll=J & Prec=FailGiveWayDriver	FirstImpact=Front	0.446	0.971	1.260
Coll=J & TrfCtrl=GW	Prec=FailGiveWayDriver	0.419	0.861	1.300
Coll=J & FirstIntAct=Car	FirstImpact=Front	0.365	0.964	1.252
Coll=J & SpdLim=40-50mph	FirstImpact=Front	0.216	1.000	1.298
Coll=J & Surf=Wet	FirstImpact=Front	0.216	1.000	1.298
Coll=J & DrvInj=Slight	FirstImpact=Front	0.189	1.000	1.298
Coll=J & DrvInj=Slight	Light=DayNSL	0.162	0.857	1.669
Coll=J & Light=DaySLUnk	Surf=Dry	0.135	1.000	1.805
Coll=J & Light=DaySLUnk	DrvInj=Uninjured	0.135	1.000	1.480
Coll=J & Light=DaySLUnk	Area=Urban	0.122	0.900	1.332
Coll=J & Light=DaySLUnk	SpdLim=30mph	0.108	0.800	1.444
Coll=J & TrfCtrl=Light	FirstImpact=Front	0.081	1.000	1.298
Coll=J & MaxInj=Uninjured	FirstImpact=Front	0.081	1.000	1.298
Coll=J & TrfCtrl=Light	Prec=FailStopOther	0.068	0.833	8.810
Coll=J & Prec=FailStopOther	SpdLim=40-50mph	0.068	1.000	3.217
Coll=J & Prec=FailStopOther	FirstImpact=Front	0.068	1.000	1.298
Coll=J & TrfCtrl=Light	SpdLim=40-50mph	0.068	0.833	2.681
Coll=J & TrfCtrl=None	Area=Urban	0.068	1.000	1.480
Coll=J & TrfCtrl=None	FirstImpact=Front	0.068	1.000	1.298
Coll=J & Prec=PoorMnvrDriver	DrvInj=Uninjured	0.054	1.000	1.480
Coll=J & Prec=PoorMnvrDriver	Surf=Dry	0.041	0.750	1.354
Coll=J & MaxInj=SeriousFatal	Area=Urban	0.041	1.000	1.480
Coll=J & MaxInj=SeriousFatal	FirstIntAct=Car	0.041	1.000	1.451
Coll=J & MaxInj=SeriousFatal	FirstImpact=Front	0.041	1.000	1.298

Table 41: All rules (up to 3 items) obtained for cluster T-C2 with collision type M (scenario T-2.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=M & Prec=FailGiveWayDriver	MaxInj=Slight	0.095	1.000	1.276
Coll=M & MaxInj=Slight	Prec=FailGiveWayDriver	0.095	0.875	1.321
Coll=M & HorizGeom=Straight	Light=DayNSL	0.081	0.750	1.461
Coll=M & RdType=SingCgw	Prec=FailGiveWayDriver	0.081	0.857	1.294
Coll=M & DrvInj=Uninjured	MaxInj=Slight	0.081	1.000	1.276
Coll=M & DrvInj=Uninjured	Prec=FailGiveWayDriver	0.068	0.833	1.259
Coll=M & TrfCtrl=GW	Prec=FailGiveWayDriver	0.068	1.000	1.510
Coll=M & TrfCtrl=GW	MaxInj=Slight	0.068	1.000	1.276
Coll=M & Area=Rural	FirstImpact=Nearside	0.054	0.800	4.933
Coll=M & Area=Rural	Light=DayNSL	0.054	0.800	1.558

Coll=M & Surf=Dry	Light=DayNSL	0.054	1.000	1.947
Coll=M & SpdLim=30mph	Prec=FailGiveWayDriver	0.054	1.000	1.510
Coll=M & SpdLim=30mph	Area=Urban	0.054	1.000	1.480
Coll=M & Area=Urban	SpdLim=30mph	0.054	1.000	1.805
Coll=M & SpdLim=30mph	MaxInj=Slight	0.054	1.000	1.276
Coll=M & Surf=Dry	MaxInj=Slight	0.054	1.000	1.276
Coll=M & Area=Urban	Prec=FailGiveWayDriver	0.054	1.000	1.510
Coll=M & Area=Urban	MaxInj=Slight	0.054	1.000	1.276
Coll=M & Surf=Dry	FirstImpact=Nearside	0.041	0.750	4.625
Coll=M & Surf=Dry	Area=Rural	0.041	0.750	2.313

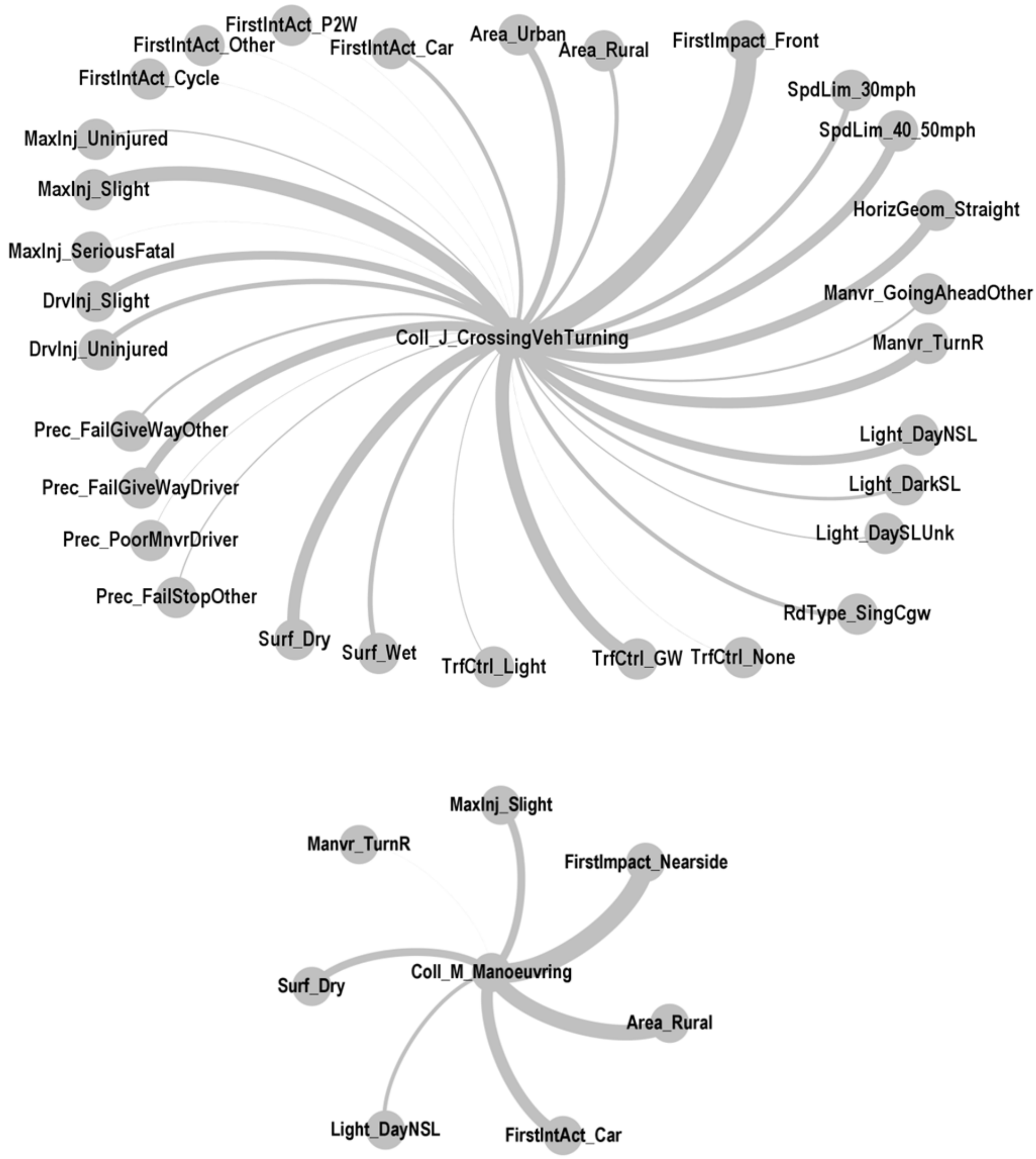


Figure 128: Weighted, directed graphs obtained from all association rules for cluster T-C2

Cluster T-C3: "The car is hit on its back, while waiting to turn right into a minor road."

Table 42: All rules (up to 3 items) obtained for cluster T-C3 with collision type F (scenario T-3.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=F & Manvr=WaitTurnR	DrvInj=Slight	0.333	0.818	1.425
Coll=F & Area=Urban	SpdLim=30mph	0.296	0.800	1.878
Coll=F & Light=DayNSL	Prec=FailStopOther	0.296	0.762	1.524
Coll=F & Area=Rural	SpdLim=60mph	0.278	0.789	2.368
Coll=F & SpdLim=60mph	DrvInj=Slight	0.222	0.800	1.394
Coll=F & Light=DarkSL	Surf=Wet	0.111	0.857	2.571
Coll=F & RdType=DualCgw	Area=Urban	0.074	1.000	1.929
Coll=F & SpdLim=40-50mph	Area=Urban	0.074	0.800	1.543
Coll=F & TrfCtrl=Light	Surf=Wet	0.056	1.000	3.000
Coll=F & Manvr=TurnR	Surf=Dry	0.056	1.000	1.543
Coll=F & RdType=DualCgw	Light=DaySLUnk	0.056	0.750	4.500
Coll=F & FirstIntAct=LGV=HGV	Surf=Wet	0.056	1.000	3.000
Coll=F & FirstIntAct=LGV=HGV	Area=Rural	0.056	1.000	2.077
Coll=F & MaxInj=Uninjured	Surf=Wet	0.037	1.000	3.000

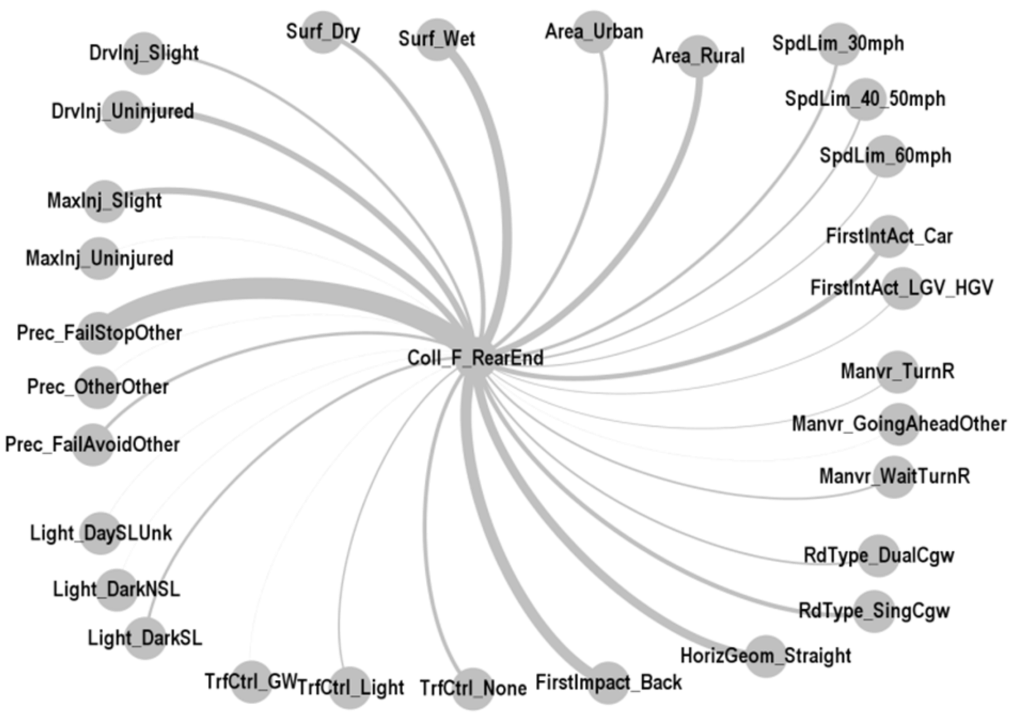


Figure 129: Weighted, directed graph obtained from all association rules for cluster T-C3

Cluster T-C4: *"The car is hit on the nearside, while turning right into a minor road."*

Table 43: All rules (up to 3 items) obtained for cluster T-C4 with collision type L (scenario T-4.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L & SpdLim=40-50mph	FirstIntAct=Car	0.276	0.842	1.357
Coll=L & SpdLim=30mph	Area=Urban	0.259	1.000	1.450
Coll=L & Area=Rural	FirstIntAct=Car	0.190	1.000	1.611
Coll=L & Light=DayNSL	Prec=FailGiveWayDriver	0.190	0.786	1.899
Coll=L & Light=DarkSL	FirstImpact=Nearside	0.155	1.000	1.289
Coll=L & Area=Rural	SpdLim=40-50mph	0.155	0.818	2.260
Coll=L & Area=Rural	Light=DayNSL	0.155	0.818	1.977
Coll=L & MaxInj=SeriousFatal	Surf=Dry	0.138	0.800	1.326
Coll=L & FirstIntAct=P2W	Area=Urban	0.086	1.000	1.450
Coll=L & FirstIntAct=P2W	FirstImpact=Nearside	0.086	1.000	1.289
Coll=L & MaxInj=Uninjured	FirstImpact=Nearside	0.069	1.000	1.289
Coll=L & MaxInj=Uninjured	Surf=Wet	0.052	0.750	2.175
Coll=L & MaxInj=Uninjured	Prec=FailGiveWayDriver	0.052	0.750	1.813
Coll=L & MaxInj=Uninjured	SpdLim=30mph	0.052	0.750	1.450
Coll=L & SpdLim=60mph	FirstImpact=Nearside	0.052	1.000	1.289
Coll=L & FirstImpact=Offside	FirstIntAct=Car	0.052	1.000	1.611
Coll=L & TrfCtrl=GW	Surf=Dry	0.034	1.000	1.657
Coll=L & DrvInj=Serious	HorizGeom=LeftSlight	0.034	1.000	11.600
Coll=L & DrvInj=Serious	Area=Rural	0.034	1.000	3.222
Coll=L & DrvInj=Serious	SpdLim=40-50mph	0.034	1.000	2.762
Coll=L & DrvInj=Serious	FirstImpact=Nearside	0.034	1.000	1.289

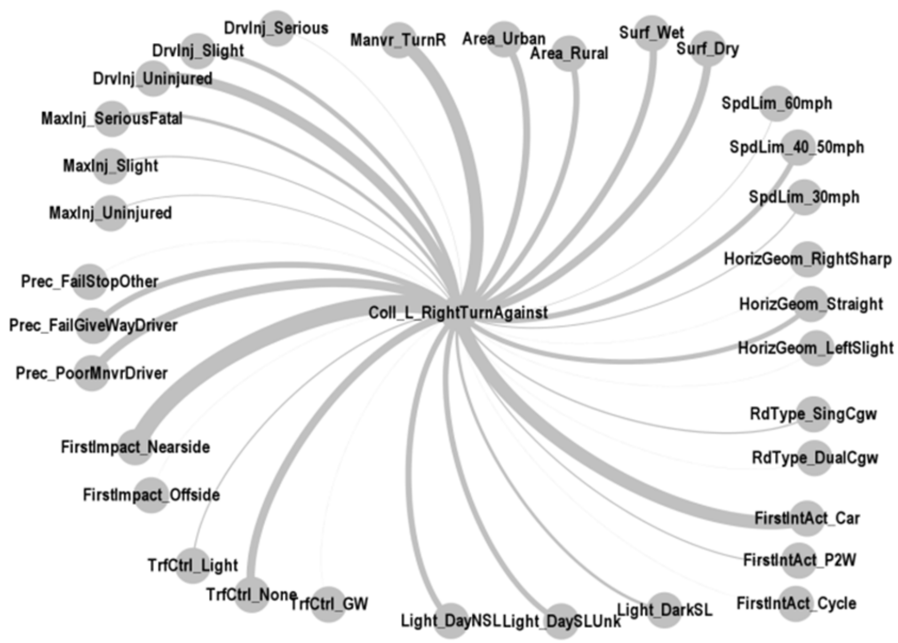


Figure 130: Weighted, directed graph obtained from all association rules for cluster T-C4

Cluster T-C5: "The car hits another car with its front, while going straight over a T-junction with a minor roads joining from the right."

Table 44: All rules (up to 3 items) obtained for cluster T-C5 with collision type F (scenario T-5.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=F & DrvInj=Slight	FirstIntAct=Car	0.218	1.000	1.261
Coll=F & Light=DaySLUnk	HorizGeom=Straight	0.184	1.000	1.279
Coll=F & Light=DaySLUnk	Area=Urban	0.161	0.875	1.313
Coll=F & SpdLim=60mph	FirstIntAct=Car	0.115	1.000	1.261
Coll=F & RdType=DualCgw	Area=Urban	0.115	1.000	1.500
Coll=F & RdType=DualCgw	FirstIntAct=Car	0.115	1.000	1.261
Coll=F & SpdLim=60mph	Area=Rural	0.103	0.900	2.700
Coll=F & TrfCtrl=Light	Area=Urban	0.080	0.875	1.313
Coll=F & SpdLim=40-50mph	RdType=DualCgw	0.080	0.875	5.075
Coll=F & SpdLim=40-50mph	Area=Urban	0.080	0.875	1.313
Coll=F & SpdLim=40-50mph	DrvInj=Uninjured	0.069	0.750	1.332
Coll=F & MaxInj=Uninjured	SpdLim=30mph	0.069	1.000	1.813
Coll=F & MaxInj=Uninjured	Area=Urban	0.069	1.000	1.500
Coll=F & MaxInj=Uninjured	HorizGeom=Straight	0.069	1.000	1.279
Coll=F & MaxInj=Uninjured	Light=DaySLUnk	0.057	0.833	3.452
Coll=F & Light=DarkSL	FirstIntAct=Car	0.057	1.000	1.261
Coll=F & Light=DarkSL	SpdLim=30mph	0.046	0.800	1.450
Coll=F & HorizGeom=LeftSlight	Area=Rural	0.034	1.000	3.000
Coll=F & HorizGeom=LeftSlight	FirstIntAct=Car	0.034	1.000	1.261

Table 45: All rules (up to 3 items) obtained for cluster T-C5 with collision type G (scenario T-5.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=G & Manvr=GoingAheadOther	Light=DayNSL	0.115	1.000	1.673
Coll=G & Surf=Dry	HorizGeom=Straight	0.103	1.000	1.279
Coll=G & MaxInj=Slight	Light=DayNSL	0.103	1.000	1.673
Coll=G & HorizGeom=Straight	Surf=Dry	0.103	0.900	1.477
Coll=G & HorizGeom=Straight	DrvInj=Uninjured	0.092	0.800	1.420
Coll=G & DrvInj=Uninjured	HorizGeom=Straight	0.092	1.000	1.279
Coll=G & Surf=Dry	DrvInj=Uninjured	0.080	0.778	1.381
Coll=G & Prec=FailAvoidDriver	Light=DayNSL	0.080	1.000	1.673
Coll=G & Prec=FailAvoidDriver	HorizGeom=Straight	0.080	1.000	1.279
Coll=G & FirstIntAct=Car	DrvInj=Uninjured	0.080	0.778	1.381
Coll=G & DrvInj=Uninjured	Surf=Dry	0.080	0.875	1.436
Coll=G & Area=Urban	Surf=Dry	0.080	0.875	1.436
Coll=G & SpdLim=30mph	Area=Urban	0.069	1.000	1.500
Coll=G & Prec=FailAvoidDriver	DrvInj=Uninjured	0.069	0.857	1.522
Coll=G & Prec=FailAvoidDriver	Surf=Dry	0.069	0.857	1.407
Coll=G & DrvInj=Uninjured	Prec=FailAvoidDriver	0.069	0.750	2.719

Coll=G & SpdLim=30mph	Surf=Dry	0.057	0.833	1.368
Coll=G & SpdLim=60mph	Prec=FailAvoidDriver	0.034	1.000	3.625
Coll=G & SpdLim=60mph	Area=Rural	0.034	1.000	3.000
Coll=G & SpdLim=60mph	DrvInj=Uninjured	0.034	1.000	1.776
Coll=G & SpdLim=60mph	Light=DayNSL	0.034	1.000	1.673
Coll=G & SpdLim=60mph	HorizGeom=Straight	0.034	1.000	1.279
Coll=G & SpdLim=60mph	FirstIntAct=Car	0.034	1.000	1.261
Coll=G & DrvInj=Slight	Light=DayNSL	0.034	1.000	1.673
Coll=G & DrvInj=Slight	Area=Urban	0.034	1.000	1.500
Coll=G & Area=Rural	SpdLim=60mph	0.034	1.000	4.833
Coll=G & Area=Rural	Prec=FailAvoidDriver	0.034	1.000	3.625
Coll=G & Area=Rural	DrvInj=Uninjured	0.034	1.000	1.776
Coll=G & Area=Rural	Light=DayNSL	0.034	1.000	1.673
Coll=G & Area=Rural	HorizGeom=Straight	0.034	1.000	1.279
Coll=G & Area=Rural	FirstIntAct=Car	0.034	1.000	1.261

Table 46: All rules (up to 3 items) obtained for cluster T-C5 with collision type M (scenario T-5.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll_M & Surf=Dry	Light=DayNSL	0.092	1.000	1.673
Coll_M & Light=DayNSL	Surf=Dry	0.092	0.889	1.459
Coll_M & MaxInj=Slight	Light=DayNSL	0.080	0.875	1.464
Coll_M & MaxInj=Slight	Surf=Dry	0.080	0.875	1.436
Coll_M & Area=Rural	HorizGeom=Straight	0.069	1.000	1.279
Coll_M & FirstIntAct=Car	Area=Rural	0.057	0.833	2.500
Coll_M & FirstIntAct=Car	Light=DayNSL	0.057	0.833	1.394
Coll_M & DrvInj=Slight	Light=DayNSL	0.057	1.000	1.673
Coll_M & Area=Rural	Light=DayNSL	0.057	0.833	1.394
Coll_M & SpdLim=30mph	Area=Urban	0.046	1.000	1.500
Coll_M & DrvInj=Uninjured	Surf=Dry	0.046	0.800	1.313
Coll_M & DrvInj=Slight	Surf=Dry	0.046	0.800	1.313
Coll_M & Area=Urban	SpdLim=30mph	0.046	0.800	1.450
Coll_M & Area=Urban	Surf=Dry	0.046	0.800	1.313
Coll_M & Surf=Wet	HorizGeom=Straight	0.034	1.000	1.279
Coll_M & SpdLim=60mph	Prec=PoorMnvrOther	0.034	0.750	8.156
Coll_M & SpdLim=60mph	DrvInj=Uninjured	0.034	0.750	1.332
Coll_M & Prec=PoorMnvrOther	SpdLim=60mph	0.034	0.750	3.625
Coll_M & FirstIntAct=LGV_HGV	HorizGeom=Straight	0.034	1.000	1.279

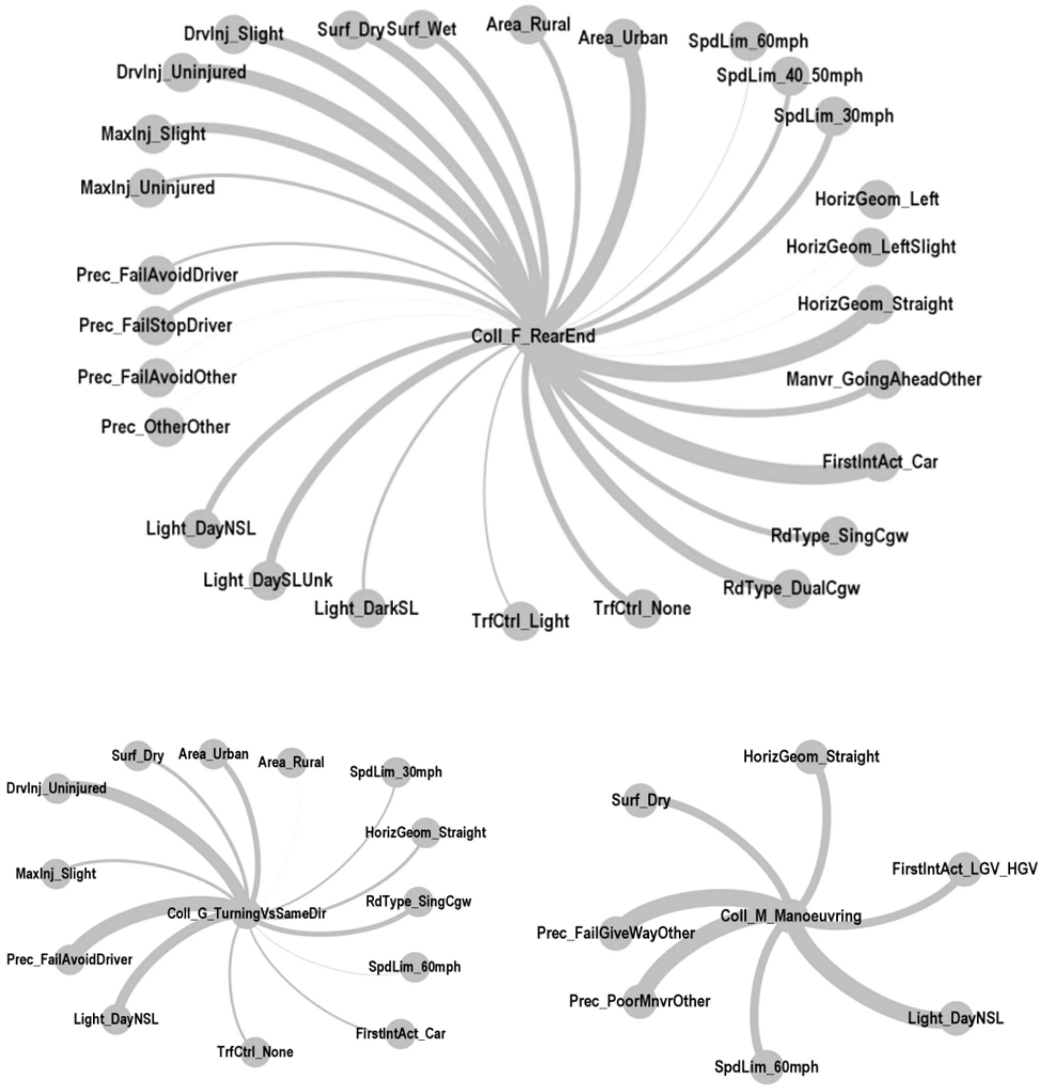


Figure 131: Weighted, directed graphs obtained from all association rules for cluster T-C5

Cluster T-C6: "The car collides with another road user, while turning left into a major road."

Table 47: Rules (up to 3 items) obtained for cluster T-C6 with collision type M (scenario T-6.1), sorted by the five highest support values

Antecedent	Consequent	Supp	Conf	Lift
Coll=M & Prec=FailGiveWayDriver	DrvInj=Uninjured	0.219	1.000	1.333
Coll=M & DrvInj=Uninjured	Prec=FailGiveWayDriver	0.219	0.778	2.263
Coll=M & HorizGeom=Straight	Prec=FailGiveWayDriver	0.219	0.778	2.263
Coll=M & Area=Urban	Surf=Dry	0.219	0.875	1.333
Coll=M & Prec=FailGiveWayDriver	Surf=Dry	0.188	0.857	1.306
Coll=M & Surf=Dry	Prec=FailGiveWayDriver	0.188	0.750	2.182
Coll=M & SpdLim=30mph	DrvInj=Uninjured	0.188	1.000	1.333
Coll=M & FirstImpact=Offside	Surf=Dry	0.156	0.833	1.270
Coll=M & Light=DayNSL	Surf=Dry	0.156	1.000	1.524
Coll=M & Light=DayNSL	Area=Urban	0.156	1.000	1.391
Coll=M & Area=Rural	Light=DarkNSL	0.094	1.000	10.667
Coll=M & FirstIntAct=Cycle	SpdLim=30mph	0.094	1.000	1.455
Coll=M & TrfCtrl=None	Light=DayNSL	0.094	1.000	1.684
Coll=M & TrfCtrl=None	Surf=Dry	0.094	1.000	1.524
Coll=M & Area=Rural	FirstIntAct=Car	0.094	1.000	1.600
Coll=M & FirstImpact=Front	Prec=FailGiveWayDriver	0.094	1.000	2.909
Coll=M & FirstImpact=Front	DrvInj=Uninjured	0.094	1.000	1.333
Coll=M & SpdLim=60mph	Light=DarkNSL	0.063	1.000	10.667
Coll=M & SpdLim=60mph	FirstIntAct=Car	0.063	1.000	1.600
Coll=M & FirstIntAct=P2W	Prec=FailGiveWayDriver	0.063	1.000	2.909
Coll=M & FirstIntAct=P2W	FirstImpact=Offside	0.063	1.000	2.667
Coll=M & FirstIntAct=P2W	Area=Urban	0.063	1.000	1.391
Coll=M & FirstIntAct=P2W	MaxInj=Slight	0.063	1.000	1.333
Coll=M & Light=DaySLUnk	Surf=Dry	0.063	1.000	1.524
Coll=M & SpdLim=40-50mph	FirstImpact=Offside	0.063	1.000	2.667
Coll=M & SpdLim=40-50mph	Surf=Dry	0.063	1.000	1.524
Coll=M & SpdLim=40-50mph	Area=Urban	0.063	1.000	1.391
Coll=M & SpdLim=40-50mph	MaxInj=Slight	0.063	1.000	1.333
Coll=M & MaxInj=Uninjured	Prec=FailGiveWayDriver	0.063	1.000	2.909
Coll=M & MaxInj=Uninjured	FirstIntAct=Car	0.063	1.000	1.600
Coll=M & MaxInj=Uninjured	Surf=Dry	0.063	1.000	1.524
Coll=M & DrvInj=Slight	FirstIntAct=Car	0.063	1.000	1.600

Table 48: Rules (up to 3 items) obtained for cluster T-C6 with collision type F (scenario T-6.2), sorted by the five highest support values

Antecedent	Consequent	Supp	Conf	Lift
Coll=F	Light=DayNSL	0.188	0.857	1.444
Coll=F	Surf=Dry	0.188	0.857	1.306

Coll=F	& Light=DayNSL	Surf=Dry	0.188	1.000	1.524
Coll=F	& Surf=Dry	Light=DayNSL	0.188	1.000	1.684
Coll=F	& Light=DayNSL	Prec=FailAvoidDriver	0.156	0.833	4.444
Coll=F	& Prec=FailAvoidDriver	Surf=Dry	0.156	1.000	1.524
Coll=F	& Surf=Dry	Prec=FailAvoidDriver	0.156	0.833	4.444
Coll=F	& DrvInj=Uninjured	Prec=FailAvoidDriver	0.156	1.000	5.333
Coll=F	& RdType=SingCgw	Prec=FailAvoidDriver	0.156	0.833	4.444
Coll=F	& DrvInj=Uninjured	Light=DayNSL	0.156	1.000	1.684
Coll=F	& HorizGeom=Straight	Light=DayNSL	0.156	1.000	1.684
Coll=F	& DrvInj=Uninjured	Surf=Dry	0.156	1.000	1.524
Coll=F	& HorizGeom=Straight	Surf=Dry	0.156	1.000	1.524
Coll=F	& HorizGeom=Straight	MaxInj=Slight	0.156	1.000	1.333
Coll=F	& FirstImpact=Front	Prec=FailAvoidDriver	0.125	1.000	5.333
Coll=F	& HorizGeom=Straight	Prec=FailAvoidDriver	0.125	0.800	4.267
Coll=F	& FirstImpact=Front	Light=DayNSL	0.125	1.000	1.684
Coll=F	& FirstImpact=Front	Surf=Dry	0.125	1.000	1.524
Coll=F	& FirstImpact=Front	MaxInj=Slight	0.125	1.000	1.333
Coll=F	& FirstImpact=Front	DrvInj=Uninjured	0.125	1.000	1.333
Coll=F	& DrvInj=Uninjured	FirstImpact=Front	0.125	0.800	1.969
Coll=F	& HorizGeom=Straight	FirstImpact=Front	0.125	0.800	1.969
Coll=F	& FirstIntAct=Car	MaxInj=Slight	0.125	1.000	1.333
Coll=F	& FirstIntAct=Car	Area=Rural	0.094	0.750	2.667
Coll=F	& SpdLim=30mph	MaxInj=Slight	0.094	1.000	1.333
Coll=F	& Prec=FailStopOther	FirstImpact=Back	0.063	1.000	8.000
Coll=F	& FirstImpact=Back	Prec=FailStopOther	0.063	1.000	10.667
Coll=F	& Prec=FailStopOther	DrvInj=Slight	0.063	1.000	4.000
Coll=F	& DrvInj=Slight	Prec=FailStopOther	0.063	1.000	10.667
Coll=F	& Prec=FailStopOther	Area=Rural	0.063	1.000	3.556
Coll=F	& Prec=FailStopOther	FirstIntAct=Car	0.063	1.000	1.600
Coll=F	& FirstIntAct=P2W	Prec=FailAvoidDriver	0.063	1.000	5.333
Coll=F	& FirstIntAct=P2W	Light=DayNSL	0.063	1.000	1.684
Coll=F	& DrvInj=Slight	FirstImpact=Back	0.063	1.000	8.000
Coll=F	& SpdLim=40-50mph	Prec=FailAvoidDriver	0.063	1.000	5.333
Coll=F	& SpdLim=40-50mph	Area=Rural	0.063	1.000	3.556
Coll=F	& SpdLim=40-50mph	Surf=Dry	0.063	1.000	1.524
Coll=F	& SpdLim=40-50mph	DrvInj=Uninjured	0.063	1.000	1.333
Coll=F	& Area=Urban	Prec=FailAvoidDriver	0.063	1.000	5.333
Coll=F	& DrvInj=Slight	Area=Rural	0.063	1.000	3.556
Coll=F	& DrvInj=Slight	FirstIntAct=Car	0.063	1.000	1.600
Coll=F	& Area=Urban	FirstImpact=Front	0.063	1.000	2.462
Coll=F	& Area=Urban	Light=DayNSL	0.063	1.000	1.684
Coll=F	& Area=Urban	Surf=Dry	0.063	1.000	1.524
Coll=F	& Area=Urban	SpdLim=30mph	0.063	1.000	1.455
Coll=F	& Area=Urban	MaxInj=Slight	0.063	1.000	1.333
Coll=F	& Area=Urban	DrvInj=Uninjured	0.063	1.000	1.333

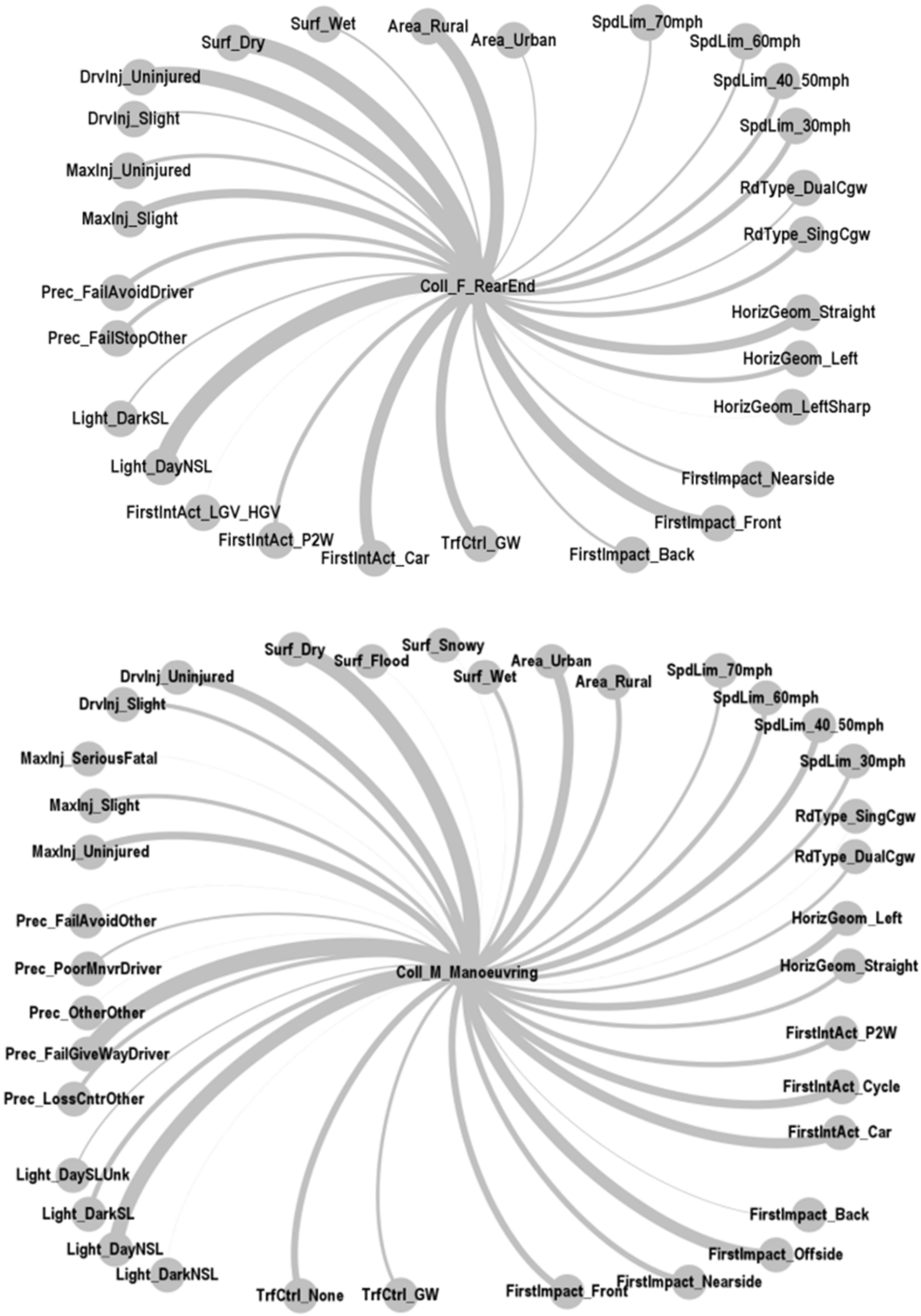


Figure 132: Weighted, directed graphs obtained from all association rules for cluster T-C6

Cluster T-C7: "The car collides with another car, while going straight over a T-junction with a minor road joining from the left."

Table 49: All rules (up to 3 items) obtained for cluster T-C7 with collision type F (scenario T-7.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=F	HorizGeom=Straight	0.289	1.000	1.267
Coll=F & TrfCtrl=None	Area=Rural	0.211	0.800	1.900
Coll=F & Light=DayNSL	Area=Rural	0.158	0.857	2.036
Coll=F & RdType=DualCgw	SpdLim=70mph	0.105	0.800	7.600
Coll=F & SpdLim=30mph	Surf=Dry	0.105	1.000	1.357
Coll=F & Prec=FailAvoidDriver	Area=Rural	0.079	1.000	2.375
Coll=F & Prec=FailAvoidDriver	Light=DayNSL	0.079	1.000	1.810
Coll=F & Area=Urban	Surf=Dry	0.079	1.000	1.357
Coll=F & Light=DarkSL	Surf=Dry	0.053	1.000	1.357
Coll=F & SpdLim=60mph	Surf=Wet	0.053	1.000	3.800
Coll=F & Surf=Wet	SpdLim=60mph	0.053	1.000	5.429
Coll=F & SpdLim=60mph	Light=DayNSL	0.053	1.000	1.810
Coll=F & Light=DaySLUnk	Area=Rural	0.053	1.000	2.375
Coll=F & Light=DaySLUnk	Surf=Dry	0.053	1.000	1.357
Coll=F & Surf=Wet	Area=Rural	0.053	1.000	2.375
Coll=F & Surf=Wet	Light=DayNSL	0.053	1.000	1.810

Table 50: All rules (up to 3 items) obtained for cluster T-C7 with collision type L (scenario T-7.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L	FirstImpact=Front	0.158	0.857	1.357
Coll=L & Surf=Dry	Light=DayNSL	0.105	0.800	1.448
Coll=L & HorizGeom=Straight	Light=DayNSL	0.105	0.800	1.448
Coll=L & Area=Urban	SpdLim=30mph	0.105	1.000	1.652
Coll=L & SpdLim=30mph	Area=Urban	0.105	1.000	1.727
Coll=L & Prec=FailGiveWayOther	HorizGeom=Straight	0.079	1.000	1.267
Coll=L & Area=Rural	Light=DayNSL	0.079	1.000	1.810
Coll=L & Area=Rural	FirstImpact=Front	0.079	1.000	1.583
Coll=L & TrfCtrl=None	Light=DayNSL	0.079	0.750	1.357
Coll=L & TrfCtrl=Light	HorizGeom=Straight	0.053	1.000	1.267
Coll=L & Prec=PoorMnvrOther	FirstImpact=Front	0.053	1.000	1.583
Coll=L & Prec=PoorMnvrOther	Surf=Dry	0.053	1.000	1.357
Coll=L & SpdLim=60mph	Light=DayNSL	0.053	1.000	1.810
Coll=L & SpdLim=60mph	FirstImpact=Front	0.053	1.000	1.583
Coll=L & Light=DaySLUnk	Area=Urban	0.053	1.000	1.727
Coll=L & Light=DaySLUnk	FirstImpact=Front	0.053	1.000	1.583
Coll=L & Surf=Wet	Prec=FailGiveWayOther	0.053	1.000	2.714
Coll=L & Surf=Wet	FirstImpact=Front	0.053	1.000	1.583
Coll=L & Surf=Wet	HorizGeom=Straight	0.053	1.000	1.267

Table 51: All rules (up to 3 items) obtained for cluster T-C7 with collision type H (scenario T-7.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H & FirstIntAct=Car	Surf=Wet	0.105	0.800	3.040
Coll=H & Area=Urban	SpdLim=30mph	0.105	1.000	1.652
Coll=H & SpdLim=30mph	Area=Urban	0.105	1.000	1.727
Coll=H & Light=DaySLUnk	Area=Urban	0.079	1.000	1.727
Coll=H & Area=Urban	Light=DaySLUnk	0.079	0.750	3.167
Coll=H & SpdLim=30mph	Light=DaySLUnk	0.079	0.750	3.167
Coll=H & Light=DaySLUnk	HorizGeom=Straight	0.079	1.000	1.267
Coll=H & HorizGeom=Straight	Light=DaySLUnk	0.079	0.750	3.167
Coll=H & Surf=Wet	Area=Urban	0.079	0.750	1.295
Coll=H & Area=Urban	Surf=Wet	0.079	0.750	2.850
Coll=H & SpdLim=30mph	Surf=Wet	0.079	0.750	2.850
Coll=H & HorizGeom=Straight	Surf=Wet	0.079	0.750	2.850
Coll=H & Prec=FailGiveWayOther	HorizGeom=Straight	0.079	1.000	1.267
Coll=H & HorizGeom=Straight	Prec=FailGiveWayOther	0.079	0.750	2.036
Coll=H & HorizGeom=Straight	Area=Urban	0.079	0.750	1.295
Coll=H & FirstImpact=Front	Light=DaySLUnk	0.053	1.000	4.222
Coll=H & FirstImpact=Front	Prec=FailGiveWayOther	0.053	1.000	2.714
Coll=H & Area=Rural	Light=DayNSL	0.053	1.000	1.810
Coll=H & FirstImpact=Front	Area=Urban	0.053	1.000	1.727
Coll=H & FirstImpact=Front	SpdLim=30mph	0.053	1.000	1.652
Coll=H & FirstImpact=Front	HorizGeom=Straight	0.053	1.000	1.267

Table 52: All rules (up to 3 items) obtained for cluster T-C7 with collision type J (scenario T-7.4), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J	Prec=FailGiveWayOther	0.132	0.833	2.262
Coll=J	Area=Urban	0.132	0.833	1.439
Coll=J	SpdLim=30mph	0.132	0.833	1.377
Coll=J & Prec=FailGiveWayOther	Surf=Dry	0.132	1.000	1.357
Coll=J & Surf=Dry	Prec=FailGiveWayOther	0.132	1.000	2.714
Coll=J & Manvr=GoingAheadOther	Prec=FailGiveWayOther	0.132	1.000	2.714
Coll=J & Manvr=GoingAheadOther	Surf=Dry	0.132	1.000	1.357
Coll=J & Area=Urban	FirstImpact=Front	0.105	0.800	1.267
Coll=J & FirstImpact=Front	Area=Urban	0.105	1.000	1.727
Coll=J & HorizGeom=Straight	Area=Urban	0.105	1.000	1.727
Coll=J & FirstIntAct=Car	FirstImpact=Front	0.105	0.800	1.267
Coll=J & Light=DayNSL	Prec=FailGiveWayOther	0.079	1.000	2.714
Coll=J & Light=DayNSL	SpdLim=30mph	0.079	1.000	1.652
Coll=J & Light=DayNSL	Surf=Dry	0.079	1.000	1.357
Coll=J & HorizGeom=RightSlight	Prec=FailGiveWayOther	0.053	1.000	2.714
Coll=J & HorizGeom=RightSlight	SpdLim=30mph	0.053	1.000	1.652
Coll=J & FirstImpact=Nearside	Prec=FailGiveWayOther	0.053	1.000	2.714
Coll=J & FirstImpact=Nearside	SpdLim=30mph	0.053	1.000	1.652
Coll=J & FirstImpact=Nearside	Surf=Dry	0.053	1.000	1.357

Table 53: All rules (up to 3 items) obtained for cluster T-C7 with collision type M (scenario T-7.5), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll_M_Manoeuvring	HorizGeom_Straight	0.105	1.000	1.267
Coll_M_Manoeuvring,Manvr_GoingAheadOther	FirstImpact_Front	0.079	1.000	1.583
Coll_M_Manoeuvring,Light_DarkNSL	SpdLim_60mph	0.053	1.000	5.429
Coll_M_Manoeuvring,SpdLim_60mph	Light_DarkNSL	0.053	1.000	12.667
Coll_M_Manoeuvring,Prec_FailGiveWayOther	Light_DarkNSL	0.053	1.000	12.667
Coll_M_Manoeuvring,Light_DarkNSL	Area_Rural	0.053	1.000	2.375
Coll_M_Manoeuvring,Area_Rural	Light_DarkNSL	0.053	1.000	12.667
Coll_M_Manoeuvring,SpdLim_60mph	Prec_FailGiveWayOther	0.053	1.000	2.714
Coll_M_Manoeuvring,Prec_FailGiveWayOther	SpdLim_60mph	0.053	1.000	5.429
Coll_M_Manoeuvring,Area_Rural	SpdLim_60mph	0.053	1.000	5.429
Coll_M_Manoeuvring,SpdLim_60mph	Surf_Dry	0.053	1.000	1.357
Coll_M_Manoeuvring,Prec_FailGiveWayOther	Area_Rural	0.053	1.000	2.375
Coll_M_Manoeuvring,Area_Rural	Prec_FailGiveWayOther	0.053	1.000	2.714
Coll_M_Manoeuvring,Prec_FailGiveWayOther	Surf_Dry	0.053	1.000	1.357
Coll_M_Manoeuvring,Area_Rural	Surf_Dry	0.053	1.000	1.357
Coll_M_Manoeuvring,Area_Urban	SpdLim_30mph	0.053	1.000	1.652
Coll_M_Manoeuvring,SpdLim_30mph	Area_Urban	0.053	1.000	1.727

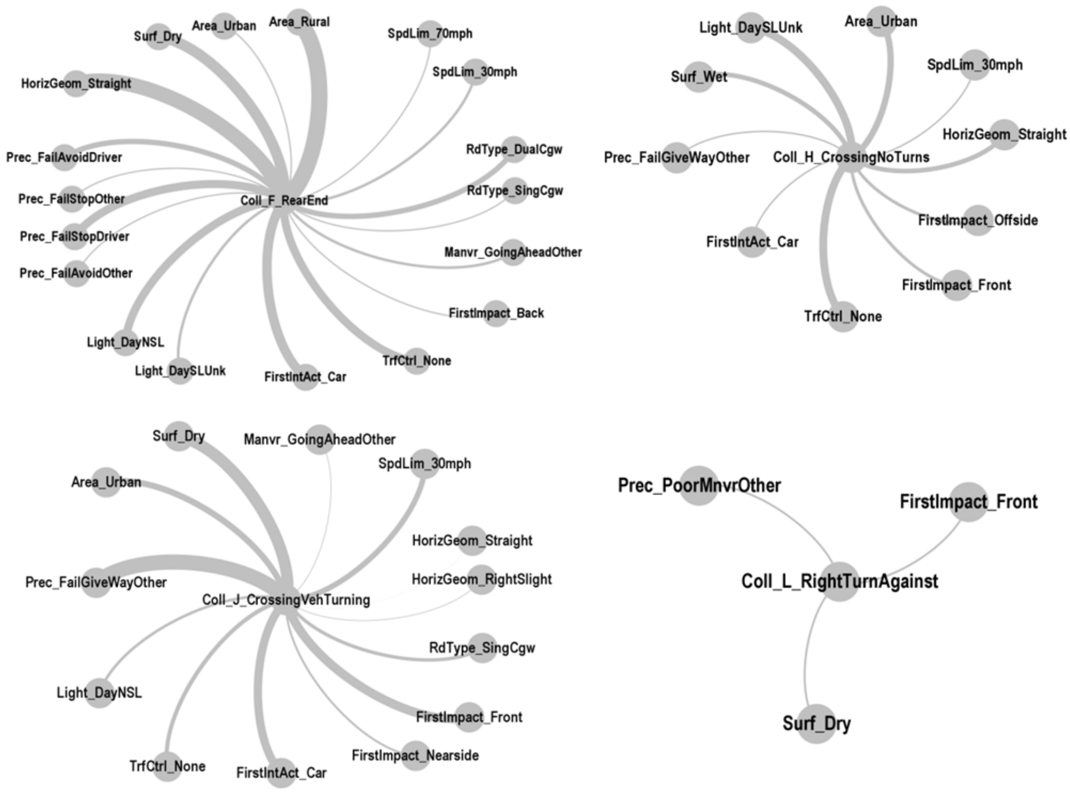


Figure 133: Weighted, directed graphs obtained from all association rules for cluster T-C7

The network graph for collision type M could not be created due to not enough rules obtained.

Cluster T-C8: *"The car is hit by another car or PTW on its offside, while turning right into a major road."*

Table 54: All rules (up to 3 items) obtained for cluster T-C8 with collision type J (scenario T-8.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J & TrfCtrl=GW	Prec=FailGiveWayDriver	0.500	0.900	1.271
Coll=J & SpdLim=30mph	Area=Urban	0.472	0.919	1.408
Coll=J & Area=Urban	SpdLim=30mph	0.472	0.895	1.400
Coll=J & SpdLim=30mph	Prec=FailGiveWayDriver	0.458	0.892	1.259
Coll=J & Surf=Wet	Prec=FailGiveWayDriver	0.278	0.909	1.283
Coll=J & Area=Rural	FirstIntAct=Car	0.181	0.929	1.311
Coll=J & Light=DaySLUnk	Area=Urban	0.167	0.857	1.313
Coll=J & Area=Rural	Light=DayNSL	0.167	0.857	1.714
Coll=J & Light=DarkSL	Area=Urban	0.139	1.000	1.532
Coll=J & FirstIntAct=P2W	Prec=FailGiveWayDriver	0.139	1.000	1.412
Coll=J & FirstIntAct=P2W	TrfCtrl=GW	0.139	1.000	1.286
Coll=J & SpdLim=40-50mph	FirstIntAct=Car	0.125	1.000	1.412
Coll=J & Light=DarkSL	SpdLim=30mph	0.125	0.900	1.409
Coll=J & SpdLim=40-50mph	DrvInj=Slight	0.097	0.778	1.697
Coll=J & SpdLim=40-50mph	Light=DayNSL	0.097	0.778	1.556
Coll=J & Prec=PoorMnvrDriver	FirstIntAct=Car	0.083	1.000	1.412
Coll=J & RdType=DualCgw	FirstIntAct=Car	0.083	1.000	1.412
Coll=J & HorizGeom=Right	SpdLim=30mph	0.042	1.000	1.565
Coll=J & Manvr=WaitTurnR	Surf=Dry	0.042	0.750	1.459

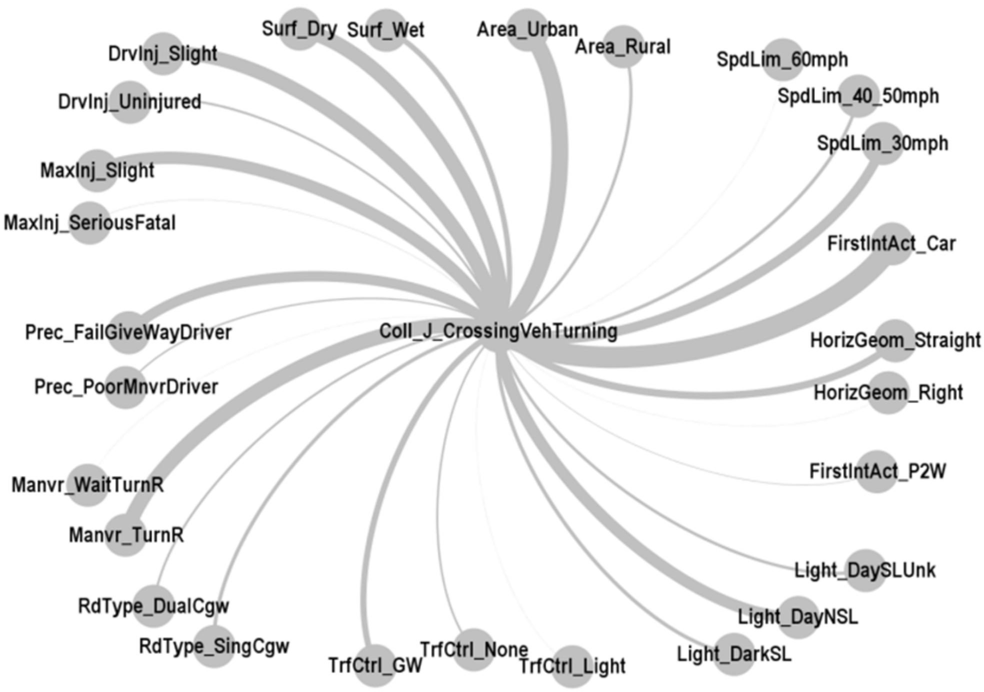


Figure 134: Weighted, directed graph obtained from all association rules for cluster T-C8

Cluster T-C9: "The car is hit by another car on its nearside, while going straight over a T-junction with a minor road joining from the left."

Table 55: All rules (up to 3 items) obtained for cluster T-C9 with collision type J (scenario T-9.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J	Area=Urban	0.450	0.900	1.286
Coll=J	FirstIntAct=Car	0.450	0.900	1.286
Coll=J & Manvr=GoingAheadOther	FirstIntAct=Car	0.425	0.944	1.349
Coll=J & TrfCtrl=None	FirstIntAct=Car	0.400	0.941	1.345
Coll=J	Prec=FailGiveWayOther	0.375	0.750	1.500
Coll=J & Manvr=GoingAheadOther	Prec=FailGiveWayOther	0.375	0.833	1.667
Coll=J & MaxInj=Slight	Prec=FailGiveWayOther	0.375	0.833	1.667
Coll=J & HorizGeom=Straight	Area=Urban	0.375	1.000	1.429
Coll=J & Prec=FailGiveWayOther	FirstIntAct=Car	0.350	0.933	1.333
Coll=J & FirstIntAct=Car	Prec=FailGiveWayOther	0.350	0.778	1.556
Coll=J & TrfCtrl=None	Prec=FailGiveWayOther	0.350	0.824	1.647
Coll=J & MaxInj=Slight	Surf=Dry	0.350	0.778	1.296
Coll=J & SpdLim=30mph	Area=Urban	0.350	1.000	1.429
Coll=J & Surf=Dry	Area=Urban	0.325	0.929	1.327
Coll=J & TrfCtrl=None	Surf=Dry	0.325	0.765	1.275
Coll=J & DrvInj=Slight	Prec=FailGiveWayOther	0.300	0.923	1.846
Coll=J & Prec=FailGiveWayOther	SpdLim=30mph	0.300	0.800	1.280
Coll=J & SpdLim=30mph	Prec=FailGiveWayOther	0.300	0.857	1.714
Coll=J & DrvInj=Slight	FirstIntAct=Car	0.300	0.923	1.319
Coll=J & SpdLim=30mph	Surf=Dry	0.300	0.857	1.429
Coll=J & Surf=Dry	Prec=FailGiveWayOther	0.275	0.786	1.571
Coll=J & Light=DayNSL	Surf=Dry	0.275	0.786	1.310
Coll=J & DrvInj=Slight	Surf=Dry	0.250	0.769	1.282
Coll=J & Surf=Wet	FirstIntAct=Car	0.150	1.000	1.429
Coll=J & DrvInj=Uninjured	SpdLim=30mph	0.125	1.000	1.600
Coll=J & DrvInj=Uninjured	Area=Urban	0.125	1.000	1.429
Coll=J & DrvInj=Uninjured	HorizGeom=Straight	0.125	1.000	1.379
Coll=J & SpdLim=40-50mph	Surf=Wet	0.100	0.800	2.133
Coll=J & DrvInj=Uninjured	Surf=Dry	0.100	0.800	1.333
Coll=J & Light=DarkSL	Area=Urban	0.075	1.000	1.429
Coll=J & Light=DarkSL	FirstIntAct=Car	0.075	1.000	1.429
Coll=J & Light=DaySLUnk	Prec=FailGiveWayOther	0.075	1.000	2.000
Coll=J & Light=DaySLUnk	SpdLim=30mph	0.075	1.000	1.600
Coll=J & Light=DaySLUnk	Area=Urban	0.075	1.000	1.429
Coll=J & Light=DaySLUnk	FirstIntAct=Car	0.075	1.000	1.429
Coll=J & Light=DaySLUnk	HorizGeom=Straight	0.075	1.000	1.379
Coll=J & Area=Rural	HorizGeom=Right	0.050	1.000	13.333
Coll=J & MaxInj=SeriousFatal	SpdLim=40-50mph	0.050	1.000	5.714
Coll=J & MaxInj=SeriousFatal	FirstIntAct=Car	0.050	1.000	1.429

Coll=J & DrvInj=Serious	SpdLim=40-50mph	0.050	1.000	5.714
Coll=J & DrvInj=Serious	FirstIntAct=Car	0.050	1.000	1.429
Coll=J & TrfCtrl=Light	Surf=Wet	0.050	1.000	2.667
Coll=J & TrfCtrl=Light	FirstIntAct=Car	0.050	1.000	1.429
Coll=J & HorizGeom=RightSlight	Prec=FailGiveWayOther	0.050	1.000	2.000
Coll=J & HorizGeom=RightSlight	DrvInj=Slight	0.050	1.000	1.667
Coll=J & HorizGeom=RightSlight	SpdLim=30mph	0.050	1.000	1.600
Coll=J & HorizGeom=RightSlight	Area=Urban	0.050	1.000	1.429
Coll=J & HorizGeom=RightSlight	FirstIntAct=Car	0.050	1.000	1.429
Coll=J & Area=Rural	SpdLim=40-50mph	0.050	1.000	5.714
Coll=J & Area=Rural	Prec=FailGiveWayOther	0.050	1.000	2.000
Coll=J & Area=Rural	DrvInj=Slight	0.050	1.000	1.667
Coll=J & Area=Rural	Light=DayNSL	0.050	1.000	1.538
Coll=J & Area=Rural	FirstIntAct=Car	0.050	1.000	1.429

Table 56: All rules (up to 3 items) obtained for cluster T-C9 with collision type M (scenario T-9.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll_M_Manoeuvring	FirstIntAct_Car	0.100	1.000	1.429
Coll_M_Manoeuvring	HorizGeom_Straight	0.100	1.000	1.379
Coll_M_Manoeuvring	DrvInj_Slight	0.075	0.750	1.250
Coll_M_Manoeuvring,Prec_FailGiveWayOther	DrvInj_Slight	0.050	1.000	1.667
Coll_M_Manoeuvring,SpdLim_30mph	DrvInj_Slight	0.050	1.000	1.667
Coll_M_Manoeuvring,Area_Urban	DrvInj_Slight	0.050	1.000	1.667
Coll_M_Manoeuvring,Surf_Dry	Light_DayNSL	0.050	1.000	1.538
Coll_M_Manoeuvring,Light_DayNSL	Surf_Dry	0.050	1.000	1.667
Coll_M_Manoeuvring,SpdLim_30mph	Area_Urban	0.050	1.000	1.429
Coll_M_Manoeuvring,Area_Urban	SpdLim_30mph	0.050	1.000	1.600

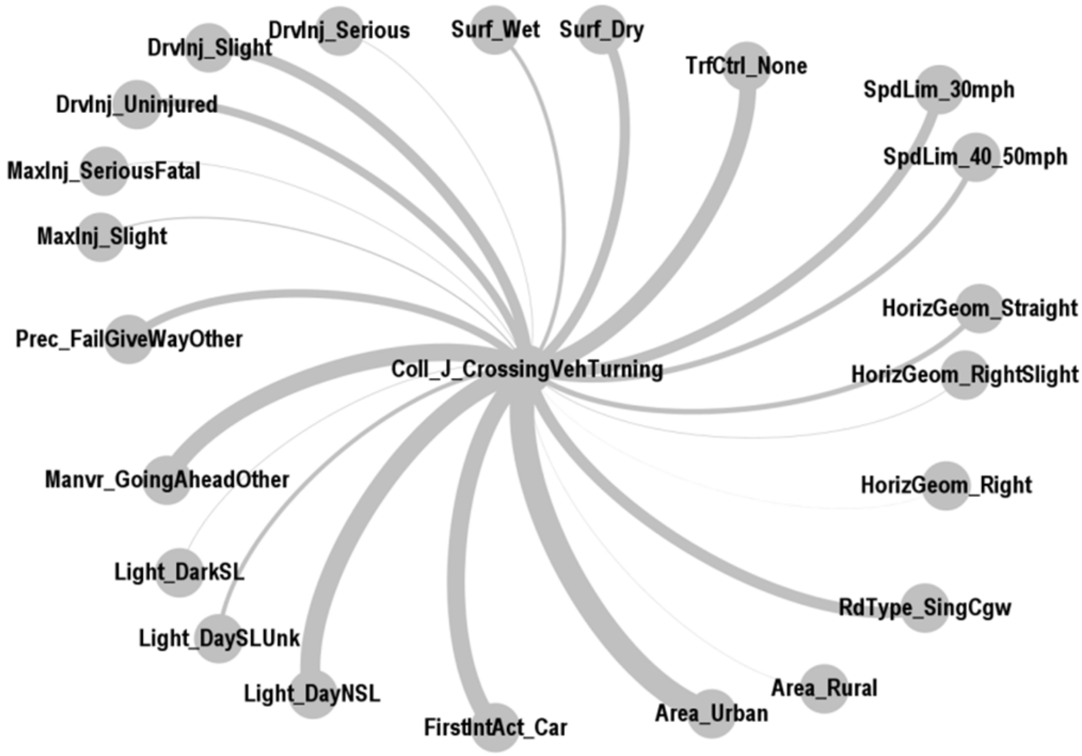


Figure 135: Weighted, directed graph obtained from all association rules for cluster T-C9

The network graph for collision type M could not be created due to not enough rules obtained.

Cluster T-C10: "The car hits another car with its front, while going straight over a T-junction with a minor road joining from the left."

Table 57: Rules (up to 3 items) obtained for cluster T-C10 with collision type L (scenario T-10.1), sorted by the five highest support values

Antecedent	Consequent	Supp	Conf	Lift
Coll=L	RdType=SingCgw	0.237	0.818	1.413
Coll=L & RdType=SingCgw	TrfCtrl=None	0.237	1.000	1.357
Coll=L & TrfCtrl=None	RdType=SingCgw	0.237	1.000	1.727
Coll=L & Manvr=GoingAheadOther	RdType=SingCgw	0.237	0.900	1.555
Coll=L & Surf=Dry	FirstIntAct=Car	0.184	1.000	1.357
Coll=L & Surf=Dry	RdType=SingCgw	0.158	0.857	1.481
Coll=L & Area=Rural	RdType=SingCgw	0.158	0.857	1.481
Coll=L & SpdLim=40-50mph	FirstIntAct=Car	0.132	1.000	1.357
Coll=L & DrvInj=Uninjured	RdType=SingCgw	0.132	1.000	1.727
Coll=L & DrvInj=Uninjured	FirstIntAct=Car	0.132	1.000	1.357
Coll=L & DrvInj=Uninjured	TrfCtrl=None	0.132	1.000	1.357
Coll=L & Prec=FailGiveWayOther	FirstIntAct=Car	0.132	1.000	1.357
Coll=L & Light=DayNSL	FirstIntAct=Car	0.132	1.000	1.357
Coll=L & Light=DaySLUnk	RdType=SingCgw	0.105	1.000	1.727
Coll=L & Light=DaySLUnk	TrfCtrl=None	0.105	1.000	1.357
Coll=L & Light=DaySLUnk	HorizGeom=Straight	0.105	1.000	1.267
Coll=L & Area=Urban	FirstIntAct=Car	0.105	1.000	1.357
Coll=L & Area=Urban	HorizGeom=Straight	0.105	1.000	1.267
Coll=L & SpdLim=40-50mph	Surf=Dry	0.105	0.800	1.448
Coll=L & DrvInj=Uninjured	Surf=Dry	0.105	0.800	1.448
Coll=L & Prec=FailGiveWayOther	Surf=Dry	0.105	0.800	1.448
Coll=L & Surf=Wet	HorizGeom=Straight	0.105	1.000	1.267
Coll=L & Light=DayNSL	Surf=Dry	0.105	0.800	1.448

Table 58: Rules (up to 3 items) obtained for cluster T-C10 with collision type J (scenario T-10.2), sorted by the five highest support values

Antecedent	Consequent	Supp	Conf	Lift
Coll=J	Prec=FailGiveWayOther	0.211	0.800	2.338
Coll=J	Light=DayNSL	0.211	0.800	1.520
Coll=J & Light=DayNSL	HorizGeom=Straight	0.211	1.000	1.267
Coll=J & HorizGeom=Straight	Light=DayNSL	0.211	1.000	1.900
Coll=J & FirstIntAct=Car	Prec=FailGiveWayOther	0.184	0.875	2.558
Coll=J & TrfCtrl=None	Prec=FailGiveWayOther	0.184	0.875	2.558
Coll=J & Area=Rural	Light=DayNSL	0.184	1.000	1.900
Coll=J & Area=Rural	HorizGeom=Straight	0.184	1.000	1.267
Coll=J & Prec=FailGiveWayOther	Surf=Wet	0.158	0.750	1.781
Coll=J & Surf=Wet	Prec=FailGiveWayOther	0.158	1.000	2.923
Coll=J & RdType=SingCgw	Prec=FailGiveWayOther	0.158	0.857	2.505
Coll=J & Area=Rural	Prec=FailGiveWayOther	0.158	0.857	2.505

Coll=J & Surf=Wet	Light=DayNSL	0.158	1.000	1.900
Coll=J & Light=DayNSL	Surf=Wet	0.158	0.750	1.781
Coll=J & Surf=Wet	Area=Rural	0.158	1.000	1.407
Coll=J & Area=Rural	Surf=Wet	0.158	0.857	2.036
Coll=J & Surf=Wet	HorizGeom=Straight	0.158	1.000	1.267
Coll=J & HorizGeom=Straight	Surf=Wet	0.158	0.750	1.781
Coll=J & TrfCtrl=None	RdType=SingCgw	0.158	0.750	1.295
Coll=J & DrvInj=Serious	Prec=FailGiveWayOther	0.105	1.000	2.923
Coll=J & SpdLim=30mph	Area=Urban	0.079	1.000	3.455
Coll=J & Area=Urban	SpdLim=30mph	0.079	1.000	3.800
Coll=J & SpdLim=30mph	Surf=Dry	0.079	1.000	1.810
Coll=J & Surf=Dry	SpdLim=30mph	0.079	0.750	2.850
Coll=J & SpdLim=30mph	RdType=SingCgw	0.079	1.000	1.727
Coll=J & SpdLim=30mph	TrfCtrl=None	0.079	1.000	1.357
Coll=J & DrvInj=Serious	SpdLim=40-50mph	0.079	0.750	2.375
Coll=J & SpdLim=40-50mph	DrvInj=Serious	0.079	1.000	3.455
Coll=J & DrvInj=Serious	Surf=Wet	0.079	0.750	1.781
Coll=J & DrvInj=Serious	RdType=SingCgw	0.079	0.750	1.295
Coll=J & Area=Urban	Surf=Dry	0.079	1.000	1.810
Coll=J & Surf=Dry	Area=Urban	0.079	0.750	2.591
Coll=J & Area=Urban	RdType=SingCgw	0.079	1.000	1.727
Coll=J & Area=Urban	TrfCtrl=None	0.079	1.000	1.357
Coll=J & SpdLim=40-50mph	Prec=FailGiveWayOther	0.079	1.000	2.923
Coll=J & SpdLim=40-50mph	Surf=Wet	0.079	1.000	2.375
Coll=J & SpdLim=40-50mph	Light=DayNSL	0.079	1.000	1.900
Coll=J & SpdLim=40-50mph	Area=Rural	0.079	1.000	1.407
Coll=J & SpdLim=40-50mph	HorizGeom=Straight	0.079	1.000	1.267
Coll=J & DrvInj=Uninjured	Light=DayNSL	0.079	1.000	1.900
Coll=J & DrvInj=Uninjured	HorizGeom=Straight	0.079	1.000	1.267
Coll=J & RdType=DualCgw	Light=DayNSL	0.079	1.000	1.900
Coll=J & RdType=DualCgw	Area=Rural	0.079	1.000	1.407
Coll=J & RdType=DualCgw	FirstIntAct=Car	0.079	1.000	1.357
Coll=J & RdType=DualCgw	HorizGeom=Straight	0.079	1.000	1.267
Coll=J & Surf=Dry	RdType=SingCgw	0.079	0.750	1.295

Table 59: Rules (up to 3 items) obtained for cluster T-C10 with collision type F (scenario T-10.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll_F_RearEnd	Surf_Dry	0.132	1.000	1.810
Coll_F_RearEnd	Light_DaySLUnk	0.105	0.800	3.040
Coll_F_RearEnd	DrvInj_Slight	0.105	0.800	2.533
Coll_F_RearEnd,Area_Rural	FirstIntAct_Car	0.105	1.000	1.357
Coll_F_RearEnd,FirstIntAct_Car	Area_Rural	0.105	1.000	1.407
Coll_F_RearEnd,Area_Rural	HorizGeom_Straight	0.105	1.000	1.267
Coll_F_RearEnd,HorizGeom_Straight	Area_Rural	0.105	1.000	1.407
Coll_F_RearEnd,Manvr_GoingAheadOther	Area_Rural	0.105	1.000	1.407

Coll_F_RearEnd,FirstIntAct_Car	HorizGeom_Straight	0.105	1.000	1.267
Coll_F_RearEnd,HorizGeom_Straight	FirstIntAct_Car	0.105	1.000	1.357
Coll_F_RearEnd,Manvr_GoingAheadOther	FirstIntAct_Car	0.105	1.000	1.357
Coll_F_RearEnd,Manvr_GoingAheadOther	HorizGeom_Straight	0.105	1.000	1.267
Coll_F_RearEnd,RdType_SingCgw	Light_DaySLUnk	0.079	1.000	3.800
Coll_F_RearEnd,RdType_SingCgw	Area_Rural	0.079	1.000	1.407
Coll_F_RearEnd,Area_Rural	RdType_SingCgw	0.079	0.750	1.295
Coll_F_RearEnd,RdType_SingCgw	FirstIntAct_Car	0.079	1.000	1.357
Coll_F_RearEnd,FirstIntAct_Car	RdType_SingCgw	0.079	0.750	1.295
Coll_F_RearEnd,RdType_SingCgw	HorizGeom_Straight	0.079	1.000	1.267
Coll_F_RearEnd,HorizGeom_Straight	RdType_SingCgw	0.079	0.750	1.295
Coll_F_RearEnd,Manvr_GoingAheadOther	RdType_SingCgw	0.079	0.750	1.295
Coll_F_RearEnd,Prec_FailAvoidDriver	DrvInj_Slight	0.053	1.000	3.167
Coll_F_RearEnd,Prec_FailAvoidDriver	RdType_DualCgw	0.053	1.000	2.533
Coll_F_RearEnd,RdType_DualCgw	Prec_FailAvoidDriver	0.053	1.000	12.667
Coll_F_RearEnd,TrfCtrl_Light	Light_DaySLUnk	0.053	1.000	3.800
Coll_F_RearEnd,TrfCtrl_Light	SpdLim_30mph	0.053	1.000	3.800
Coll_F_RearEnd,SpdLim_30mph	TrfCtrl_Light	0.053	1.000	6.333
Coll_F_RearEnd,TrfCtrl_Light	RdType_SingCgw	0.053	1.000	1.727
Coll_F_RearEnd,TrfCtrl_Light	Area_Rural	0.053	1.000	1.407
Coll_F_RearEnd,SpdLim_30mph	Light_DaySLUnk	0.053	1.000	3.800
Coll_F_RearEnd,SpdLim_30mph	RdType_SingCgw	0.053	1.000	1.727
Coll_F_RearEnd,SpdLim_30mph	Area_Rural	0.053	1.000	1.407
Coll_F_RearEnd,SpdLim_30mph	FirstIntAct_Car	0.053	1.000	1.357
Coll_F_RearEnd,SpdLim_30mph	HorizGeom_Straight	0.053	1.000	1.267
Coll_F_RearEnd,RdType_DualCgw	DrvInj_Slight	0.053	1.000	3.167
Coll_F_RearEnd,TrfCtrl_None	DrvInj_Slight	0.053	1.000	3.167
Coll_F_RearEnd,TrfCtrl_None	Area_Rural	0.053	1.000	1.407
Coll_F_RearEnd,TrfCtrl_None	FirstIntAct_Car	0.053	1.000	1.357
Coll_F_RearEnd,TrfCtrl_None	HorizGeom_Straight	0.053	1.000	1.267

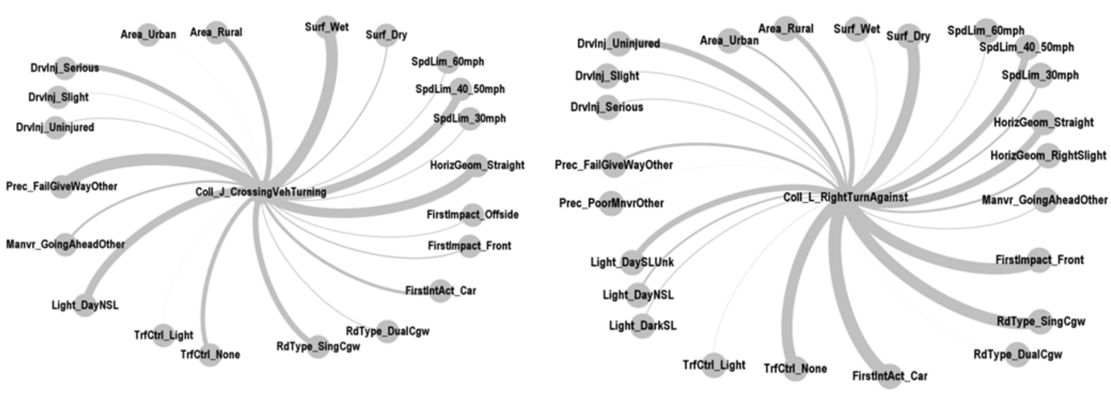


Figure 136: Weighted, directed graphs obtained from all association rules for cluster T-C10

The network graph for collision type F could not be created due to not enough rules obtained.

Cluster T-C11: "The car is hit by another car on its offside, while going straight over a T-junction with a minor road joining from the right."

Table 60: All rules (up to 3 items) obtained for cluster T-C11 with collision type M (scenario T-11.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=M & Manvr=GoingAheadOther	Prec=FailGiveWayOther	0.188	0.857	2.743
Coll=M & FirstIntAct=Car	Surf=Wet	0.188	0.750	2.000
Coll=M & SpdLim=30mph	Area=Urban	0.188	0.857	1.371
Coll=M & Prec=PoorMnvrDriver	Manvr=Other	0.156	1.000	3.556
Coll=M & Light=DarkSL	FirstIntAct=Car	0.094	1.000	1.524
Coll=M & Light=DarkSL	HorizGeom=Straight	0.094	1.000	1.524
Coll=M & SpdLim=40-50mph	HorizGeom=Straight	0.094	1.000	1.524
Coll=M & HorizGeom=RightSlight	Manvr=Other	0.063	1.000	3.556
Coll=M & HorizGeom=RightSlight	Light=DayNSL	0.063	1.000	2.133
Coll=M & HorizGeom=RightSlight	DrvInj=Uninjured	0.063	1.000	1.882
Coll=M & HorizGeom=RightSlight	SpdLim=30mph	0.063	1.000	1.684
Coll=M & HorizGeom=LeftSlight	Prec=FailGiveWayOther	0.063	1.000	3.200
Coll=M & HorizGeom=LeftSlight	DrvInj=Slight	0.063	1.000	2.286
Coll=M & HorizGeom=LeftSlight	Light=DayNSL	0.063	1.000	2.133
Coll=M & HorizGeom=LeftSlight	SpdLim=30mph	0.063	1.000	1.684
Coll=M & HorizGeom=LeftSlight	Area=Urban	0.063	1.000	1.600
Coll=M & Light=DaySLUnk	Surf=Wet	0.063	1.000	2.667
Coll=M & Light=DaySLUnk	HorizGeom=Straight	0.063	1.000	1.524
Coll=M & MaxInj=SeriousFatal	FirstIntAct=Cycle	0.031	1.000	32.000
Coll=M & Prec=OtherDriver	Surf=Flood	0.031	1.000	32.000
Coll=M & MaxInj=Uninjured	Light=DarkSLUnk	0.031	1.000	32.000
Coll=M & Prec=FailStopDriver	Light=DarkSLUnk	0.031	1.000	32.000
Coll=M & MaxInj=Uninjured	SpdLim=70mph	0.031	1.000	32.000
Coll=M & Prec=FailStopDriver	SpdLim=70mph	0.031	1.000	32.000
Coll=M & Prec=OtherDriver	FirstIntAct=LGV=HGV	0.031	1.000	8.000
Coll=M & Prec=OtherDriver	Manvr=Other	0.031	1.000	3.556
Coll=M & Prec=OtherDriver	DrvInj=Slight	0.031	1.000	2.286
Coll=M & Prec=OtherDriver	HorizGeom=Straight	0.031	1.000	1.524
Coll=M & MaxInj=Uninjured	Prec=FailStopDriver	0.031	1.000	16.000
Coll=M & Prec=FailStopDriver	MaxInj=Uninjured	0.031	1.000	16.000
Coll=M & MaxInj=Uninjured	TrfCtrl=GW	0.031	1.000	10.667
Coll=M & MaxInj=Uninjured	RdType=DualCgw	0.031	1.000	10.667
Coll=M & MaxInj=Uninjured	FirstIntAct=LGV=HGV	0.031	1.000	8.000
Coll=M & Prec=FailStopDriver	TrfCtrl=GW	0.031	1.000	10.667
Coll=M & Prec=FailStopDriver	RdType=DualCgw	0.031	1.000	10.667
Coll=M & Prec=FailStopDriver	FirstIntAct=LGV=HGV	0.031	1.000	8.000
Coll=M & Prec=FailStopDriver	DrvInj=Uninjured	0.031	1.000	1.882
Coll=M & Prec=FailStopDriver	Area=Urban	0.031	1.000	1.600
Coll=M & Light=DarkNSL	RdType=DualCgw	0.031	1.000	10.667

Coll=M & Light=DarkNSL	Prec=FailGiveWayOther	0.031	1.000	3.200
Coll=M & Light=DarkNSL	Surf=Wet	0.031	1.000	2.667
Coll=M & Light=DarkNSL	DrvInj=Uninjured	0.031	1.000	1.882
Coll=M & FirstIntAct=P2W	SpdLim=60mph	0.031	1.000	8.000
Coll=M & FirstIntAct=P2W	Prec=PoorMnvrDriver	0.031	1.000	4.000
Coll=M & FirstIntAct=P2W	Manvr=Other	0.031	1.000	3.556
Coll=M & FirstIntAct=P2W	Area=Rural	0.031	1.000	2.667
Coll=M & FirstIntAct=P2W	Light=DayNSL	0.031	1.000	2.133
Coll=M & FirstIntAct=P2W	DrvInj=Uninjured	0.031	1.000	1.882
Coll=M & MaxInj=SeriousFatal	SpdLim=40-50mph	0.031	1.000	4.000
Coll=M & MaxInj=SeriousFatal	Prec=FailGiveWayOther	0.031	1.000	3.200
Coll=M & MaxInj=SeriousFatal	Area=Rural	0.031	1.000	2.667
Coll=M & MaxInj=SeriousFatal	Light=DayNSL	0.031	1.000	2.133
Coll=M & MaxInj=SeriousFatal	DrvInj=Uninjured	0.031	1.000	1.882
Coll=M & MaxInj=SeriousFatal	Surf=Dry	0.031	1.000	1.684
Coll=M & MaxInj=SeriousFatal	HorizGeom=Straight	0.031	1.000	1.524

Table 61: All rules (up to 3 items) obtained for cluster T-C11 with collision type F (scenario T-11.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll_F=RearEnd	HorizGeom=Straight	0.156	1.000	1.524
Coll_F=RearEnd	SpdLim=30mph	0.125	0.800	1.347
Coll_F=RearEnd	Surf=Dry	0.125	0.800	1.347
Coll_F=RearEnd & Light=DaySLUnk	SpdLim=30mph	0.094	1.000	1.684
Coll_F=RearEnd & SpdLim=30mph	Light=DaySLUnk	0.094	0.750	4.000
Coll_F=RearEnd & Light=DaySLUnk	Manvr=GoingAheadOther	0.094	1.000	1.391
Coll_F=RearEnd & Manvr=GoingAheadOther	Light=DaySLUnk	0.094	0.750	4.000
Coll_F=RearEnd & Area=Rural	Surf=Dry	0.094	1.000	1.684
Coll_F=RearEnd & Surf=Dry	Area=Rural	0.094	0.750	2.000
Coll_F=RearEnd & Area=Rural	FirstIntAct=Car	0.094	1.000	1.524
Coll_F=RearEnd & FirstIntAct=Car	Area=Rural	0.094	0.750	2.000
Coll_F=RearEnd & DrvInj=Slight	Surf=Dry	0.094	1.000	1.684
Coll_F=RearEnd & Surf=Dry	DrvInj=Slight	0.094	0.750	1.714
Coll_F=RearEnd & DrvInj=Slight	Manvr=GoingAheadOther	0.094	1.000	1.391
Coll_F=RearEnd & Manvr=GoingAheadOther	DrvInj=Slight	0.094	0.750	1.714
Coll_F=RearEnd & Area=Urban	Light=DaySLUnk	0.063	1.000	5.333
Coll_F=RearEnd & DrvInj=Uninjured	SpdLim=30mph	0.063	1.000	1.684
Coll_F=RearEnd & DrvInj=Uninjured	FirstIntAct=Car	0.063	1.000	1.524
Coll_F=RearEnd & Area=Urban	SpdLim=30mph	0.063	1.000	1.684
Coll_F=RearEnd & Area=Urban	Manvr=GoingAheadOther	0.063	1.000	1.391
Coll_F=RearEnd & Surf=Wet	Prec=FailAvoidDriver	0.031	1.000	16.000
Coll_F=RearEnd & FirstIntAct=P2W	Prec=FailStopOther	0.031	1.000	16.000
Coll_F=RearEnd & Manvr=Other	Prec=FailStopOther	0.031	1.000	16.000
Coll_F=RearEnd & Light=DayNSL	Prec=FailStopOther	0.031	1.000	16.000
Coll_F=RearEnd & Prec=FailStopDriver	Light=DarkNSL	0.031	1.000	16.000

Coll_F=RearEnd	& Light=DarkNSL	Prec=FailStopDriver	0.031	1.000	16.000
Coll_F=RearEnd	& Prec=FailStopDriver	SpdLim=60mph	0.031	1.000	8.000
Coll_F=RearEnd	& SpdLim=60mph	Prec=FailStopDriver	0.031	1.000	16.000
Coll_F=RearEnd	& Prec=FailStopDriver	Area=Rural	0.031	1.000	2.667
Coll_F=RearEnd	& Prec=FailStopDriver	DrvInj=Slight	0.031	1.000	2.286
Coll_F=RearEnd	& Prec=FailStopDriver	FirstIntAct=Car	0.031	1.000	1.524
Coll_F=RearEnd	& SpdLim=60mph	Light=DarkNSL	0.031	1.000	16.000
Coll_F=RearEnd	& Light=DarkNSL	DrvInj=Slight	0.031	1.000	2.286
Coll_F=RearEnd	& Light=DarkNSL	Surf=Dry	0.031	1.000	1.684
Coll_F=RearEnd	& FirstIntAct=P2W	Light=DaySLUnk	0.031	1.000	5.333
Coll_F=RearEnd	& FirstIntAct=P2W	DrvInj=Slight	0.031	1.000	2.286
Coll_F=RearEnd	& FirstIntAct=P2W	SpdLim=30mph	0.031	1.000	1.684
Coll_F=RearEnd	& FirstIntAct=P2W	Area=Urban	0.031	1.000	1.600
Coll_F=RearEnd	& FirstIntAct=P2W	Manvr=GoingAheadOther	0.031	1.000	1.391
Coll_F=RearEnd	& SpdLim=60mph	DrvInj=Slight	0.031	1.000	2.286
Coll_F=RearEnd	& SpdLim=60mph	Surf=Dry	0.031	1.000	1.684
Coll_F=RearEnd	& SpdLim=60mph	FirstIntAct=Car	0.031	1.000	1.524
Coll_F=RearEnd	& SpdLim=60mph	Manvr=GoingAheadOther	0.031	1.000	1.391
Coll_F=RearEnd	& Surf=Wet	Light=DaySLUnk	0.031	1.000	5.333
Coll_F=RearEnd	& Manvr=Other	Area=Rural	0.031	1.000	2.667
Coll_F=RearEnd	& Manvr=Other	Light=DayNSL	0.031	1.000	2.133
Coll_F=RearEnd	& Light=DayNSL	Manvr=Other	0.031	1.000	3.556
Coll_F=RearEnd	& Manvr=Other	DrvInj=Uninjured	0.031	1.000	1.882
Coll_F=RearEnd	& Manvr=Other	SpdLim=30mph	0.031	1.000	1.684
Coll_F=RearEnd	& Manvr=Other	Surf=Dry	0.031	1.000	1.684
Coll_F=RearEnd	& Manvr=Other	FirstIntAct=Car	0.031	1.000	1.524
Coll_F=RearEnd	& Light=DayNSL	Area=Rural	0.031	1.000	2.667
Coll_F=RearEnd	& Surf=Wet	DrvInj=Uninjured	0.031	1.000	1.882
Coll_F=RearEnd	& Surf=Wet	SpdLim=30mph	0.031	1.000	1.684
Coll_F=RearEnd	& Surf=Wet	Area=Urban	0.031	1.000	1.600
Coll_F=RearEnd	& Surf=Wet	FirstIntAct=Car	0.031	1.000	1.524
Coll_F=RearEnd	& Surf=Wet	Manvr=GoingAheadOther	0.031	1.000	1.391
Coll_F=RearEnd	& Light=DayNSL	DrvInj=Uninjured	0.031	1.000	1.882
Coll_F=RearEnd	& Light=DayNSL	SpdLim=30mph	0.031	1.000	1.684
Coll_F=RearEnd	& Light=DayNSL	Surf=Dry	0.031	1.000	1.684
Coll_F=RearEnd	& Light=DayNSL	FirstIntAct=Car	0.031	1.000	1.524

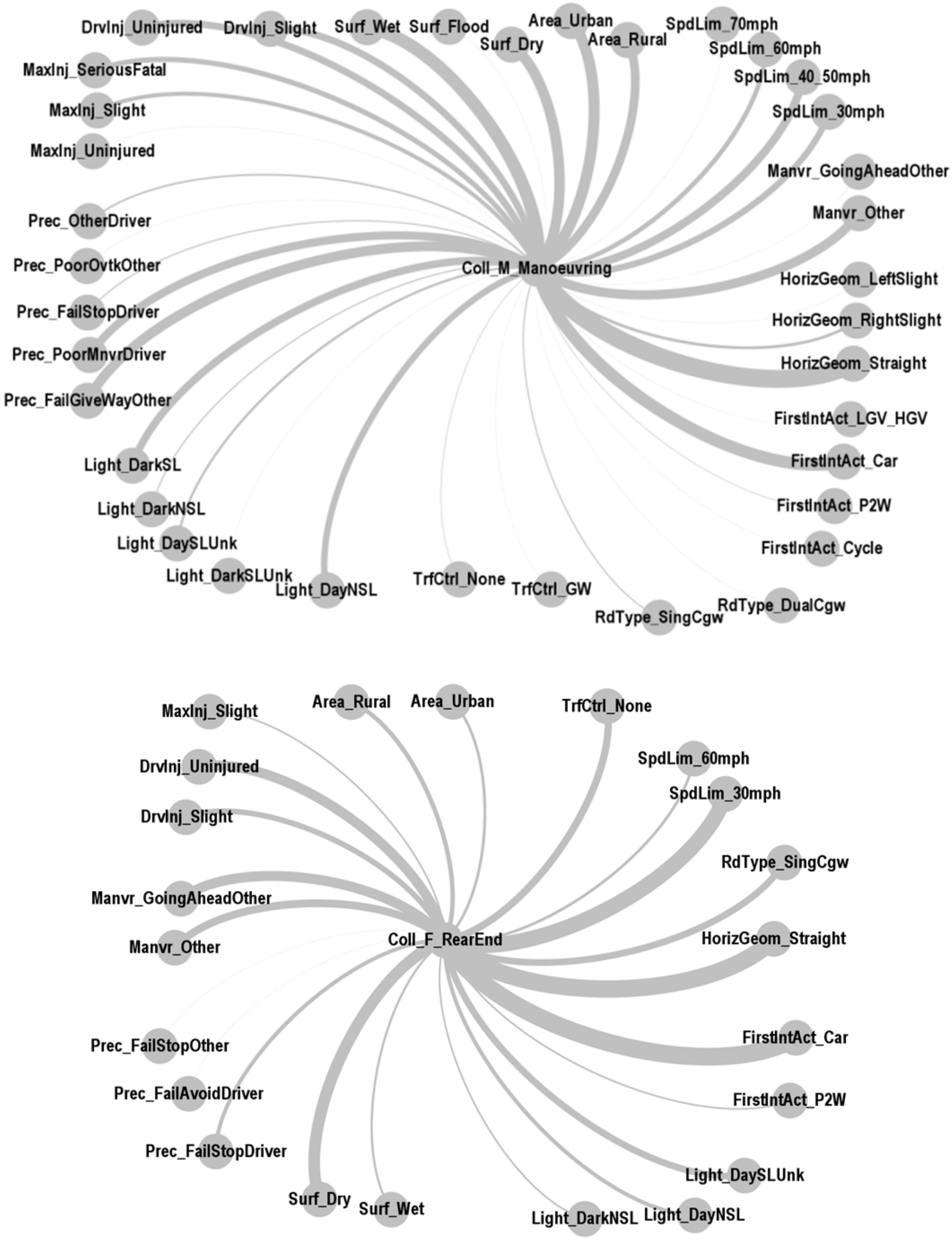


Figure 137: Weighted, directed graphs obtained from all association rules for cluster T-C11

Cluster T-C12: "The car collides with a PTW, while turning right into minor or major road."

Table 62: All rules (up to 3 items) obtained for cluster T-C12 with collision type J (scenario T-12.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J	Prec=FailGiveWayDriver	0.382	0.813	1.625
Coll=J & HorizGeom=Straight	Prec=FailGiveWayDriver	0.353	0.857	1.714
Coll=J & TrfCtrl=GW	Prec=FailGiveWayDriver	0.294	0.909	1.818
Coll=J & Prec=FailGiveWayDriver	TrfCtrl=GW	0.294	0.769	1.538
Coll=J & FirstImpact=Offside	Prec=FailGiveWayDriver	0.294	0.909	1.818
Coll=J & DrvInj=Uninjured	Prec=FailGiveWayDriver	0.294	0.833	1.667
Coll=J & FirstImpact=Offside	TrfCtrl=GW	0.265	0.818	1.636
Coll=J & Light=DayNSL	Prec=FailGiveWayDriver	0.235	0.889	1.778
Coll=J & Area=Rural	Prec=FailGiveWayDriver	0.206	1.000	2.000
Coll=J & Area=Rural	FirstImpact=Offside	0.206	1.000	1.417
Coll=J & Light=DayNSL	TrfCtrl=GW	0.206	0.778	1.556
Coll=J & SpdLim=30mph	Prec=FailGiveWayDriver	0.206	0.875	1.750
Coll=J & Area=Rural	TrfCtrl=GW	0.176	0.857	1.714
Coll=J & SpdLim=30mph	TrfCtrl=GW	0.176	0.750	1.500
Coll=J & FirstImpact=Front	Area=Urban	0.147	1.000	1.417
Coll=J & FirstImpact=Front	DrvInj=Uninjured	0.147	1.000	1.259
Coll=J & TrfCtrl=None	DrvInj=Uninjured	0.147	1.000	1.259
Coll=J & SpdLim=60mph	Prec=FailGiveWayDriver	0.088	1.000	2.000
Coll=J & SpdLim=60mph	Light=DayNSL	0.088	1.000	1.700
Coll=J & SpdLim=60mph	FirstImpact=Offside	0.088	1.000	1.417
Coll=J & Light=DaySLUnk	TrfCtrl=GW	0.088	0.750	1.500
Coll=J & Light=DarkSL	Prec=FailGiveWayDriver	0.059	1.000	2.000
Coll=J & Light=DarkSL	FirstImpact=Offside	0.059	1.000	1.417
Coll=J & Prec=PoorOvtkOther	SpdLim=40-50mph	0.059	1.000	4.250
Coll=J & Prec=PoorOvtkOther	FirstImpact=Front	0.059	1.000	3.778
Coll=J & Prec=PoorOvtkOther	TrfCtrl=None	0.059	1.000	2.125
Coll=J & Prec=PoorOvtkOther	Area=Urban	0.059	1.000	1.417
Coll=J & Prec=PoorOvtkOther	DrvInj=Uninjured	0.059	1.000	1.259

Table 63: All rules (up to 3 items) obtained for cluster T-C12 with collision type G (scenario T-12.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=G	TrfCtrl=None	0.206	1.000	2.125
Coll=G	FirstImpact=Offside	0.206	1.000	1.417
Coll=G	Light=DayNSL	0.176	0.857	1.457
Coll=G	SpdLim=30mph	0.176	0.857	1.325
Coll=G & FirstIntAct=P2W	Light=DayNSL	0.176	1.000	1.700
Coll=G & SpdLim=30mph	Area=Urban	0.176	1.000	1.417
Coll=G & Area=Urban	SpdLim=30mph	0.176	1.000	1.545

Coll=G & DrvInj=Uninjured	SpdLim=30mph	0.147	1.000	1.545
Coll=G & DrvInj=Uninjured	Area=Urban	0.147	1.000	1.417
Coll=G & Prec=PoorOvtkOther	DrvInj=Slight	0.059	1.000	11.333
Coll=G & Prec=PoorOvtkOther	Light=DayNSL	0.059	1.000	1.700

Table 64: All rules (up to 3 items) obtained for cluster T-C12 with collision type M (scenario T-12.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=M	SpdLim=30mph	0.147	1.000	1.545
Coll=M	Area=Urban	0.147	1.000	1.417
Coll=M	DrvInj=Uninjured	0.147	1.000	1.259
Coll=M	Prec=FailGiveWayDriver	0.118	0.800	1.600
Coll=M & TrfCtrl=GW	Prec=FailGiveWayDriver	0.088	1.000	2.000
Coll=M & TrfCtrl=GW	Light=DayNSL	0.088	1.000	1.700
Coll=M & Light=DayNSL	TrfCtrl=GW	0.088	1.000	2.000
Coll=M & Prec=FailGiveWayDriver	Light=DayNSL	0.088	0.750	1.275
Coll=M & Light=DayNSL	Prec=FailGiveWayDriver	0.088	1.000	2.000
Coll=M & TrfCtrl=None	FirstImpact=Offside	0.059	1.000	1.417

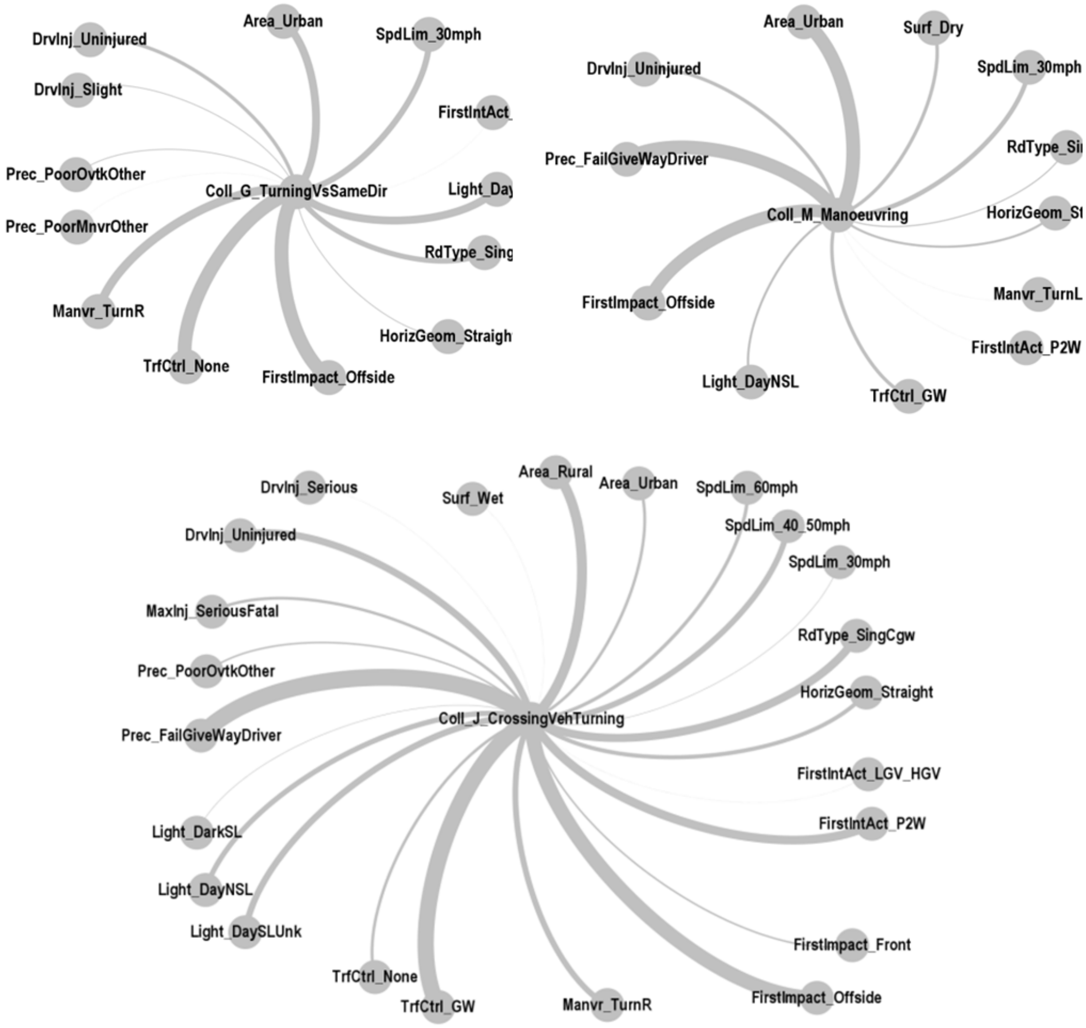


Figure 138: Weighted, directed graphs obtained from all association rules for cluster T-C12

Cluster T-C13: "The car hits another car or PTW with its front, while turning right into a minor road."

Table 65: All rules (up to 3 items) obtained for cluster T-C13 with collision type L (scenario T-13.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L & SpdLim=30mph	Surf=Dry	0.406	0.867	1.261
Coll=L & Prec=FailGiveWayDriver	MaxInj=Slight	0.344	1.000	1.333
Coll=L & Area=Urban	Surf=Dry	0.344	0.917	1.333
Coll=L & DrvInj=Slight	MaxInj=Slight	0.281	1.000	1.333
Coll=L & Prec=PoorMnvrDriver	Surf=Dry	0.250	0.889	1.293
Coll=L & DrvInj=Slight	Surf=Dry	0.250	0.889	1.293
Coll=L & Surf=Wet	FirstIntAct=Car	0.188	1.000	1.524
Coll=L & Surf=Wet	Area=Rural	0.156	0.833	2.051
Coll=L & MaxInj=Uninjured	Prec=PoorMnvrDriver	0.125	1.000	2.909
Coll=L & TrfCtrl=Light	DrvInj=Uninjured	0.094	1.000	1.684
Coll=L & TrfCtrl=Light	MaxInj=Slight	0.094	1.000	1.333
Coll=L & Light=DarkSL	MaxInj=Slight	0.094	1.000	1.333
Coll=L & MaxInj=Uninjured	Area=Rural	0.094	0.750	1.846
Coll=L & MaxInj=SeriousFatal	Light=DaySLUnk	0.063	1.000	4.571
Coll=L & MaxInj=SeriousFatal	SpdLim=30mph	0.063	1.000	1.455
Coll=L & TrfCtrl=GW	FirstIntAct=P2W	0.031	1.000	5.333
Coll=L & TrfCtrl=GW	Light=DaySLUnk	0.031	1.000	4.571
Coll=L & TrfCtrl=GW	DrvInj=Uninjured	0.031	1.000	1.684
Coll=L & TrfCtrl=GW	Area=Urban	0.031	1.000	1.684
Coll=L & TrfCtrl=GW	SpdLim=30mph	0.031	1.000	1.455

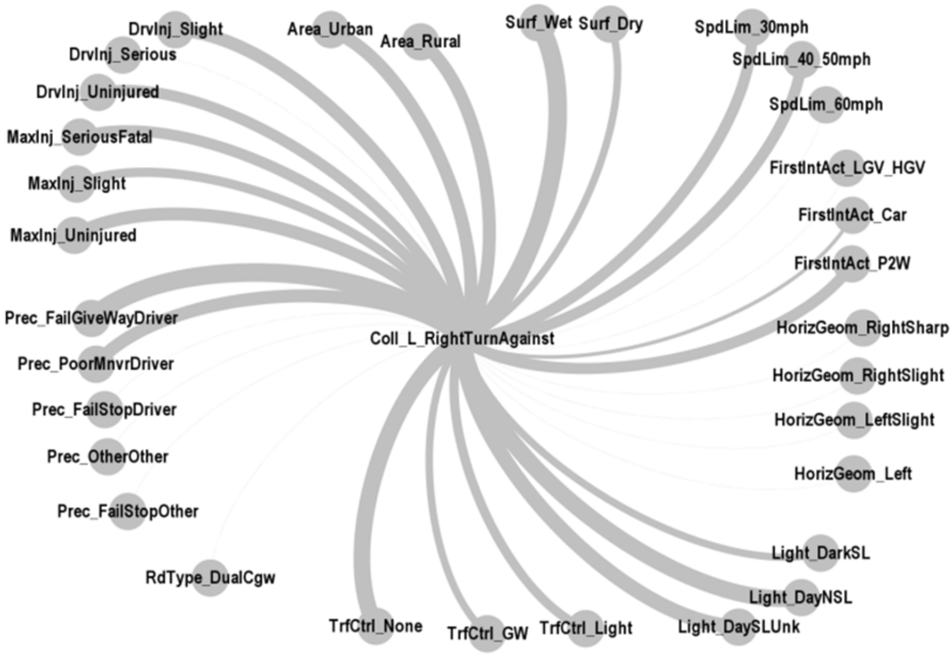


Figure 139: Weighted, directed graph obtained from all association rules for cluster T-C13

Cluster X-C1: "The car hits another road user with its front, while going straight over a crossroad with minor roads joining left and right."

Table 66: All rules (up to 3 items) obtained for cluster X-C1 with collision type H (scenario X-1.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H & SpdLim=30mph	RdType=SingCgw	0.202	1.000	1.308
Coll=H & Prec=FailGiveWayOther	TrfCtrl=None	0.193	1.000	2.017
Coll=H & TrfCtrl=None	Prec=FailGiveWayOther	0.193	0.767	2.401
Coll=H & Area=Rural	TrfCtrl=None	0.109	0.765	1.542
Coll=H & SpdLim=60mph	RdType=SingCgw	0.101	1.000	1.308
Coll=H & SpdLim=60mph	Surf=Dry	0.092	0.917	1.330
Coll=H & MaxInj=SeriousFatal	RdType=SingCgw	0.084	1.000	1.308
Coll=H & Prec=FailStopDriver	TrfCtrl=Light	0.084	1.000	1.983
Coll=H & SpdLim=60mph	DrvInj=Slight	0.076	0.750	1.684
Coll=H & Light=DarkSL	RdType=SingCgw	0.076	1.000	1.308
Coll=H & Light=DarkSL	SpdLim=30mph	0.067	0.889	1.603
Coll=H & FirstIntAct=LGV=HGV	DrvInj=Slight	0.034	1.000	2.245
Coll=H & FirstIntAct=LGV=HGV	MaxInj=Slight	0.034	1.000	1.337

Table 67: All rules (up to 3 items) obtained for cluster X-C1 with collision type L (scenario X-1.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L & Manvr=GoingAheadOther	TrfCtrl=Light	0.126	0.789	1.566
Coll=L & SpdLim=30mph	Area=Urban	0.126	1.000	1.266
Coll=L & DrvInj=Uninjured	TrfCtrl=Light	0.109	0.813	1.611
Coll=L & Surf=Dry	TrfCtrl=Light	0.109	0.765	1.517
Coll=L & Light=DayNSL	RdType=SingCgw	0.101	1.000	1.308
Coll=L & Surf=Wet	MaxInj=Slight	0.092	1.000	1.337
Coll=L & DrvInj=Slight	MaxInj=Slight	0.092	1.000	1.337
Coll=L & DrvInj=Slight	RdType=SingCgw	0.092	1.000	1.308
Coll=L & Manvr=TurnR	MaxInj=Slight	0.076	1.000	1.337
Coll=L & Light=DaySLUnk	Area=Urban	0.076	1.000	1.266
Coll=L & Manvr=TurnR	SpdLim=30mph	0.067	0.889	1.603
Coll=L & SpdLim=40-50mph	TrfCtrl=Light	0.067	1.000	1.983
Coll=L & Prec=FailGiveWayOther	TrfCtrl=Light	0.067	0.889	1.763
Coll=L & TrfCtrl=None	MaxInj=Slight	0.067	1.000	1.337
Coll=L & TrfCtrl=None	RdType=SingCgw	0.067	1.000	1.308
Coll=L & TrfCtrl=None	SpdLim=30mph	0.050	0.750	1.352
Coll=L & Prec=FailGiveWayDriver	Manvr=TurnR	0.042	1.000	9.154
Coll=L & Prec=FailGiveWayDriver	SpdLim=30mph	0.042	1.000	1.803
Coll=L & Prec=FailGiveWayDriver	MaxInj=Slight	0.042	1.000	1.337
Coll=L & Light=DarkSL	TrfCtrl=Light	0.042	1.000	1.983
Coll=L & Light=DarkSL	MaxInj=Slight	0.042	1.000	1.337
Coll=L & Light=DarkSL	Area=Urban	0.042	1.000	1.266

Coll=L & Area=Rural	RdType=SingCgw	0.042	1.000	1.308
Coll=L & Prec=PoorMnvrOther	RdType=SingCgw	0.034	1.000	1.308
Coll=L & Light=DarkSL	Manvr=TurnR	0.034	0.800	7.323
Coll=L & Light=DarkSL	DrvInj=Uninjured	0.034	0.800	1.535
Coll=L & Area=Rural	DrvInj=Slight	0.034	0.800	1.796
Coll=L & Area=Rural	Light=DayNSL	0.034	0.800	1.561

Table 68: All rules (up to 3 items) obtained for cluster X-C1 with collision type F (scenario X-1.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=F	Area=Urban	0.143	1.000	1.266
Coll=F	Light=DayNSL	0.109	0.765	1.492
Coll=F & MaxInj=Slight	Light=DayNSL	0.092	0.846	1.651
Coll=F & FirstIntAct=Car	Light=DayNSL	0.092	0.786	1.533
Coll=F & MaxInj=Slight	TrfCtrl=Light	0.084	0.769	1.526
Coll=F & SpdLim=30mph	RdType=SingCgw	0.076	1.000	1.308
Coll=F & RdType=SingCgw	SpdLim=30mph	0.076	1.000	1.803
Coll=F & FirstImpact=Back	Surf=Dry	0.067	0.889	1.290
Coll=F & RdType=DualCgw	TrfCtrl=Light	0.067	1.000	1.983
Coll=F & RdType=DualCgw	MaxInj=Slight	0.067	1.000	1.337
Coll=F & FirstImpact=Back	Light=DayNSL	0.059	0.778	1.517
Coll=F & RdType=DualCgw	Surf=Dry	0.059	0.875	1.270
Coll=F & SpdLim=30mph	Light=DayNSL	0.059	0.778	1.517
Coll=F & RdType=SingCgw	Light=DayNSL	0.059	0.778	1.517
Coll=F & RdType=DualCgw	SpdLim=40-50mph	0.050	0.750	3.078
Coll=F & SpdLim=40-50mph	RdType=DualCgw	0.050	1.000	4.958
Coll=F & SpdLim=40-50mph	TrfCtrl=Light	0.050	1.000	1.983
Coll=F & SpdLim=40-50mph	Light=DayNSL	0.050	1.000	1.951
Coll=F & SpdLim=40-50mph	MaxInj=Slight	0.050	1.000	1.337
Coll=F & DrvInj=Slight	MaxInj=Slight	0.050	1.000	1.337
Coll=F & FirstImpact=Front	TrfCtrl=Light	0.050	0.750	1.488
Coll=F & FirstImpact=Front	DrvInj=Uninjured	0.050	0.750	1.440
Coll=F & DrvInj=Slight	Light=DayNSL	0.042	0.833	1.626
Coll=F & TrfCtrl=None	Light=DayNSL	0.042	1.000	1.951
Coll=F & TrfCtrl=None	SpdLim=30mph	0.042	1.000	1.803
Coll=F & TrfCtrl=None	Surf=Dry	0.042	1.000	1.451
Coll=F & TrfCtrl=None	RdType=SingCgw	0.042	1.000	1.308
Coll=F & Prec=FailAvoidOther	TrfCtrl=Light	0.034	0.800	1.587
Coll=F & MaxInj=Uninjured	SpdLim=30mph	0.034	1.000	1.803
Coll=F & MaxInj=Uninjured	Surf=Dry	0.034	1.000	1.451
Coll=F & MaxInj=Uninjured	RdType=SingCgw	0.034	1.000	1.308
Coll=F & Prec=FailStopDriver	Light=DayNSL	0.034	1.000	1.951

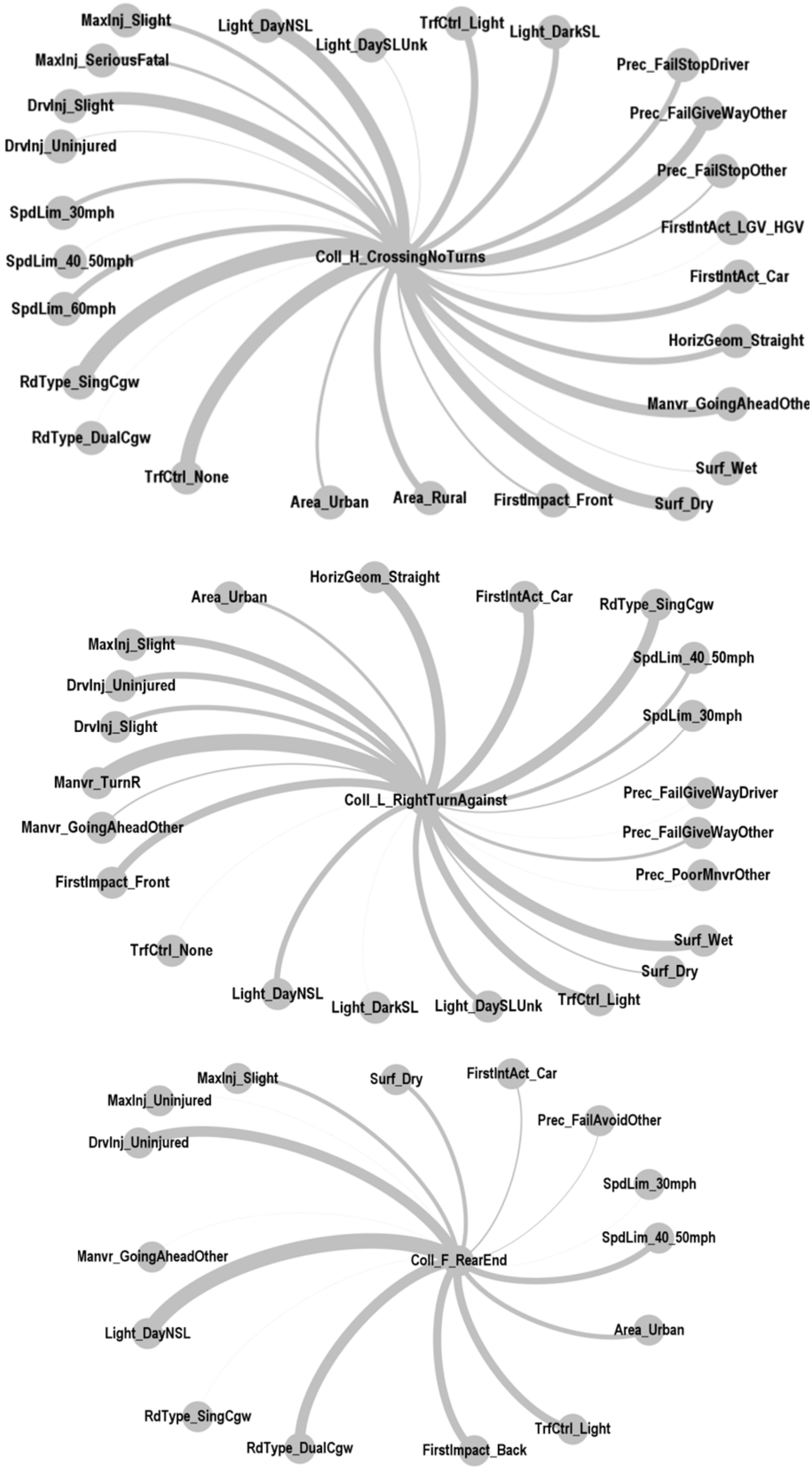


Figure 140: Weighted, directed graphs obtained from all association rules for cluster X-C1

Cluster X-C2: "The car hits another road user with its front, while crossing or turning into a major road."

Table 69: All rules (up to 3 items) obtained for cluster X-C2 with collision type H (scenario X-2.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H	Manvr=GoingAheadOther	0.340	0.810	1.619
Coll=H & Area=Urban	SpdLim=30mph	0.300	0.833	1.302
Coll=H & RdType=SingCgw	Manvr=GoingAheadOther	0.280	0.824	1.647
Coll=H & FirstIntAct=Car	Manvr=GoingAheadOther	0.260	0.813	1.625
Coll=H & Surf=Dry	Manvr=GoingAheadOther	0.240	0.857	1.714
Coll=H & DrvInj=Uninjured	SpdLim=30mph	0.240	0.800	1.250
Coll=H & TrfCtrl=GW	Prec=FailGiveWayDriver	0.200	0.909	1.748
Coll=H & Prec=FailGiveWayDriver	TrfCtrl=GW	0.200	0.769	1.748
Coll=H & MaxInj=Slight	SpdLim=30mph	0.200	0.833	1.302
Coll=H & Light=DayNSL	Prec=FailGiveWayDriver	0.180	0.900	1.731
Coll=H & TrfCtrl=GW	SpdLim=30mph	0.180	0.818	1.278
Coll=H & Surf=Wet	SpdLim=30mph	0.140	1.000	1.563
Coll=H & Light=DarkSL	Manvr=GoingAheadOther	0.100	0.833	1.667
Coll=H & DrvInj=Slight	Manvr=GoingAheadOther	0.100	1.000	2.000
Coll=H & TrfCtrl=Light	Manvr=GoingAheadOther	0.100	1.000	2.000
Coll=H & TrfCtrl=Light	Surf=Dry	0.100	1.000	1.429
Coll=H & Prec=FailStopDriver	Manvr=GoingAheadOther	0.080	1.000	2.000
Coll=H & DrvInj=Slight	Light=DayNSL	0.080	0.800	2.105
Coll=H & SpdLim=40-50mph	Manvr=GoingAheadOther	0.080	1.000	2.000
Coll=H & SpdLim=40-50mph	Surf=Dry	0.080	1.000	1.429
Coll=H & SpdLim=40-50mph	FirstIntAct=Car	0.080	1.000	1.351
Coll=H & Light=DaySLUnk	Manvr=GoingAheadOther	0.080	1.000	2.000
Coll=H & Light=DaySLUnk	SpdLim=30mph	0.080	1.000	1.563
Coll=H & Light=DaySLUnk	MaxInj=Slight	0.080	1.000	1.515
Coll=H & Light=DaySLUnk	FirstIntAct=Car	0.080	1.000	1.351
Coll=H & FirstIntAct=P2W	Light=DayNSL	0.060	0.750	1.974
Coll=H & FirstIntAct=P2W	TrfCtrl=GW	0.060	0.750	1.705
Coll=H & SpdLim=40-50mph	MaxInj=SeriousFatal	0.060	0.750	5.357
Coll=H & TrfCtrl=Stop	Manvr=GoingAheadOther	0.060	1.000	2.000
Coll=H & TrfCtrl=Stop	MaxInj=Slight	0.060	1.000	1.515
Coll=H & Area=Rural	Light=DayNSL	0.060	1.000	2.632
Coll=H & Area=Rural	Manvr=GoingAheadOther	0.060	1.000	2.000
Coll=H & Area=Rural	Prec=FailGiveWayDriver	0.060	1.000	1.923
Coll=H & Area=Rural	Surf=Dry	0.060	1.000	1.429
Coll=H & SpdLim=40-50mph	DrvInj=Slight	0.060	0.750	3.125
Coll=H & RdType=OneWayStr	TrfCtrl=None	0.040	1.000	25.000
Coll=H & Prec=FailStopOther	Light=DarkSL	0.040	1.000	4.167
Coll=H & Prec=FailStopOther	SpdLim=40-50mph	0.040	1.000	3.846
Coll=H & Prec=FailStopOther	Manvr=GoingAheadOther	0.040	1.000	2.000

Coll=H & SpdLim=60mph	Manvr=GoingAheadOther	0.040	1.000	2.000
Coll=H & SpdLim=60mph	DrvInj=Uninjured	0.040	1.000	1.351
Coll=H & RdType=OneWayStr	MaxInj=SeriousFatal	0.040	1.000	7.143
Coll=H & RdType=OneWayStr	Light=DarkSL	0.040	1.000	4.167
Coll=H & RdType=OneWayStr	Surf=Wet	0.040	1.000	3.333
Coll=H & RdType=DualCgw	MaxInj=Slight	0.040	1.000	1.515
Coll=H & Manvr=TurnL	Light=DayNSL	0.040	1.000	2.632
Coll=H & Manvr=TurnL	TrfCtrl=GW	0.040	1.000	2.273
Coll=H & Manvr=TurnL	Prec=FailGiveWayDriver	0.040	1.000	1.923
Coll=H & Manvr=TurnL	SpdLim=30mph	0.040	1.000	1.563
Coll=H & Manvr=TurnL	MaxInj=Slight	0.040	1.000	1.515
Coll=H & Manvr=TurnL	DrvInj=Uninjured	0.040	1.000	1.351
Coll=H & Manvr=TurnL	FirstIntAct=Car	0.040	1.000	1.351
Coll=H & MaxInj=Uninjured	TrfCtrl=GW	0.040	1.000	2.273
Coll=H & MaxInj=Uninjured	Manvr=GoingAheadOther	0.040	1.000	2.000
Coll=H & MaxInj=Uninjured	SpdLim=30mph	0.040	1.000	1.563
Coll=H & MaxInj=Uninjured	Surf=Dry	0.040	1.000	1.429
Coll=H & MaxInj=Uninjured	FirstIntAct=Car	0.040	1.000	1.351
Coll=H & Manvr=TurnR	Prec=FailGiveWayDriver	0.040	1.000	1.923
Coll=H & Manvr=TurnR	SpdLim=30mph	0.040	1.000	1.563
Coll=H & Manvr=TurnR	DrvInj=Uninjured	0.040	1.000	1.351

Table 70: All rules (up to 3 items) obtained for cluster X-C2 with collision type F (scenario X-2.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=F	MaxInj=Slight	0.100	0.833	1.263
Coll=F & HorizGeom=LeftSlight	Manvr=TurnL	0.080	1.000	5.556
Coll=F & Manvr=TurnL	HorizGeom=LeftSlight	0.080	1.000	8.333
Coll=F & MaxInj=Slight	HorizGeom=LeftSlight	0.080	0.800	6.667
Coll=F & RdType=SingCgw	HorizGeom=LeftSlight	0.080	0.800	6.667
Coll=F & Manvr=TurnL	MaxInj=Slight	0.080	1.000	1.515
Coll=F & MaxInj=Slight	Manvr=TurnL	0.080	0.800	4.444
Coll=F & RdType=SingCgw	Manvr=TurnL	0.080	0.800	4.444
Coll=F & FirstIntAct=Car	MaxInj=Slight	0.080	1.000	1.515
Coll=F & FirstImpact=Front	MaxInj=Slight	0.080	1.000	1.515
Coll=F & DrvInj=Uninjured	Prec=FailAvoidDriver	0.060	0.750	12.500
Coll=F & FirstImpact=Front	Prec=FailAvoidDriver	0.060	0.750	12.500
Coll=F & FirstIntAct=Car	HorizGeom=LeftSlight	0.060	0.750	6.250
Coll=F & FirstImpact=Front	HorizGeom=LeftSlight	0.060	0.750	6.250
Coll=F & FirstIntAct=Car	Manvr=TurnL	0.060	0.750	4.167
Coll=F & FirstImpact=Front	Manvr=TurnL	0.060	0.750	4.167
Coll=F & SpdLim=40-50mph	TrfCtrl=Light	0.060	1.000	2.632
Coll=F & TrfCtrl=Light	SpdLim=40-50mph	0.060	1.000	3.846
Coll=F & SpdLim=40-50mph	MaxInj=Slight	0.060	1.000	1.515
Coll=F & SpdLim=40-50mph	Surf=Dry	0.060	1.000	1.429
Coll=F & Surf=Dry	SpdLim=40-50mph	0.060	1.000	3.846

Coll=F & SpdLim=40-50mph	FirstIntAct=Car	0.060	1.000	1.351
Coll=F & FirstIntAct=Car	SpdLim=40-50mph	0.060	0.750	2.885
Coll=F & Area=Urban	SpdLim=40-50mph	0.060	0.750	2.885
Coll=F & Surf=Wet	TrfCtrl=GW	0.060	1.000	2.273
Coll=F & TrfCtrl=GW	Surf=Wet	0.060	1.000	3.333
Coll=F & TrfCtrl=Light	MaxInj=Slight	0.060	1.000	1.515
Coll=F & TrfCtrl=Light	Surf=Dry	0.060	1.000	1.429
Coll=F & Surf=Dry	TrfCtrl=Light	0.060	1.000	2.632
Coll=F & TrfCtrl=Light	FirstIntAct=Car	0.060	1.000	1.351
Coll=F & FirstIntAct=Car	TrfCtrl=Light	0.060	0.750	1.974
Coll=F & Area=Urban	TrfCtrl=Light	0.060	0.750	1.974
Coll=F & Surf=Dry	MaxInj=Slight	0.060	1.000	1.515
Coll=F & Surf=Dry	FirstIntAct=Car	0.060	1.000	1.351
Coll=F & Light=DarkNSL	SpdLim=70mph	0.040	1.000	25.000
Coll=F & Area=Rural	SpdLim=70mph	0.040	1.000	25.000
Coll=F & Light=DarkNSL	HorizGeom=LeftSlight	0.040	1.000	8.333
Coll=F & Area=Rural	Light=DarkNSL	0.040	1.000	12.500
Coll=F & Light=DarkNSL	Manvr=TurnL	0.040	1.000	5.556
Coll=F & Light=DarkNSL	Surf=Wet	0.040	1.000	3.333
Coll=F & Light=DarkNSL	TrfCtrl=GW	0.040	1.000	2.273
Coll=F & FirstImpact=Back	Prec=FailAvoidOther	0.040	1.000	12.500
Coll=F & DrvInj=Slight	Prec=FailAvoidOther	0.040	1.000	12.500
Coll=F & FirstIntAct=LGV=HGV	Surf=Wet	0.040	1.000	3.333
Coll=F & FirstIntAct=LGV=HGV	TrfCtrl=GW	0.040	1.000	2.273
Coll=F & Area=Rural	HorizGeom=LeftSlight	0.040	1.000	8.333
Coll=F & DrvInj=Slight	HorizGeom=LeftSlight	0.040	1.000	8.333
Coll=F & Light=DayNSL	HorizGeom=LeftSlight	0.040	1.000	8.333
Coll=F & Area=Rural	Manvr=TurnL	0.040	1.000	5.556
Coll=F & Area=Rural	Surf=Wet	0.040	1.000	3.333
Coll=F & Area=Rural	TrfCtrl=GW	0.040	1.000	2.273
Coll=F & Area=Rural	MaxInj=Slight	0.040	1.000	1.515
Coll=F & DrvInj=Slight	Manvr=TurnL	0.040	1.000	5.556
Coll=F & Light=DayNSL	Manvr=TurnL	0.040	1.000	5.556
Coll=F & DrvInj=Slight	MaxInj=Slight	0.040	1.000	1.515
Coll=F & Light=DayNSL	SpdLim=40-50mph	0.040	1.000	3.846
Coll=F & Light=DayNSL	TrfCtrl=Light	0.040	1.000	2.632
Coll=F & Light=DayNSL	MaxInj=Slight	0.040	1.000	1.515
Coll=F & Light=DayNSL	Surf=Dry	0.040	1.000	1.429
Coll=F & Light=DayNSL	FirstIntAct=Car	0.040	1.000	1.351
Coll=F & HorizGeom=Straight	DrvInj=Uninjured	0.040	1.000	1.351

Table 71: All rules (up to 3 items) obtained for cluster X-C2 with collision type J (scenario X-2.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J	Manvr=TurnR	0.080	1.000	3.125
Coll=J	TrfCtrl=GW	0.080	1.000	2.273

Coll=J	Prec=FailGiveWayDriver	0.080	1.000	1.923
Coll=J & MaxInj=Uninjured	SpdLim=30mph	0.040	1.000	1.563
Coll=J & MaxInj=Uninjured	FirstIntAct=Car	0.040	1.000	1.351
Coll=J & Surf=Wet	SpdLim=30mph	0.040	1.000	1.563
Coll=J & Surf=Wet	FirstIntAct=Car	0.040	1.000	1.351
Coll=J & Light=DayNSL	MaxInj=Slight	0.040	1.000	1.515
Coll=J & MaxInj=Slight	Light=DayNSL	0.040	1.000	2.632
Coll=J & SpdLim=30mph	FirstIntAct=Car	0.060	1.000	1.351
Coll=J & FirstIntAct=Car	SpdLim=30mph	0.060	1.000	1.563
Coll=J & Surf=Dry	DrvInj=Uninjured	0.040	1.000	1.351

Table 72: All rules (up to 3 items) obtained for cluster X-C2 with collision type L (scenario X-2.4), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L	TrfCtrl=Light	0.160	1.000	2.632
Coll=L	FirstIntAct=Car	0.160	1.000	1.351
Coll=L	MaxInj=Slight	0.140	0.875	1.326
Coll=L	Surf=Dry	0.140	0.875	1.250
Coll=L & Area=Urban	Surf=Dry	0.140	1.000	1.429
Coll=L & RdType=SingCgw	MaxInj=Slight	0.120	1.000	1.515
Coll=L & Manvr=GoingAheadOther	Surf=Dry	0.100	1.000	1.429
Coll=L & SpdLim=30mph	MaxInj=Slight	0.100	1.000	1.515
Coll=L & SpdLim=30mph	Surf=Dry	0.100	1.000	1.429
Coll=L & RdType=SingCgw	SpdLim=30mph	0.100	0.833	1.302
Coll=L & DrvInj=Uninjured	Surf=Dry	0.100	1.000	1.429
Coll=L & Prec=FailGiveWayOther	MaxInj=Slight	0.080	1.000	1.515
Coll=L & Prec=FailGiveWayOther	Surf=Dry	0.080	1.000	1.429
Coll=L & Light=DaySLUnk	Surf=Dry	0.080	1.000	1.429
Coll=L & DrvInj=Uninjured	Light=DaySLUnk	0.080	0.800	2.857
Coll=L & DrvInj=Uninjured	Manvr=GoingAheadOther	0.080	0.800	1.600
Coll=L & DrvInj=Uninjured	SpdLim=30mph	0.080	0.800	1.250
Coll=L & DrvInj=Slight	MaxInj=Slight	0.060	1.000	1.515
Coll=L & Light=DaySLUnk	Manvr=GoingAheadOther	0.060	0.750	1.500
Coll=L & Manvr=TurnR	MaxInj=Slight	0.060	1.000	1.515
Coll=L & Prec=FailGiveWayDriver	MaxInj=Slight	0.060	1.000	1.515
Coll=L & RdType=DualCgw	SpdLim=40-50mph	0.040	1.000	3.846
Coll=L & Light=DayNSL	SpdLim=30mph	0.040	1.000	1.563
Coll=L & Light=DayNSL	MaxInj=Slight	0.040	1.000	1.515
Coll=L & Light=DayNSL	Surf=Dry	0.040	1.000	1.429

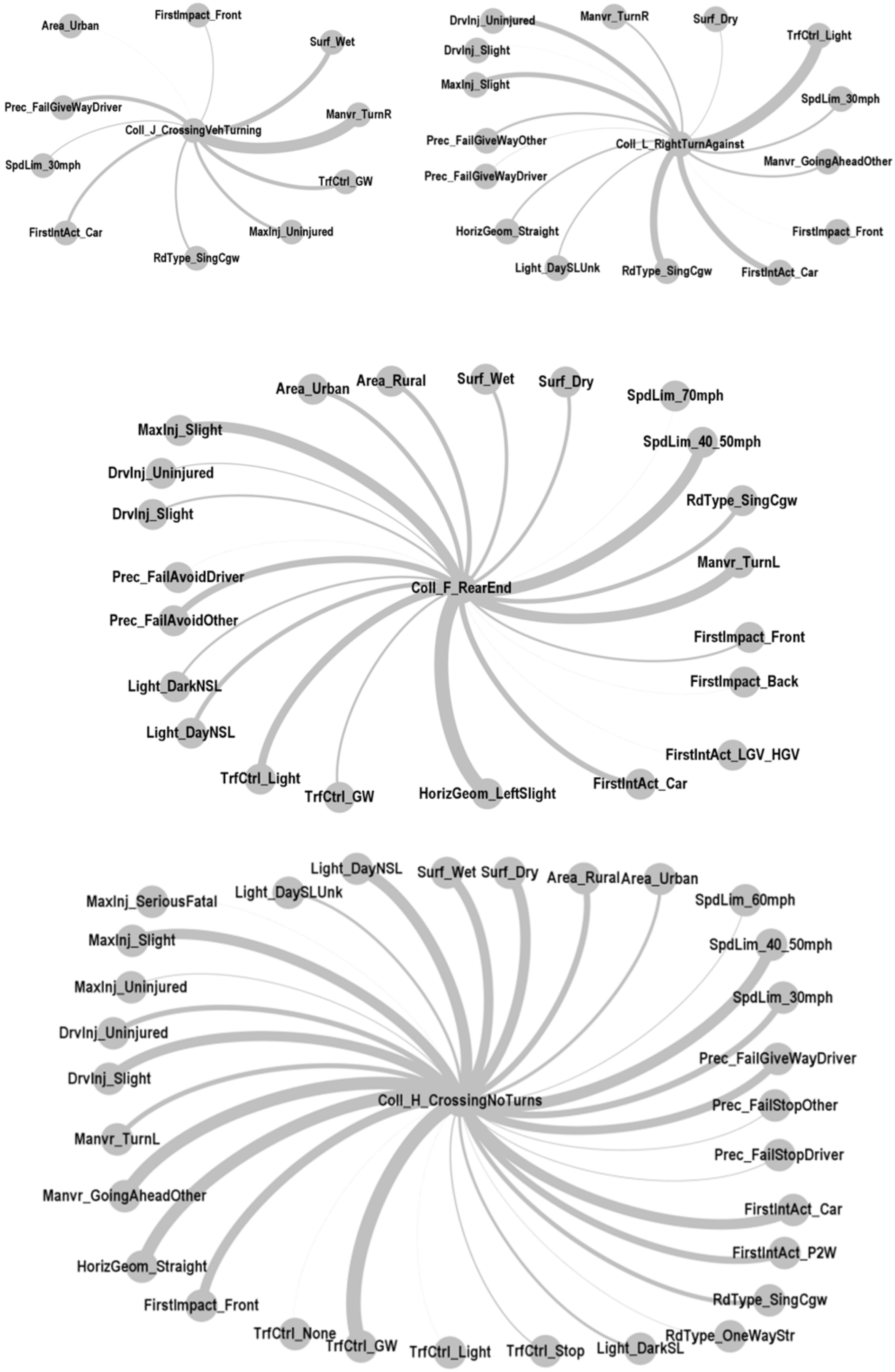


Figure 141: Weighted, directed graphs obtained from all association rules for cluster X-C2

Cluster X-C3: "The car is turning right into a minor road, when being hit on its offside by another vehicle."

Table 73: All rules (up to 3 items) obtained for cluster X-C3 with collision type L (scenario X-3.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L	Manvr=TurnR	0.341	0.933	2.251
Coll=L	TrfCtrl=Light	0.293	0.800	1.367
Coll=L & Manvr=TurnR	TrfCtrl=Light	0.293	0.857	1.464
Coll=L & TrfCtrl=Light	Manvr=TurnR	0.293	1.000	2.412
Coll=L & DrvInj=Uninjured	Manvr=TurnR	0.220	1.000	2.412
Coll=L & Light=DayNSL	Manvr=TurnR	0.220	1.000	2.412
Coll=L & Prec=FailGiveWayDriver	Manvr=TurnR	0.195	1.000	2.412
Coll=L & Prec=FailGiveWayDriver	TrfCtrl=Light	0.195	1.000	1.708
Coll=L & SpdLim=40-50mph	Manvr=TurnR	0.195	1.000	2.412
Coll=L & SpdLim=40-50mph	TrfCtrl=Light	0.195	1.000	1.708
Coll=L & Area=Urban	Manvr=TurnR	0.195	1.000	2.412
Coll=L & Light=DayNSL	TrfCtrl=Light	0.195	0.889	1.519
Coll=L & Area=Urban	TrfCtrl=Light	0.195	1.000	1.708
Coll=L & Area=Urban	Prec=FailGiveWayDriver	0.171	0.875	3.261
Coll=L & Surf=Wet	Manvr=TurnR	0.122	1.000	2.412
Coll=L & Surf=Wet	MaxInj=Slight	0.122	1.000	1.281
Coll=L & Prec=PoorMnvrDriver	Manvr=TurnR	0.098	1.000	2.412
Coll=L & DrvInj=Slight	FirstIntAct=Car	0.098	1.000	1.323
Coll=L & Prec=PoorMnvrDriver	Area=Rural	0.073	0.750	2.563
Coll=L & SpdLim=60mph	Area=Rural	0.073	0.750	2.563
Coll=L & SpdLim=60mph	TrfCtrl=None	0.073	0.750	1.922
Coll=L & TrfCtrl=None	SpdLim=60mph	0.073	1.000	8.200
Coll=L & SpdLim=60mph	DrvInj=Uninjured	0.073	0.750	1.618
Coll=L & Light=DaySLUnk	Prec=FailGiveWayDriver	0.073	1.000	3.727
Coll=L & Light=DaySLUnk	Manvr=TurnR	0.073	1.000	2.412
Coll=L & Light=DaySLUnk	TrfCtrl=Light	0.073	1.000	1.708
Coll=L & Light=DaySLUnk	MaxInj=Slight	0.073	1.000	1.281
Coll=L & SpdLim=30mph	Prec=FailGiveWayDriver	0.073	1.000	3.727
Coll=L & TrfCtrl=None	Area=Rural	0.073	1.000	3.417
Coll=L & TrfCtrl=None	FirstIntAct=Car	0.073	1.000	1.323
Coll=L & TrfCtrl=None	MaxInj=Slight	0.073	1.000	1.281
Coll=L & SpdLim=30mph	Manvr=TurnR	0.073	1.000	2.412
Coll=L & SpdLim=30mph	TrfCtrl=Light	0.073	1.000	1.708
Coll=L & SpdLim=30mph	Area=Urban	0.073	1.000	1.414
Coll=L & SpdLim=30mph	MaxInj=Slight	0.073	1.000	1.281
Coll=L & MaxInj=SeriousFatal	DrvInj=Serious	0.049	1.000	10.250
Coll=L & DrvInj=Serious	Manvr=TurnR	0.049	1.000	2.412
Coll=L & DrvInj=Serious	TrfCtrl=Light	0.049	1.000	1.708
Coll=L & DrvInj=Serious	Surf=Dry	0.049	1.000	1.323

Coll=L & MaxInj=Uninjured	Manvr=TurnR	0.049	1.000	2.412
Coll=L & MaxInj=Uninjured	TrfCtrl=Light	0.049	1.000	1.708
Coll=L & MaxInj=Uninjured	Light=DayNSL	0.049	1.000	1.708
Coll=L & MaxInj=Uninjured	Surf=Dry	0.049	1.000	1.323
Coll=L & RdType=DualCgw	Surf=Wet	0.049	1.000	4.100
Coll=L & RdType=DualCgw	Prec=FailGiveWayDriver	0.049	1.000	3.727
Coll=L & RdType=DualCgw	Manvr=TurnR	0.049	1.000	2.412
Coll=L & RdType=DualCgw	DrvInj=Uninjured	0.049	1.000	2.158
Coll=L & MaxInj=SeriousFatal	Manvr=TurnR	0.049	1.000	2.412
Coll=L & MaxInj=SeriousFatal	TrfCtrl=Light	0.049	1.000	1.708
Coll=L & MaxInj=SeriousFatal	Surf=Dry	0.049	1.000	1.323

Table 74: All rules (up to 3 items) obtained for cluster X-C3 with collision type H (scenario X-3.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H	Manvr=GoingAheadOther	0.366	1.000	1.864
Coll=H & MaxInj=Slight	Surf=Dry	0.293	1.000	1.323
Coll=H & RdType=SingCgw	TrfCtrl=None	0.244	0.833	2.135
Coll=H & Area=Urban	Surf=Dry	0.244	1.000	1.323
Coll=H & Area=Urban	MaxInj=Slight	0.244	1.000	1.281
Coll=H & DrvInj=Slight	Surf=Dry	0.220	1.000	1.323
Coll=H & MaxInj=Slight	DrvInj=Slight	0.220	0.750	1.708
Coll=H & RdType=SingCgw	Light=DayNSL	0.220	0.750	1.281
Coll=H & TrfCtrl=None	Light=DayNSL	0.195	0.800	1.367
Coll=H & Light=DayNSL	TrfCtrl=None	0.195	0.889	2.278
Coll=H & SpdLim=30mph	Surf=Dry	0.195	1.000	1.323
Coll=H & SpdLim=30mph	MaxInj=Slight	0.195	1.000	1.281
Coll=H & SpdLim=30mph	TrfCtrl=None	0.171	0.875	2.242
Coll=H & Prec=FailGiveWayOther	TrfCtrl=None	0.146	1.000	2.563
Coll=H & Prec=FailGiveWayOther	Surf=Dry	0.146	1.000	1.323
Coll=H & SpdLim=30mph	Light=DayNSL	0.146	0.750	1.281
Coll=H & Prec=FailGiveWayOther	SpdLim=30mph	0.122	0.833	1.708
Coll=H & Prec=FailGiveWayOther	Light=DayNSL	0.122	0.833	1.424
Coll=H & Area=Rural	TrfCtrl=None	0.122	1.000	2.563
Coll=H & Area=Rural	FirstIntAct=Car	0.122	1.000	1.323
Coll=H & TrfCtrl=Light	Area=Urban	0.122	1.000	1.414
Coll=H & TrfCtrl=Light	Surf=Dry	0.122	1.000	1.323
Coll=H & TrfCtrl=Light	MaxInj=Slight	0.122	1.000	1.281
Coll=H & Light=DarkSL	Surf=Dry	0.098	1.000	1.323
Coll=H & Area=Rural	Light=DayNSL	0.098	0.800	1.367
Coll=H & TrfCtrl=Light	DrvInj=Slight	0.098	0.800	1.822
Coll=H & DrvInj=Uninjured	Light=DayNSL	0.098	1.000	1.708
Coll=H & Prec=FailStopDriver	Light=DarkSL	0.073	1.000	6.833
Coll=H & Light=DarkSL	Prec=FailStopDriver	0.073	0.750	7.688
Coll=H & Prec=FailStopDriver	DrvInj=Slight	0.073	1.000	2.278
Coll=H & Prec=FailStopDriver	MaxInj=Slight	0.073	1.000	1.281

Coll=H & RdType=DualCgw	DrvInj=Slight	0.073	1.000	2.278
Coll=H & RdType=DualCgw	FirstIntAct=Car	0.073	1.000	1.323
Coll=H & RdType=DualCgw	Surf=Dry	0.073	1.000	1.323
Coll=H & MaxInj=SeriousFatal	Area=Rural	0.073	1.000	3.417
Coll=H & MaxInj=SeriousFatal	TrfCtrl=None	0.073	1.000	2.563
Coll=H & MaxInj=SeriousFatal	Light=DayNSL	0.073	1.000	1.708
Coll=H & DrvInj=Uninjured	TrfCtrl=None	0.073	0.750	1.922
Coll=H & DrvInj=Uninjured	SpdLim=30mph	0.073	0.750	1.538
Coll=H & DrvInj=Serious	Area=Rural	0.049	1.000	3.417
Coll=H & DrvInj=Serious	TrfCtrl=None	0.049	1.000	2.563
Coll=H & DrvInj=Serious	Light=DayNSL	0.049	1.000	1.708
Coll=H & Prec=FailStopOther	Light=DaySLUnk	0.049	1.000	5.125
Coll=H & Light=DaySLUnk	Prec=FailStopOther	0.049	1.000	8.200
Coll=H & Prec=FailStopOther	Area=Urban	0.049	1.000	1.414
Coll=H & Prec=FailStopOther	FirstIntAct=Car	0.049	1.000	1.323
Coll=H & Prec=FailStopOther	Surf=Dry	0.049	1.000	1.323
Coll=H & Surf=Wet	MaxInj=SeriousFatal	0.049	1.000	8.200
Coll=H & Light=DaySLUnk	DrvInj=Slight	0.049	1.000	2.278
Coll=H & Light=DaySLUnk	Surf=Dry	0.049	1.000	1.323
Coll=H & Light=DaySLUnk	MaxInj=Slight	0.049	1.000	1.281
Coll=H & Surf=Wet	Area=Rural	0.049	1.000	3.417
Coll=H & Surf=Wet	TrfCtrl=None	0.049	1.000	2.563
Coll=H & Surf=Wet	SpdLim=40-50mph	0.049	1.000	2.563
Coll=H & Surf=Wet	Light=DayNSL	0.049	1.000	1.708
Coll=H & Surf=Wet	FirstIntAct=Car	0.049	1.000	1.323

Table 75: All rules (up to 3 items) obtained for cluster X-C3 with collision type J (scenario X-3.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J	SpdLim=30mph	0.146	1.000	2.050
Coll=J	Area=Urban	0.146	1.000	1.414
Coll=J & Surf=Wet	FirstIntAct=Car	0.073	1.000	1.323
Coll=J & TrfCtrl=None	Manvr=GoingAheadOther	0.073	1.000	1.864
Coll=J & Manvr=GoingAheadOther	TrfCtrl=None	0.073	1.000	2.563
Coll=J & TrfCtrl=None	MaxInj=Slight	0.073	1.000	1.281
Coll=J & Manvr=TurnR	FirstIntAct=Car	0.073	1.000	1.323
Coll=J & DrvInj=Slight	FirstIntAct=Car	0.073	1.000	1.323
Coll=J & Manvr=GoingAheadOther	MaxInj=Slight	0.073	1.000	1.281
Coll=J & Surf=Dry	MaxInj=Slight	0.073	1.000	1.281
Coll=J & Light=DaySLUnk	Surf=Dry	0.049	1.000	1.323
Coll=J & Light=DaySLUnk	MaxInj=Slight	0.049	1.000	1.281
Coll=J & Prec=FailGiveWayOther	TrfCtrl=None	0.049	1.000	2.563
Coll=J & Prec=FailGiveWayOther	DrvInj=Slight	0.049	1.000	2.278
Coll=J & Prec=FailGiveWayOther	Manvr=GoingAheadOther	0.049	1.000	1.864
Coll=J & Prec=FailGiveWayOther	FirstIntAct=Car	0.049	1.000	1.323
Coll=J & Prec=FailGiveWayOther	MaxInj=Slight	0.049	1.000	1.281

Coll=J & Prec=FailGiveWayDriver	Manvr=TurnR	0.049	1.000	2.412
Coll=J & Prec=FailGiveWayDriver	DrvInj=Uninjured	0.049	1.000	2.158
Coll=J & Prec=FailGiveWayDriver	FirstIntAct=Car	0.049	1.000	1.323
Coll=J & TrfCtrl=Light	Manvr=TurnR	0.049	1.000	2.412
Coll=J & TrfCtrl=Light	FirstIntAct=Car	0.049	1.000	1.323
Coll=J & TrfCtrl=Light	MaxInj=Slight	0.049	1.000	1.281

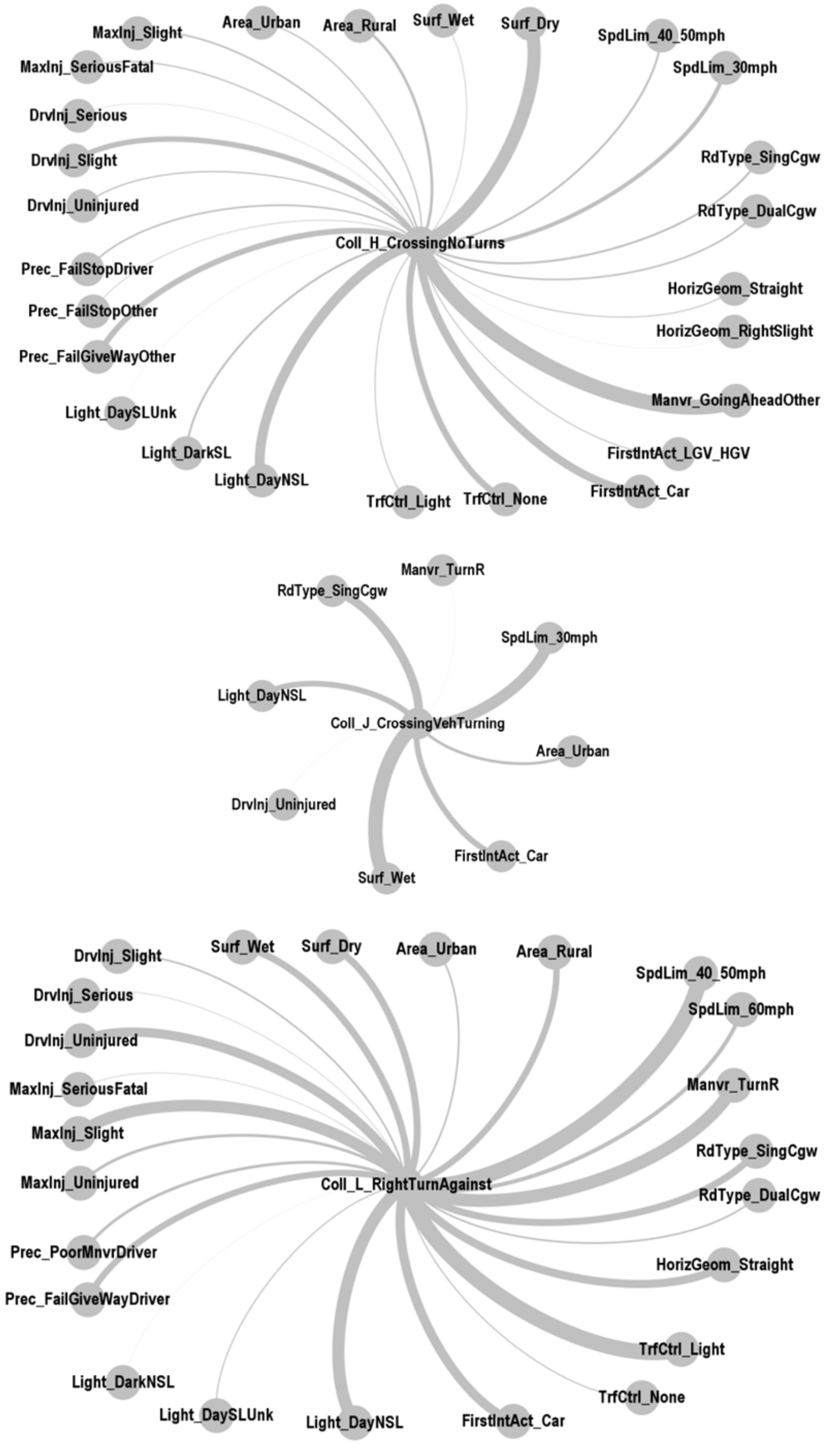


Figure 142: Weighted, directed graphs obtained from all association rules for cluster X-C3

Cluster X-C4: "The car is crossing or turning into a major road, when being hit on its offside by another vehicle."

Table 76: All rules (up to 3 items) obtained for cluster X-C4 with collision type H (scenario X-4.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H	Manvr=GoingAheadOther	0.561	0.958	1.310
Coll=H & Area=Urban	Manvr=GoingAheadOther	0.390	1.000	1.367
Coll=H & MaxInj=Slight	Manvr=GoingAheadOther	0.341	1.000	1.367
Coll=H & TrfCtrl=Light	Manvr=GoingAheadOther	0.293	1.000	1.367
Coll=H & Area=Urban	DrvInj=Uninjured	0.293	0.750	1.337
Coll=H & SpdLim=30mph	Manvr=GoingAheadOther	0.268	1.000	1.367
Coll=H & Prec=FailGiveWayDriver	TrfCtrl=GW	0.268	0.917	1.708
Coll=H & TrfCtrl=GW	Prec=FailGiveWayDriver	0.268	1.000	2.050
Coll=H & TrfCtrl=Light	Area=Urban	0.244	0.833	1.367
Coll=H & Light=DayNSL	Surf=Dry	0.244	1.000	1.577
Coll=H & Light=DayNSL	Manvr=GoingAheadOther	0.244	1.000	1.367
Coll=H & SpdLim=30mph	DrvInj=Uninjured	0.220	0.818	1.458
Coll=H & Prec=FailGiveWayDriver	DrvInj=Uninjured	0.220	0.750	1.337
Coll=H & TrfCtrl=GW	DrvInj=Uninjured	0.220	0.818	1.458
Coll=H & SpdLim=40-50mph	Manvr=GoingAheadOther	0.195	1.000	1.367
Coll=H & Light=DarkSL	Manvr=GoingAheadOther	0.171	1.000	1.367
Coll=H & DrvInj=Slight	Manvr=GoingAheadOther	0.171	1.000	1.367
Coll=H & SpdLim=40-50mph	TrfCtrl=Light	0.146	0.750	1.809
Coll=H & DrvInj=Slight	Surf=Dry	0.146	0.857	1.352
Coll=H & Surf=Wet	Manvr=GoingAheadOther	0.146	1.000	1.367
Coll=H & Area=Rural	Prec=FailGiveWayDriver	0.146	0.750	1.538
Coll=H & SpdLim=60mph	Surf=Dry	0.122	1.000	1.577
Coll=H & Prec=FailGiveWayOther	TrfCtrl=Light	0.122	1.000	2.412
Coll=H & Prec=FailGiveWayOther	Area=Urban	0.122	1.000	1.640
Coll=H & Prec=FailStopOther	Manvr=GoingAheadOther	0.122	1.000	1.367
Coll=H & MaxInj=SeriousFatal	Surf=Dry	0.122	1.000	1.577
Coll=H & MaxInj=SeriousFatal	Manvr=GoingAheadOther	0.122	1.000	1.367
Coll=H & Surf=Wet	DrvInj=Uninjured	0.122	0.833	1.486
Coll=H & SpdLim=60mph	Light=DayNSL	0.098	0.800	1.640
Coll=H & Prec=FailGiveWayOther	Surf=Dry	0.098	0.800	1.262
Coll=H & Prec=FailStopOther	Surf=Dry	0.098	0.800	1.262
Coll=H & Light=DarkNSL	Prec=FailGiveWayDriver	0.073	1.000	2.050
Coll=H & Light=DaySLUnk	Manvr=GoingAheadOther	0.073	1.000	1.367
Coll=H & RdType=DualCgw	SpdLim=40-50mph	0.073	1.000	2.929
Coll=H & RdType=DualCgw	TrfCtrl=Light	0.073	1.000	2.412
Coll=H & RdType=DualCgw	DrvInj=Uninjured	0.073	1.000	1.783
Coll=H & RdType=DualCgw	Area=Urban	0.073	1.000	1.640
Coll=H & RdType=DualCgw	Manvr=GoingAheadOther	0.073	1.000	1.367

Table 77: All rules (up to 3 items) obtained for cluster X-C4 with collision type J (scenario X-4.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=J & HorizGeom=Straight	TrfCtrl=GW	0.171	0.778	1.449
Coll=J & RdType=SingCgw	TrfCtrl=GW	0.146	0.857	1.597
Coll=J & Manvr=TurnR	Prec=FailGiveWayDriver	0.122	0.833	1.708
Coll=J & Prec=FailGiveWayDriver	Manvr=TurnR	0.122	0.833	3.796
Coll=J & Manvr=TurnR	Light=DayNSL	0.122	0.833	1.708
Coll=J & Light=DayNSL	Manvr=TurnR	0.122	1.000	4.556
Coll=J & Prec=FailGiveWayDriver	Light=DayNSL	0.122	0.833	1.708
Coll=J & Light=DayNSL	Prec=FailGiveWayDriver	0.122	1.000	2.050
Coll=J & Surf=Wet	Manvr=TurnR	0.098	1.000	4.556
Coll=J & Area=Rural	Manvr=TurnR	0.098	0.800	3.644
Coll=J & DrvInj=Slight	MaxInj=Slight	0.098	1.000	1.577
Coll=J & MaxInj=Slight	DrvInj=Slight	0.098	0.800	2.343
Coll=J & Area=Rural	Prec=FailGiveWayDriver	0.098	0.800	1.640
Coll=J & Area=Rural	Light=DayNSL	0.098	0.800	1.640
Coll=J & Light=DayNSL	Area=Rural	0.098	0.800	2.050
Coll=J & Area=Rural	TrfCtrl=GW	0.098	0.800	1.491
Coll=J & SpdLim=30mph	Area=Urban	0.098	1.000	1.640
Coll=J & Area=Urban	SpdLim=30mph	0.098	0.800	1.822
Coll=J & Light=DayNSL	TrfCtrl=GW	0.098	0.800	1.491
Coll=J & Area=Urban	Surf=Dry	0.098	0.800	1.262
Coll=J & Manvr=GoingAheadOther	Surf=Dry	0.098	1.000	1.577
Coll=J & MaxInj=SeriousFatal	Manvr=TurnR	0.073	1.000	4.556
Coll=J & MaxInj=SeriousFatal	Area=Rural	0.073	1.000	2.563
Coll=J & MaxInj=SeriousFatal	Prec=FailGiveWayDriver	0.073	1.000	2.050
Coll=J & MaxInj=SeriousFatal	Light=DayNSL	0.073	1.000	2.050
Coll=J & MaxInj=SeriousFatal	TrfCtrl=GW	0.073	1.000	1.864
Coll=J & Light=DarkSL	Area=Urban	0.073	1.000	1.640
Coll=J & Surf=Wet	Area=Rural	0.073	0.750	1.922
Coll=J & Surf=Wet	Prec=FailGiveWayDriver	0.073	0.750	1.538
Coll=J & Surf=Wet	Light=DayNSL	0.073	0.750	1.538
Coll=J & Surf=Wet	TrfCtrl=GW	0.073	0.750	1.398
Coll=J & SpdLim=30mph	TrfCtrl=GW	0.073	0.750	1.398
Coll=J & DrvInj=Uninjured	Area=Urban	0.073	1.000	1.640
Coll=J & DrvInj=Uninjured	Surf=Dry	0.073	1.000	1.577
Coll=J & FirstIntAct=LGV=HGV	DrvInj=Fatal	0.049	1.000	20.500
Coll=J & FirstIntAct=LGV=HGV	MaxInj=SeriousFatal	0.049	1.000	5.125
Coll=J & FirstIntAct=LGV=HGV	Manvr=TurnR	0.049	1.000	4.556
Coll=J & FirstIntAct=LGV=HGV	SpdLim=40-50mph	0.049	1.000	2.929
Coll=J & FirstIntAct=LGV=HGV	Area=Rural	0.049	1.000	2.563
Coll=J & FirstIntAct=LGV=HGV	Prec=FailGiveWayDriver	0.049	1.000	2.050
Coll=J & FirstIntAct=LGV=HGV	Light=DayNSL	0.049	1.000	2.050
Coll=J & FirstIntAct=LGV=HGV	TrfCtrl=GW	0.049	1.000	1.864
Coll=J & Prec=FailStopOther	RdType=DualCgw	0.049	1.000	6.833

Coll=J & TrfCtrl=Light	RdType=DualCgw	0.049	1.000	6.833
Coll=J & MaxInj=Uninjured	Area=Urban	0.049	1.000	1.640
Coll=J & MaxInj=Uninjured	Surf=Dry	0.049	1.000	1.577
Coll=J & MaxInj=Uninjured	Manvr=GoingAheadOther	0.049	1.000	1.367
Coll=J & SpdLim=60mph	DrvInj=Slight	0.049	1.000	2.929
Coll=J & SpdLim=60mph	Area=Rural	0.049	1.000	2.563
Coll=J & SpdLim=60mph	MaxInj=Slight	0.049	1.000	1.577
Coll=J & Prec=FailStopOther	Light=DarkSL	0.049	1.000	3.727
Coll=J & TrfCtrl=Light	Prec=FailStopOther	0.049	1.000	5.125
Coll=J & Prec=FailStopOther	Area=Urban	0.049	1.000	1.640
Coll=J & Prec=FailStopOther	Surf=Dry	0.049	1.000	1.577
Coll=J & Prec=FailStopOther	Manvr=GoingAheadOther	0.049	1.000	1.367
Coll=J & TrfCtrl=Light	Light=DarkSL	0.049	1.000	3.727
Coll=J & TrfCtrl=Light	Area=Urban	0.049	1.000	1.640
Coll=J & TrfCtrl=Light	Surf=Dry	0.049	1.000	1.577
Coll=J & TrfCtrl=Light	Manvr=GoingAheadOther	0.049	1.000	1.367

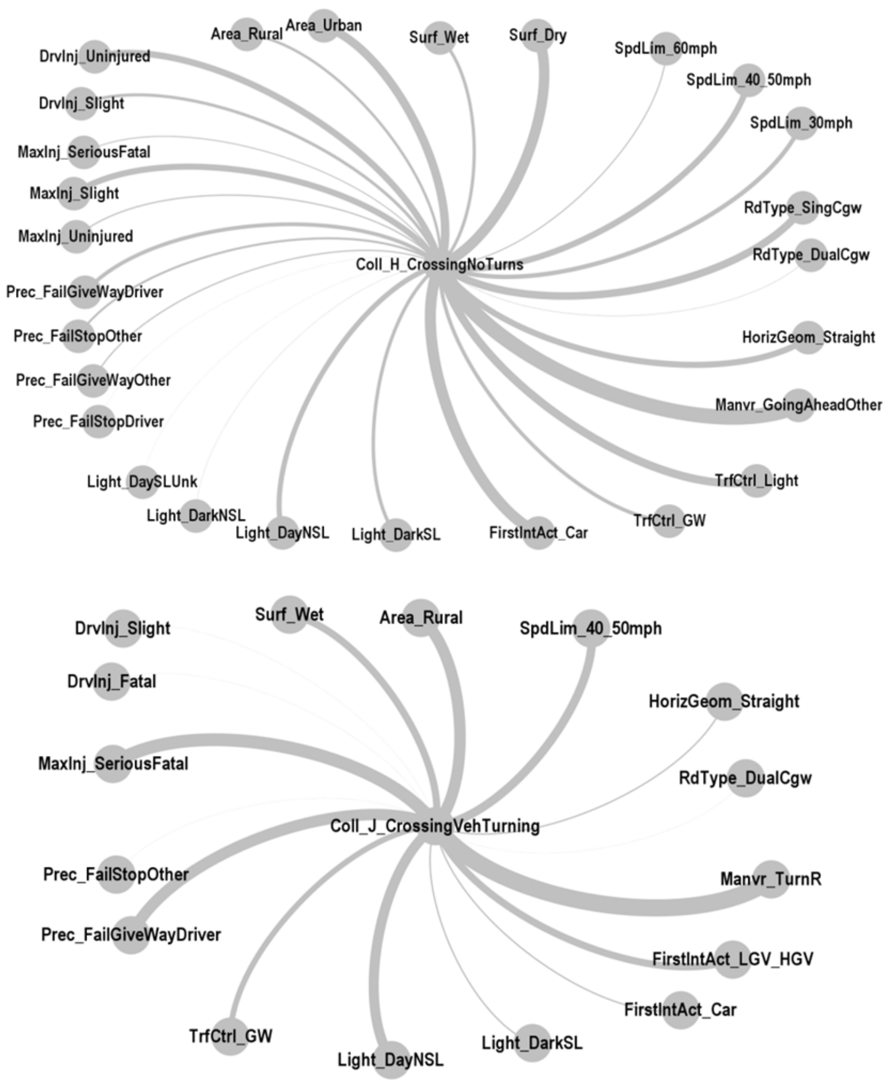


Figure 143: Weighted, directed graphs obtained from all association rules for cluster X-C4

Cluster X-C5: “The car is going straight on a road broken by a major road, when being hit on its nearside by another vehicle.”

Table 78: All rules (up to 3 items) obtained for cluster X-C5 with collision type H (scenario X-5.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H	Manvr=GoingAheadOther	0.700	1.000	1.364
Coll=H & TrfCtrl=GW	Prec=FailGiveWayDriver	0.467	0.933	1.474
Coll=H & Prec=FailGiveWayDriver	TrfCtrl=GW	0.467	0.933	1.474
Coll=H & MaxInj=Slight	Surf=Dry	0.467	0.933	1.333
Coll=H & SpdLim=30mph	RdType=SingCgw	0.333	1.000	1.304
Coll=H & Area=Rural	DrvInj=Slight	0.300	0.818	1.888
Coll=H & DrvInj=Slight	Area=Rural	0.300	0.818	1.888
Coll=H & Area=Rural	TrfCtrl=GW	0.300	0.818	1.292
Coll=H & Area=Urban	SpdLim=30mph	0.267	0.800	1.714
Coll=H & SpdLim=30mph	TrfCtrl=GW	0.267	0.800	1.263
Coll=H & SpdLim=30mph	Prec=FailGiveWayDriver	0.267	0.800	1.263
Coll=H & DrvInj=Uninjured	Prec=FailGiveWayDriver	0.267	0.800	1.263
Coll=H & Surf=Wet	RdType=SingCgw	0.200	1.000	1.304
Coll=H & Surf=Wet	TrfCtrl=GW	0.167	0.833	1.316
Coll=H & TrfCtrl=Light	Area=Urban	0.133	1.000	1.765
Coll=H & MaxInj=Uninjured	Surf=Wet	0.100	1.000	3.333
Coll=H & MaxInj=Uninjured	SpdLim=30mph	0.100	1.000	2.143
Coll=H & MaxInj=Uninjured	RdType=SingCgw	0.100	1.000	1.304
Coll=H & Light=DarkSL	SpdLim=30mph	0.100	1.000	2.143
Coll=H & Light=DarkSL	RdType=SingCgw	0.100	1.000	1.304
Coll=H & RdType=DualCgw	Surf=Dry	0.100	1.000	1.429
Coll=H & RdType=DualCgw	MaxInj=Slight	0.100	1.000	1.304
Coll=H & TrfCtrl=Light	DrvInj=Uninjured	0.100	0.750	1.324
Coll=H & SpdLim=40-50mph	RdType=DualCgw	0.067	1.000	5.000
Coll=H & SpdLim=40-50mph	TrfCtrl=Light	0.067	1.000	3.750
Coll=H & SpdLim=40-50mph	Area=Urban	0.067	1.000	1.765
Coll=H & SpdLim=40-50mph	DrvInj=Uninjured	0.067	1.000	1.765
Coll=H & SpdLim=40-50mph	Surf=Dry	0.067	1.000	1.429
Coll=H & SpdLim=40-50mph	MaxInj=Slight	0.067	1.000	1.304
Coll=H & Prec=FailGiveWayOther	SpdLim=30mph	0.033	1.000	2.143
Coll=H & Prec=FailGiveWayOther	Area=Urban	0.033	1.000	1.765
Coll=H & Prec=FailGiveWayOther	Light=DayNSL	0.033	1.000	1.765
Coll=H & Prec=FailGiveWayOther	Surf=Dry	0.033	1.000	1.429
Coll=H & Light=DarkNSL	MaxInj=SeriousFatal	0.033	1.000	10.000
Coll=H & Light=DarkNSL	SpdLim=60mph	0.033	1.000	3.750
Coll=H & Light=DarkNSL	TrfCtrl=GW	0.033	1.000	1.579
Coll=H & Light=DarkNSL	Prec=FailGiveWayDriver	0.033	1.000	1.579

Table 79: All rules (up to 3 items) obtained for cluster X-C5 with collision type L (scenario X-5.2), sorted by support

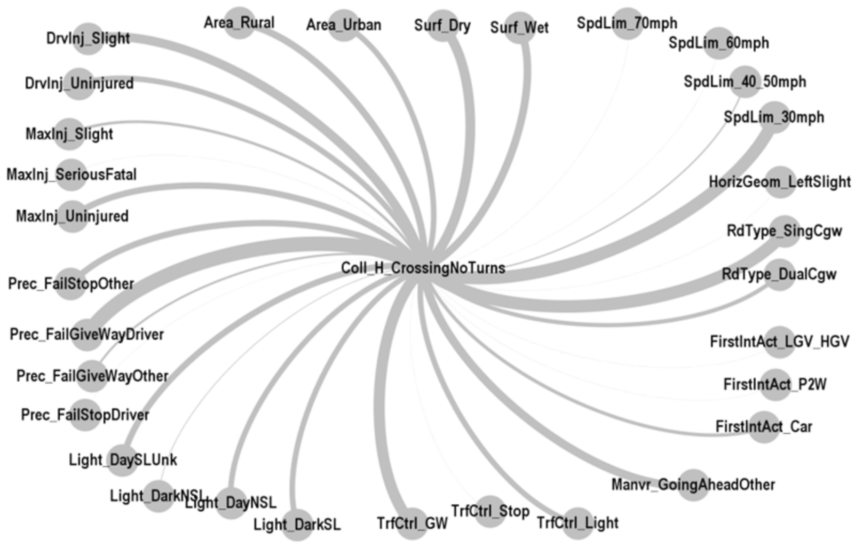
Antecedent	Consequent	Supp	Conf	Lift
Coll=L	SpdLim=40-50mph	0.100	1.000	4.286
Coll=L	TrfCtrl=Light	0.100	1.000	3.750
Coll=L & RdType=DualCgw	Manvr=TurnR	0.067	1.000	4.286
Coll=L & Manvr=TurnR	RdType=DualCgw	0.067	1.000	5.000
Coll=L & RdType=DualCgw	Area=Urban	0.067	1.000	1.765
Coll=L & Area=Urban	RdType=DualCgw	0.067	1.000	5.000
Coll=L & RdType=DualCgw	Surf=Dry	0.067	1.000	1.429
Coll=L & Surf=Dry	RdType=DualCgw	0.067	1.000	5.000
Coll=L & Manvr=TurnR	Area=Urban	0.067	1.000	1.765
Coll=L & Area=Urban	Manvr=TurnR	0.067	1.000	4.286
Coll=L & Manvr=TurnR	Surf=Dry	0.067	1.000	1.429
Coll=L & Surf=Dry	Manvr=TurnR	0.067	1.000	4.286
Coll=L & DrvInj=Slight	MaxInj=Slight	0.067	1.000	1.304
Coll=L & MaxInj=Slight	DrvInj=Slight	0.067	1.000	2.308
Coll=L & Area=Urban	Surf=Dry	0.067	1.000	1.429
Coll=L & Surf=Dry	Area=Urban	0.067	1.000	1.765
Coll=L & MaxInj=Uninjured	Prec=PoorMnvrDriver	0.033	1.000	30.000
Coll=L & Light=DaySLUnk	Prec=PoorMnvrDriver	0.033	1.000	30.000
Coll=L & DrvInj=Uninjured	Prec=PoorMnvrDriver	0.033	1.000	30.000
Coll=L & Prec=FailGiveWayOther	Light=DarkNSL	0.033	1.000	15.000
Coll=L & Light=DarkNSL	Prec=FailGiveWayOther	0.033	1.000	15.000
Coll=L & Prec=FailGiveWayOther	Surf=Wet	0.033	1.000	3.333
Coll=L & Surf=Wet	Prec=FailGiveWayOther	0.033	1.000	15.000
Coll=L & Prec=FailGiveWayOther	Area=Rural	0.033	1.000	2.308
Coll=L & Area=Rural	Prec=FailGiveWayOther	0.033	1.000	15.000
Coll=L & Manvr=GoingAheadOther	Prec=FailGiveWayOther	0.033	1.000	15.000
Coll=L & RdType=SingCgw	Prec=FailGiveWayOther	0.033	1.000	15.000
Coll=L & Surf=Wet	Light=DarkNSL	0.033	1.000	15.000
Coll=L & Area=Rural	Light=DarkNSL	0.033	1.000	15.000
Coll=L & Manvr=GoingAheadOther	Light=DarkNSL	0.033	1.000	15.000
Coll=L & RdType=SingCgw	Light=DarkNSL	0.033	1.000	15.000
Coll=L & Light=DarkNSL	MaxInj=Slight	0.033	1.000	1.304
Coll=L & MaxInj=Uninjured	RdType=DualCgw	0.033	1.000	5.000
Coll=L & MaxInj=Uninjured	Light=DaySLUnk	0.033	1.000	5.000
Coll=L & Light=DaySLUnk	MaxInj=Uninjured	0.033	1.000	7.500
Coll=L & MaxInj=Uninjured	Manvr=TurnR	0.033	1.000	4.286
Coll=L & MaxInj=Uninjured	Area=Urban	0.033	1.000	1.765
Coll=L & DrvInj=Uninjured	MaxInj=Uninjured	0.033	1.000	7.500
Coll=L & MaxInj=Uninjured	Surf=Dry	0.033	1.000	1.429
Coll=L & Light=DarkSL	RdType=DualCgw	0.033	1.000	5.000
Coll=L & Light=DarkSL	Manvr=TurnR	0.033	1.000	4.286
Coll=L & Light=DarkSL	DrvInj=Slight	0.033	1.000	2.308
Coll=L & Light=DarkSL	Area=Urban	0.033	1.000	1.765

Coll=L & Light=DarkSL	Prec=FailGiveWayDriver	0.033	1.000	1.579
Coll=L & Prec=FailGiveWayDriver	Light=DarkSL	0.033	1.000	6.000
Coll=L & Light=DarkSL	Surf=Dry	0.033	1.000	1.429
Coll=L & Light=DarkSL	MaxInj=Slight	0.033	1.000	1.304
Coll=L & Light=DaySLUnk	RdType=DualCgw	0.033	1.000	5.000
Coll=L & DrvInj=Uninjured	RdType=DualCgw	0.033	1.000	5.000
Coll=L & Prec=FailGiveWayDriver	RdType=DualCgw	0.033	1.000	5.000
Coll=L & Light=DaySLUnk	Manvr=TurnR	0.033	1.000	4.286
Coll=L & Light=DaySLUnk	Area=Urban	0.033	1.000	1.765
Coll=L & Light=DaySLUnk	DrvInj=Uninjured	0.033	1.000	1.765
Coll=L & DrvInj=Uninjured	Light=DaySLUnk	0.033	1.000	5.000
Coll=L & Light=DaySLUnk	Surf=Dry	0.033	1.000	1.429
Coll=L & DrvInj=Uninjured	Manvr=TurnR	0.033	1.000	4.286
Coll=L & Prec=FailGiveWayDriver	Manvr=TurnR	0.033	1.000	4.286
Coll=L & Surf=Wet	Area=Rural	0.033	1.000	2.308
Coll=L & Area=Rural	Surf=Wet	0.033	1.000	3.333
Coll=L & Surf=Wet	DrvInj=Slight	0.033	1.000	2.308
Coll=L & Surf=Wet	Manvr=GoingAheadOther	0.033	1.000	1.364
Coll=L & Manvr=GoingAheadOther	Surf=Wet	0.033	1.000	3.333
Coll=L & Surf=Wet	RdType=SingCgw	0.033	1.000	1.304
Coll=L & RdType=SingCgw	Surf=Wet	0.033	1.000	3.333
Coll=L & Surf=Wet	MaxInj=Slight	0.033	1.000	1.304
Coll=L & Area=Rural	DrvInj=Slight	0.033	1.000	2.308
Coll=L & Area=Rural	Manvr=GoingAheadOther	0.033	1.000	1.364
Coll=L & Manvr=GoingAheadOther	Area=Rural	0.033	1.000	2.308
Coll=L & Area=Rural	RdType=SingCgw	0.033	1.000	1.304
Coll=L & RdType=SingCgw	Area=Rural	0.033	1.000	2.308
Coll=L & Area=Rural	MaxInj=Slight	0.033	1.000	1.304
Coll=L & Prec=FailGiveWayDriver	DrvInj=Slight	0.033	1.000	2.308
Coll=L & Manvr=GoingAheadOther	DrvInj=Slight	0.033	1.000	2.308
Coll=L & RdType=SingCgw	DrvInj=Slight	0.033	1.000	2.308
Coll=L & DrvInj=Uninjured	Area=Urban	0.033	1.000	1.765
Coll=L & Prec=FailGiveWayDriver	Area=Urban	0.033	1.000	1.765
Coll=L & DrvInj=Uninjured	Surf=Dry	0.033	1.000	1.429
Coll=L & Prec=FailGiveWayDriver	Surf=Dry	0.033	1.000	1.429
Coll=L & Prec=FailGiveWayDriver	MaxInj=Slight	0.033	1.000	1.304
Coll=L & Manvr=GoingAheadOther	RdType=SingCgw	0.033	1.000	1.304
Coll=L & RdType=SingCgw	Manvr=GoingAheadOther	0.033	1.000	1.364
Coll=L & Manvr=GoingAheadOther	MaxInj=Slight	0.033	1.000	1.304
Coll=L & RdType=SingCgw	MaxInj=Slight	0.033	1.000	1.304

Table 80: All rules (up to 3 items) obtained for cluster X-C5 with collision type M (scenario X-5.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=M	DrvInj=Uninjured	0.133	1.000	1.765
Coll=M	Light=DayNSL	0.133	1.000	1.765

Coll=M	MaxInj=Slight	0.133	1.000	1.304
Coll=M	Manvr=TurnR	0.100	0.750	3.214
Coll=M	Area=Urban	0.100	0.750	1.324
Coll=M & Manvr=TurnR	Prec=FailGiveWayDriver	0.100	1.000	1.579
Coll=M & Prec=FailGiveWayDriver	Manvr=TurnR	0.100	1.000	4.286
Coll=M & FirstIntAct=Car	Manvr=TurnR	0.100	1.000	4.286
Coll=M & Area=Urban	Surf=Dry	0.100	1.000	1.429
Coll=M & Surf=Dry	Area=Urban	0.100	1.000	1.765
Coll=M & TrfCtrl=GW	RdType=SingCgw	0.100	1.000	1.304
Coll=M & RdType=SingCgw	TrfCtrl=GW	0.100	1.000	1.579
Coll=M & FirstIntAct=Car	Prec=FailGiveWayDriver	0.100	1.000	1.579
Coll=M & SpdLim=40-50mph	Manvr=TurnR	0.067	1.000	4.286
Coll=M & SpdLim=40-50mph	TrfCtrl=GW	0.067	1.000	1.579
Coll=M & SpdLim=40-50mph	Prec=FailGiveWayDriver	0.067	1.000	1.579
Coll=M & SpdLim=40-50mph	RdType=SingCgw	0.067	1.000	1.304
Coll=M & SpdLim=30mph	Area=Urban	0.067	1.000	1.765
Coll=M & SpdLim=30mph	Surf=Dry	0.067	1.000	1.429
Coll=M & Surf=Wet	SpdLim=40-50mph	0.033	1.000	4.286
Coll=M & Area=Rural	SpdLim=40-50mph	0.033	1.000	4.286
Coll=M & Surf=Wet	Manvr=TurnR	0.033	1.000	4.286
Coll=M & Area=Rural	Manvr=TurnR	0.033	1.000	4.286
Coll=M & Surf=Wet	Area=Rural	0.033	1.000	2.308
Coll=M & Area=Rural	Surf=Wet	0.033	1.000	3.333
Coll=M & Surf=Wet	TrfCtrl=GW	0.033	1.000	1.579
Coll=M & Surf=Wet	Prec=FailGiveWayDriver	0.033	1.000	1.579
Coll=M & Surf=Wet	RdType=SingCgw	0.033	1.000	1.304
Coll=M & Area=Rural	TrfCtrl=GW	0.033	1.000	1.579
Coll=M & Area=Rural	Prec=FailGiveWayDriver	0.033	1.000	1.579
Coll=M & Area=Rural	RdType=SingCgw	0.033	1.000	1.304



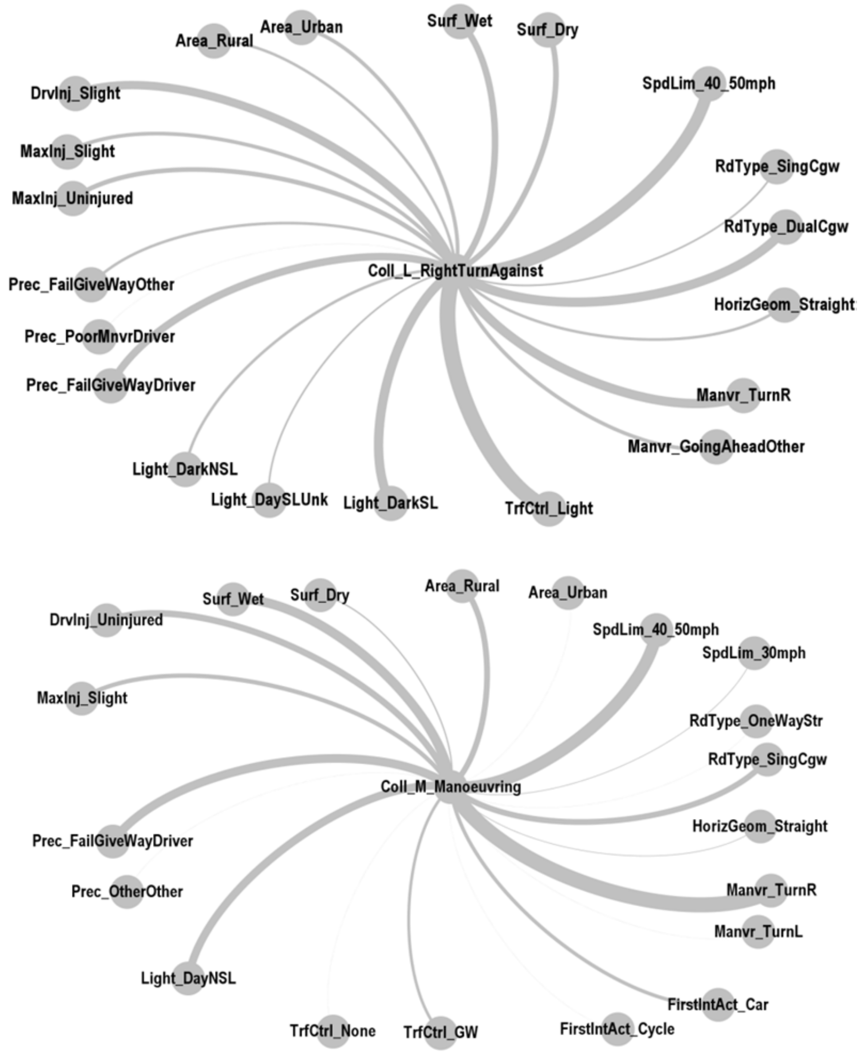


Figure 144: Weighted, directed graphs obtained from all association rules for cluster X-C5

Cluster X-C6: "The car is going straight over a junction with minor roads joining from the left and right, when being hit on its offside by another car or goods vehicle."

Table 81: All rules (up to 3 items) obtained for cluster X-C6 with collision type H (scenario X-6.1), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=H & FirstIntAct=Car	Light=DayNSL	0.345	0.769	1.394
Coll=H & RdType=SingCgw	Prec=FailGiveWayOther	0.310	0.818	1.318
Coll=H & TrfCtrl=None	Prec=FailGiveWayOther	0.276	0.889	1.432
Coll=H & Prec=FailGiveWayOther	TrfCtrl=None	0.276	0.800	1.450
Coll=H & Light=DayNSL	Prec=FailGiveWayOther	0.276	0.800	1.289
Coll=H & Prec=FailGiveWayOther	Light=DayNSL	0.276	0.800	1.450
Coll=H & Area=Urban	SpdLim=30mph	0.276	1.000	1.611
Coll=H & DrvInj=Slight	Prec=FailGiveWayOther	0.276	0.800	1.289
Coll=H & MaxInj=Slight	Prec=FailGiveWayOther	0.276	0.800	1.289
Coll=H & Surf=Wet	SpdLim=30mph	0.241	0.875	1.410
Coll=H & SpdLim=30mph	Surf=Wet	0.241	0.778	1.880
Coll=H & Area=Urban	Surf=Wet	0.241	0.875	2.115
Coll=H & Area=Rural	TrfCtrl=None	0.172	0.833	1.510
Coll=H & Area=Rural	Light=DayNSL	0.172	0.833	1.510
Coll=H & Area=Rural	Surf=Dry	0.172	0.833	1.422
Coll=H & Surf=Dry	Area=Rural	0.172	0.833	3.021
Coll=H & Surf=Dry	TrfCtrl=None	0.172	0.833	1.510
Coll=H & Surf=Dry	Light=DayNSL	0.172	0.833	1.510
Coll=H & Surf=Dry	Prec=FailGiveWayOther	0.172	0.833	1.343
Coll=H & TrfCtrl=Light	Surf=Wet	0.138	0.800	1.933
Coll=H & TrfCtrl=Light	Light=DayNSL	0.138	0.800	1.450
Coll=H & TrfCtrl=Light	SpdLim=30mph	0.138	0.800	1.289
Coll=H & Prec=FailStopOther	Surf=Wet	0.103	1.000	2.417
Coll=H & Prec=FailStopOther	TrfCtrl=Light	0.103	1.000	2.231
Coll=H & MaxInj=SeriousFatal	Light=DayNSL	0.103	1.000	1.813
Coll=H & DrvInj=Uninjured	SpdLim=30mph	0.103	1.000	1.611
Coll=H & RdType=DualCgw	SpdLim=70mph	0.069	1.000	14.500
Coll=H & SpdLim=60mph	TrfCtrl=None	0.069	1.000	1.813
Coll=H & SpdLim=60mph	Light=DayNSL	0.069	1.000	1.813
Coll=H & HorizGeom=RightSlight	Area=Rural	0.069	1.000	3.625
Coll=H & HorizGeom=RightSlight	MaxInj=Slight	0.069	1.000	1.318
Coll=H & RdType=DualCgw	Area=Rural	0.069	1.000	3.625
Coll=H & RdType=DualCgw	Light=DayNSL	0.069	1.000	1.813
Coll=H & Light=DaySLUnk	Surf=Wet	0.069	1.000	2.417
Coll=H & Light=DaySLUnk	TrfCtrl=None	0.069	1.000	1.813
Coll=H & Light=DaySLUnk	SpdLim=30mph	0.069	1.000	1.611
Coll=H & Light=DaySLUnk	Prec=FailGiveWayOther	0.069	1.000	1.611
Coll=H & Light=DaySLUnk	Area=Urban	0.069	1.000	1.381
Coll=H & Light=DaySLUnk	DrvInj=Slight	0.069	1.000	1.318

Coll=H & Light=DaySLUnk	MaxInj=Slight	0.069	1.000	1.318
Coll=H & Light=DaySLUnk	RdType=SingCgw	0.069	1.000	1.261
Coll=H & Light=DarkSL	Prec=OtherDriver	0.034	1.000	29.000
Coll=H & SpdLim=40-50mph	HorizGeom=Right	0.034	1.000	29.000
Coll=H & FirstIntAct=LGV=HGV	Light=DarkNSL	0.034	1.000	29.000
Coll=H & FirstIntAct=LGV=HGV	RdType=OneWayStr	0.034	1.000	29.000
Coll=H & Manvr=TurnR	Prec=FailStopOther	0.034	1.000	7.250
Coll=H & Manvr=TurnR	MaxInj=SeriousFatal	0.034	1.000	7.250
Coll=H & Manvr=TurnR	Surf=Wet	0.034	1.000	2.417
Coll=H & Manvr=TurnR	SpdLim=30mph	0.034	1.000	1.611
Coll=H & Manvr=TurnR	RdType=SingCgw	0.034	1.000	1.261
Coll=H & FirstIntAct=LGV=HGV	Prec=FailStopOther	0.034	1.000	7.250
Coll=H & FirstIntAct=LGV=HGV	Surf=Wet	0.034	1.000	2.417
Coll=H & Light=DarkSL	HorizGeom=RightSlight	0.034	1.000	9.667
Coll=H & Light=DarkSL	DrvInj=Uninjured	0.034	1.000	4.833
Coll=H & Light=DarkSL	Area=Rural	0.034	1.000	3.625
Coll=H & Light=DarkSL	TrfCtrl=None	0.034	1.000	1.813
Coll=H & Light=DarkSL	Surf=Dry	0.034	1.000	1.706
Coll=H & Light=DarkSL	MaxInj=Slight	0.034	1.000	1.318
Coll=H & MaxInj=Uninjured	Surf=Wet	0.034	1.000	2.417
Coll=H & MaxInj=Uninjured	TrfCtrl=None	0.034	1.000	1.813
Coll=H & MaxInj=Uninjured	Light=DayNSL	0.034	1.000	1.813
Coll=H & MaxInj=Uninjured	SpdLim=30mph	0.034	1.000	1.611
Coll=H & MaxInj=Uninjured	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=H & SpdLim=40-50mph	MaxInj=SeriousFatal	0.034	1.000	7.250
Coll=H & SpdLim=40-50mph	Area=Rural	0.034	1.000	3.625
Coll=H & SpdLim=40-50mph	TrfCtrl=None	0.034	1.000	1.813
Coll=H & SpdLim=40-50mph	Light=DayNSL	0.034	1.000	1.813
Coll=H & SpdLim=40-50mph	Surf=Dry	0.034	1.000	1.706
Coll=H & SpdLim=40-50mph	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=H & SpdLim=40-50mph	DrvInj=Slight	0.034	1.000	1.318
Coll=H & SpdLim=40-50mph	RdType=SingCgw	0.034	1.000	1.261

Table 82: All rules (up to 3 items) obtained for cluster X-C6 with collision type L (scenario X-6.2), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=L & FirstIntAct=Car	DrvInj=Slight	0.138	1.000	1.318
Coll=L & Light=DaySLUnk	DrvInj=Slight	0.103	1.000	1.318
Coll=L & DrvInj=Slight	Light=DaySLUnk	0.103	0.750	2.417
Coll=L & FirstIntAct=Car	Light=DaySLUnk	0.103	0.750	2.417
Coll=L & Surf=Wet	SpdLim=30mph	0.103	1.000	1.611
Coll=L & SpdLim=30mph	Surf=Wet	0.103	1.000	2.417
Coll=L & Surf=Wet	Area=Urban	0.103	1.000	1.381
Coll=L & Area=Urban	Surf=Wet	0.103	0.750	1.813
Coll=L & SpdLim=30mph	Area=Urban	0.103	1.000	1.381
Coll=L & MaxInj=Slight	DrvInj=Slight	0.103	1.000	1.318

Coll=L & Light=DarkSL	Surf=Wet	0.069	1.000	2.417
Coll=L & Light=DarkSL	TrfCtrl=Light	0.069	1.000	2.231
Coll=L & Light=DarkSL	Area=Urban	0.069	1.000	1.381
Coll=L & Surf=Dry	Light=DaySLUnk	0.069	1.000	3.222
Coll=L & Surf=Dry	TrfCtrl=Light	0.069	1.000	2.231
Coll=L & Surf=Dry	DrvInj=Slight	0.069	1.000	1.318
Coll=L & MaxInj=SeriousFatal	Prec=LossCntrOther	0.034	1.000	29.000
Coll=L & SpdLim=40-50mph	Prec=LossCntrOther	0.034	1.000	29.000
Coll=L & FirstIntAct=Other	Light=DarkSL	0.034	1.000	9.667
Coll=L & FirstIntAct=Other	MaxInj=Uninjured	0.034	1.000	9.667
Coll=L & MaxInj=Uninjured	FirstIntAct=Other	0.034	1.000	14.500
Coll=L & FirstIntAct=Other	DrvInj=Uninjured	0.034	1.000	4.833
Coll=L & DrvInj=Uninjured	FirstIntAct=Other	0.034	1.000	14.500
Coll=L & FirstIntAct=Other	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & FirstIntAct=Other	SpdLim=30mph	0.034	1.000	1.611
Coll=L & FirstIntAct=Other	Area=Urban	0.034	1.000	1.381
Coll=L & Area=Rural	SpdLim=60mph	0.034	1.000	9.667
Coll=L & SpdLim=60mph	Light=DaySLUnk	0.034	1.000	3.222
Coll=L & SpdLim=60mph	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & MaxInj=Uninjured	Light=DarkSL	0.034	1.000	9.667
Coll=L & DrvInj=Uninjured	Light=DarkSL	0.034	1.000	9.667
Coll=L & DrvInj=Uninjured	MaxInj=Uninjured	0.034	1.000	9.667
Coll=L & MaxInj=Uninjured	Surf=Wet	0.034	1.000	2.417
Coll=L & MaxInj=Uninjured	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & MaxInj=Uninjured	SpdLim=30mph	0.034	1.000	1.611
Coll=L & MaxInj=Uninjured	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=L & MaxInj=SeriousFatal	SpdLim=40-50mph	0.034	1.000	4.833
Coll=L & SpdLim=40-50mph	MaxInj=SeriousFatal	0.034	1.000	7.250
Coll=L & MaxInj=SeriousFatal	Light=DaySLUnk	0.034	1.000	3.222
Coll=L & MaxInj=SeriousFatal	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & MaxInj=SeriousFatal	Surf=Dry	0.034	1.000	1.706
Coll=L & MaxInj=SeriousFatal	Area=Urban	0.034	1.000	1.381
Coll=L & MaxInj=SeriousFatal	DrvInj=Slight	0.034	1.000	1.318
Coll=L & DrvInj=Uninjured	Surf=Wet	0.034	1.000	2.417
Coll=L & DrvInj=Uninjured	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & DrvInj=Uninjured	SpdLim=30mph	0.034	1.000	1.611
Coll=L & DrvInj=Uninjured	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=L & DrvInj=Uninjured	Area=Urban	0.034	1.000	1.381
Coll=L & SpdLim=40-50mph	Light=DaySLUnk	0.034	1.000	3.222
Coll=L & SpdLim=40-50mph	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & SpdLim=40-50mph	Surf=Dry	0.034	1.000	1.706
Coll=L & SpdLim=40-50mph	Area=Urban	0.034	1.000	1.381
Coll=L & SpdLim=40-50mph	DrvInj=Slight	0.034	1.000	1.318
Coll=L & Area=Rural	Light=DaySLUnk	0.034	1.000	3.222
Coll=L & Area=Rural	TrfCtrl=Light	0.034	1.000	2.231
Coll=L & Area=Rural	Surf=Dry	0.034	1.000	1.706

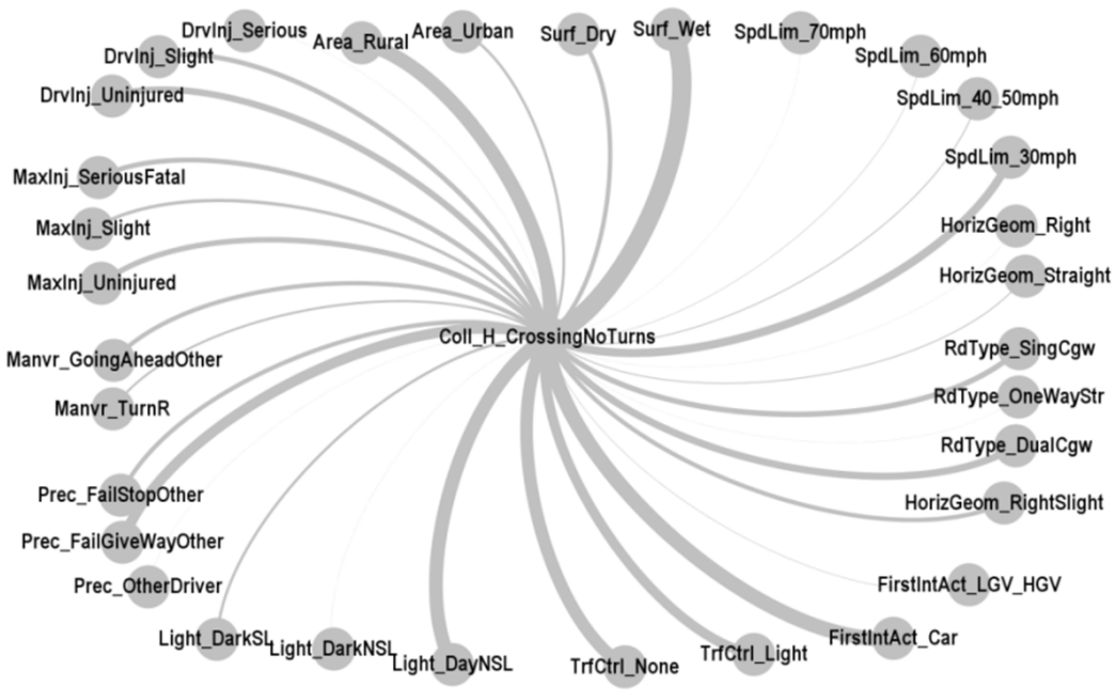
Coll=L & Area=Rural	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=L & Area=Rural	DrvInj=Slight	0.034	1.000	1.318
Coll=L & Area=Rural	MaxInj=Slight	0.034	1.000	1.318
Coll=L & TrfCtrl=None	Light=DaySLUnk	0.034	1.000	3.222
Coll=L & TrfCtrl=None	Surf=Wet	0.034	1.000	2.417
Coll=L & TrfCtrl=None	SpdLim=30mph	0.034	1.000	1.611
Coll=L & TrfCtrl=None	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=L & TrfCtrl=None	Area=Urban	0.034	1.000	1.381
Coll=L & TrfCtrl=None	DrvInj=Slight	0.034	1.000	1.318
Coll=L & TrfCtrl=None	MaxInj=Slight	0.034	1.000	1.318

Table 83: All rules (up to 3 items) obtained for cluster X-C6 with collision type M (scenario X-6.3), sorted by support

Antecedent	Consequent	Supp	Conf	Lift
Coll=M & Surf=Dry	Area=Urban	0.172	1.000	1.381
Coll=M & Area=Urban	Surf=Dry	0.172	1.000	1.706
Coll=M & Surf=Dry	RdType=SingCgw	0.172	1.000	1.261
Coll=M & RdType=SingCgw	Surf=Dry	0.172	1.000	1.706
Coll=M & FirstIntAct=Car	Surf=Dry	0.172	1.000	1.706
Coll=M & Area=Urban	RdType=SingCgw	0.172	1.000	1.261
Coll=M & RdType=SingCgw	Area=Urban	0.172	1.000	1.381
Coll=M & FirstIntAct=Car	Area=Urban	0.172	1.000	1.381
Coll=M & DrvInj=Slight	MaxInj=Slight	0.172	1.000	1.318
Coll=M & MaxInj=Slight	DrvInj=Slight	0.172	1.000	1.318
Coll=M & Manvr=GoingAheadOther	DrvInj=Slight	0.172	1.000	1.318
Coll=M & HorizGeom=Straight	DrvInj=Slight	0.172	1.000	1.318
Coll=M & Manvr=GoingAheadOther	MaxInj=Slight	0.172	1.000	1.318
Coll=M & HorizGeom=Straight	MaxInj=Slight	0.172	1.000	1.318
Coll=M & FirstIntAct=Car	RdType=SingCgw	0.172	1.000	1.261
Coll=M & Light=DayNSL	Prec=FailGiveWayOther	0.138	1.000	1.611
Coll=M & Prec=FailGiveWayOther	Light=DayNSL	0.138	1.000	1.813
Coll=M & Light=DayNSL	DrvInj=Slight	0.138	1.000	1.318
Coll=M & DrvInj=Slight	Light=DayNSL	0.138	0.800	1.450
Coll=M & Light=DayNSL	MaxInj=Slight	0.138	1.000	1.318
Coll=M & MaxInj=Slight	Light=DayNSL	0.138	0.800	1.450
Coll=M & Manvr=GoingAheadOther	Light=DayNSL	0.138	0.800	1.450
Coll=M & HorizGeom=Straight	Light=DayNSL	0.138	0.800	1.450
Coll=M & Prec=FailGiveWayOther	DrvInj=Slight	0.138	1.000	1.318
Coll=M & DrvInj=Slight	Prec=FailGiveWayOther	0.138	0.800	1.289
Coll=M & Prec=FailGiveWayOther	MaxInj=Slight	0.138	1.000	1.318
Coll=M & MaxInj=Slight	Prec=FailGiveWayOther	0.138	0.800	1.289
Coll=M & Manvr=GoingAheadOther	Prec=FailGiveWayOther	0.138	0.800	1.289
Coll=M & HorizGeom=Straight	Prec=FailGiveWayOther	0.138	0.800	1.289
Coll=M & SpdLim=30mph	Surf=Dry	0.103	1.000	1.706
Coll=M & SpdLim=30mph	Area=Urban	0.103	1.000	1.381
Coll=M & SpdLim=30mph	DrvInj=Slight	0.103	1.000	1.318

Coll=M	& SpdLim=30mph	MaxInj=Slight	0.103	1.000	1.318
Coll=M	& SpdLim=30mph	RdType=SingCgw	0.103	1.000	1.261
Coll=M	& Light=DaySLUnk	Surf=Dry	0.069	1.000	1.706
Coll=M	& Light=DaySLUnk	Area=Urban	0.069	1.000	1.381
Coll=M	& Light=DaySLUnk	RdType=SingCgw	0.069	1.000	1.261
Coll=M	& Prec=PoorMnvrDriver	Manvr=Other	0.034	1.000	29.000
Coll=M	& HorizGeom=RightSlight	Manvr=Other	0.034	1.000	29.000
Coll=M	& MaxInj=Uninjured	Manvr=Other	0.034	1.000	29.000
Coll=M	& DrvInj=Uninjured	Manvr=Other	0.034	1.000	29.000
Coll=M	& FirstIntAct=Other	RdType=DualCgw	0.034	1.000	5.800
Coll=M	& RdType=DualCgw	FirstIntAct=Other	0.034	1.000	14.500
Coll=M	& FirstIntAct=Other	SpdLim=40-50mph	0.034	1.000	4.833
Coll=M	& FirstIntAct=Other	Area=Rural	0.034	1.000	3.625
Coll=M	& Area=Rural	FirstIntAct=Other	0.034	1.000	14.500
Coll=M	& Surf=Wet	FirstIntAct=Other	0.034	1.000	14.500
Coll=M	& FirstIntAct=Other	Light=DayNSL	0.034	1.000	1.813
Coll=M	& FirstIntAct=Other	DrvInj=Slight	0.034	1.000	1.318
Coll=M	& FirstIntAct=Other	MaxInj=Slight	0.034	1.000	1.318
Coll=M	& Prec=PoorMnvrDriver	HorizGeom=RightSlight	0.034	1.000	9.667
Coll=M	& HorizGeom=RightSlight	Prec=PoorMnvrDriver	0.034	1.000	14.500
Coll=M	& Prec=PoorMnvrDriver	MaxInj=Uninjured	0.034	1.000	9.667
Coll=M	& MaxInj=Uninjured	Prec=PoorMnvrDriver	0.034	1.000	14.500
Coll=M	& Prec=PoorMnvrDriver	DrvInj=Uninjured	0.034	1.000	4.833
Coll=M	& DrvInj=Uninjured	Prec=PoorMnvrDriver	0.034	1.000	14.500
Coll=M	& Prec=PoorMnvrDriver	Light=DaySLUnk	0.034	1.000	3.222
Coll=M	& Prec=PoorMnvrDriver	RdType=SingCgw	0.034	1.000	1.261
Coll=M	& HorizGeom=RightSlight	MaxInj=Uninjured	0.034	1.000	9.667
Coll=M	& MaxInj=Uninjured	HorizGeom=RightSlight	0.034	1.000	9.667
Coll=M	& HorizGeom=RightSlight	DrvInj=Uninjured	0.034	1.000	4.833
Coll=M	& DrvInj=Uninjured	HorizGeom=RightSlight	0.034	1.000	9.667
Coll=M	& HorizGeom=RightSlight	SpdLim=40-50mph	0.034	1.000	4.833
Coll=M	& HorizGeom=RightSlight	Light=DaySLUnk	0.034	1.000	3.222
Coll=M	& HorizGeom=RightSlight	Area=Urban	0.034	1.000	1.381
Coll=M	& HorizGeom=RightSlight	RdType=SingCgw	0.034	1.000	1.261
Coll=M	& DrvInj=Uninjured	MaxInj=Uninjured	0.034	1.000	9.667
Coll=M	& MaxInj=Uninjured	SpdLim=40-50mph	0.034	1.000	4.833
Coll=M	& MaxInj=Uninjured	Light=DaySLUnk	0.034	1.000	3.222
Coll=M	& MaxInj=Uninjured	Surf=Dry	0.034	1.000	1.706
Coll=M	& Prec=FailStopOther	Light=DaySLUnk	0.034	1.000	3.222
Coll=M	& Prec=FailStopOther	Surf=Dry	0.034	1.000	1.706
Coll=M	& Prec=FailStopOther	SpdLim=30mph	0.034	1.000	1.611
Coll=M	& Prec=FailStopOther	Area=Urban	0.034	1.000	1.381
Coll=M	& Prec=FailStopOther	DrvInj=Slight	0.034	1.000	1.318
Coll=M	& Prec=FailStopOther	MaxInj=Slight	0.034	1.000	1.318
Coll=M	& Prec=FailStopOther	RdType=SingCgw	0.034	1.000	1.261
Coll=M	& RdType=DualCgw	SpdLim=40-50mph	0.034	1.000	4.833

Coll=M & RdType=DualCgw	Area=Rural	0.034	1.000	3.625
Coll=M & Area=Rural	RdType=DualCgw	0.034	1.000	5.800
Coll=M & RdType=DualCgw	Surf=Wet	0.034	1.000	2.417
Coll=M & Surf=Wet	RdType=DualCgw	0.034	1.000	5.800
Coll=M & RdType=DualCgw	Light=DayNSL	0.034	1.000	1.813
Coll=M & RdType=DualCgw	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=M & RdType=DualCgw	DrvInj=Slight	0.034	1.000	1.318
Coll=M & RdType=DualCgw	MaxInj=Slight	0.034	1.000	1.318
Coll=M & DrvInj=Uninjured	SpdLim=40-50mph	0.034	1.000	4.833
Coll=M & DrvInj=Uninjured	Light=DaySLUnk	0.034	1.000	3.222
Coll=M & DrvInj=Uninjured	Surf=Dry	0.034	1.000	1.706
Coll=M & DrvInj=Uninjured	Area=Urban	0.034	1.000	1.381
Coll=M & Area=Rural	SpdLim=40-50mph	0.034	1.000	4.833
Coll=M & Surf=Wet	SpdLim=40-50mph	0.034	1.000	4.833
Coll=M & Area=Rural	Surf=Wet	0.034	1.000	2.417
Coll=M & Surf=Wet	Area=Rural	0.034	1.000	3.625
Coll=M & Area=Rural	Light=DayNSL	0.034	1.000	1.813
Coll=M & Area=Rural	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=M & Area=Rural	DrvInj=Slight	0.034	1.000	1.318
Coll=M & Area=Rural	MaxInj=Slight	0.034	1.000	1.318
Coll=M & Surf=Wet	Light=DayNSL	0.034	1.000	1.813
Coll=M & Surf=Wet	Prec=FailGiveWayOther	0.034	1.000	1.611
Coll=M & Surf=Wet	DrvInj=Slight	0.034	1.000	1.318
Coll=M & Surf=Wet	MaxInj=Slight	0.034	1.000	1.318



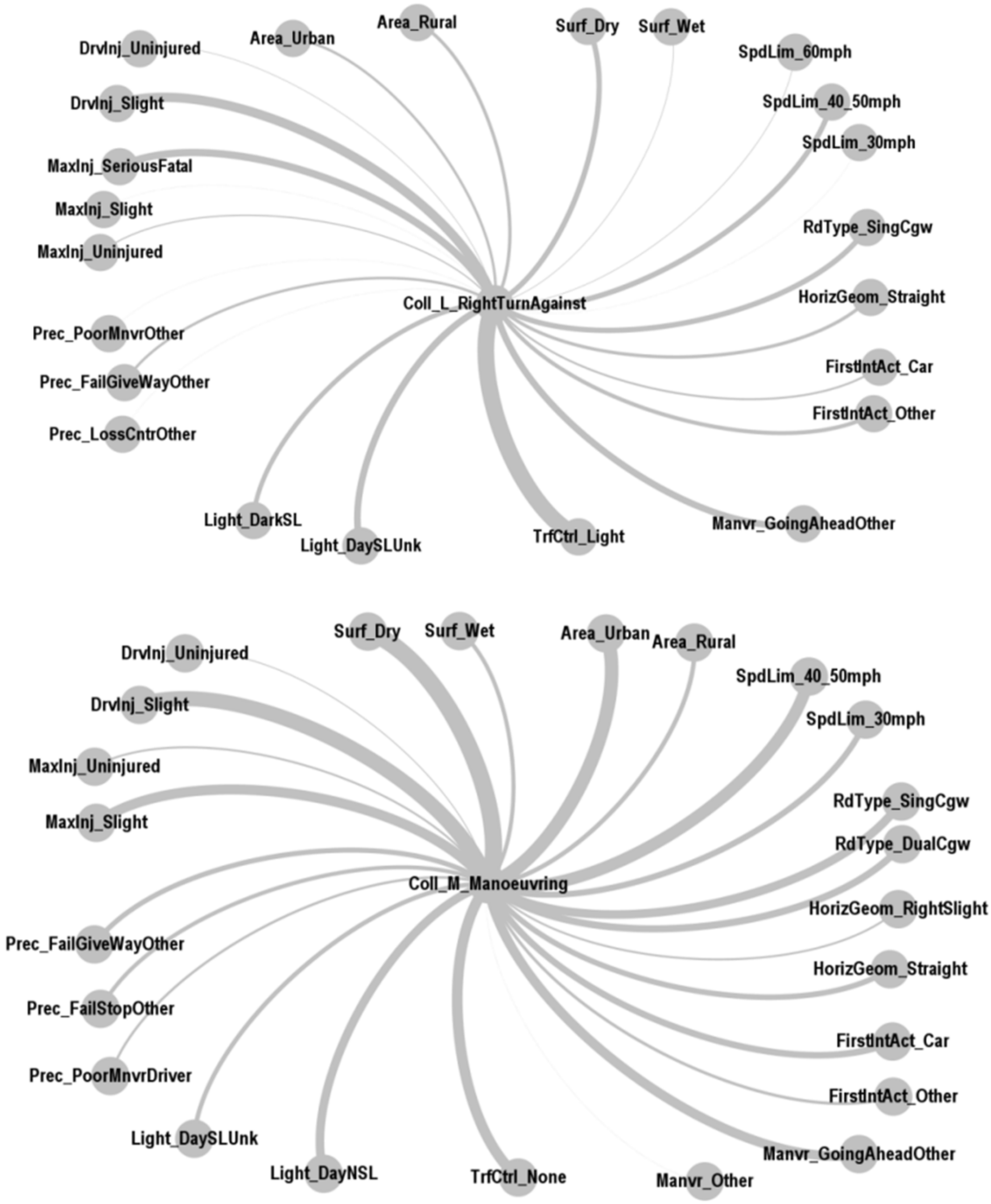
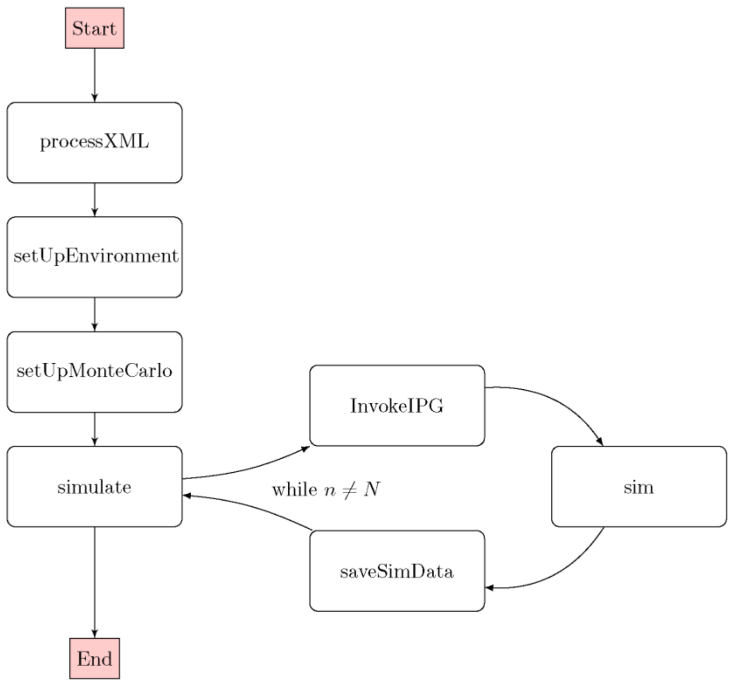
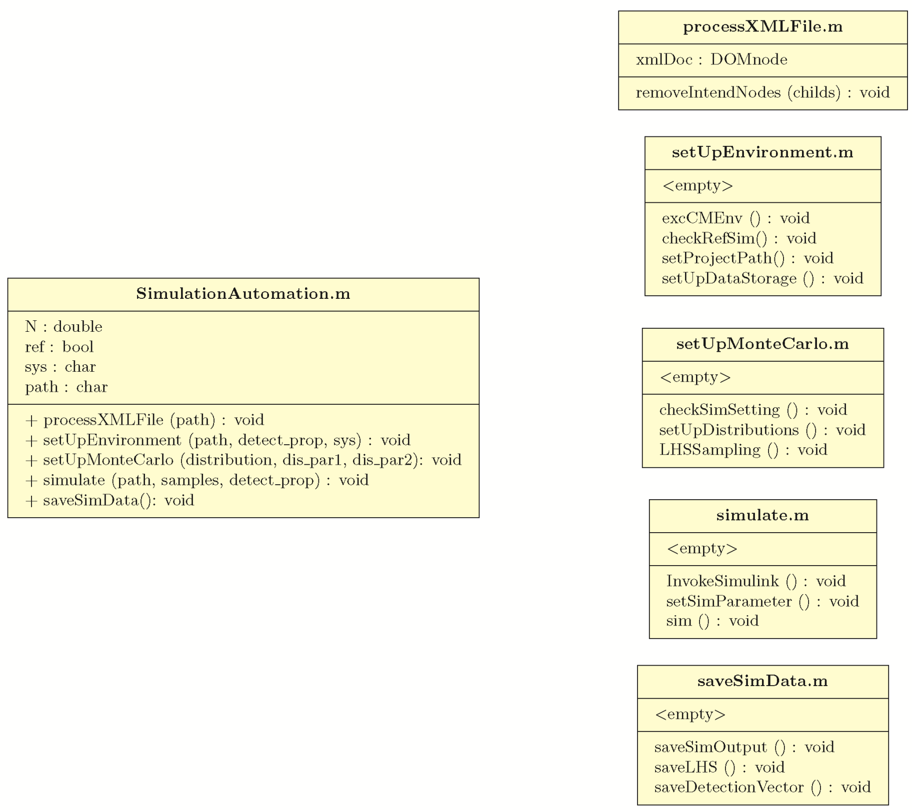


Figure 145: Weighted, directed graphs obtained from all association rules for cluster X-C6

Appendix G: Simulation automation in MATLAB

The following two graphs show the simulation automation script as flowchart (top) and class diagram (bottom), which were realised in MATLAB.





Appendix H: Publications

Three publications are attached on the following pages, which were published during the development of this thesis:

1. Nitsche, P., Mocanu, I., Reinthaler, M., 2014a. Requirements on Tomorrow‘s Road Infrastructure for Highly Automated Driving, in: The 3rd International Conference on Connected Vehicles & Expo (ICCVE 2014). Vienna, Austria.
2. Mocanu, I., Nitsche, P., Saleh, P., 2015. Highly automated driving and its requirements on road planning and design, in: Proceedings of the 25th PIARC World Road Congress. Seoul, Korea.
3. Nitsche, P., Thomas, P., Stuetz, R., Welsh, R., 2017. Pre-crash scenarios at road junctions: A clustering method for car crash data. Accident Analysis & Prevention 107, 137–151.

Requirements on Tomorrow's Road Infrastructure for Highly Automated Driving

Philippe Nitsche, Isabela Mocanu and Martin Reinthaler
Mobility Department, AIT Austrian Institute of Technology
Vienna, Austria

***Abstract*—This paper presents the results of a study on the requirements on road infrastructure regarding increased use of highly automated vehicles. Based on the outcome of a literature review and a web questionnaire, factors that most influence the performance of automated driving systems are given. Requirements for future road design and planning are recommended to ensure a safe and efficient operation.**

Keywords—automated driving, road infrastructure, lane assistance, collision avoidance, speed control

I. Introduction

Automation of road vehicles has gained an increased presence on the agendas of companies and public authorities. Two major factors are paramount towards the successful future implementation of Automated Driving Systems (ADS): legal framework and technical reliability (cf. [1]). The latter involves ensuring full functionality under varying road infrastructure and transnational differences. This paper investigates the interplay between road infrastructure and automated vehicles, with the aim of defining requirements for safe and efficient future road traffic. The study answers questions such as: What are the most influencing factors to be considered in future road design for optimum functionality of ADS? What are the measures concerning road infrastructure to enable an accelerated deployment of self-driving vehicles?

II. Study Design

A. Literature review

A comprehensive literature search of 30 studies was conducted to identify current ADS and sensor technologies. By reviewing developments of the automotive and sensor industry, as well as research projects, conference and journal papers (e.g. [2]–[5]), an extensive description of the most promising ADS was compiled, along with the infrastructure-related factors that may influence their performance. The focus was on partial, conditional and highly automated driving, i.e. automation level 2-4 as given in [6], where the driver can be out of the loop, but may take back control. The literature review resulted in three distinct groups of ADS, which differ in their purpose of use, technologies applied and market readiness, namely 1) lane assistance, 2) collision avoidance and 3) speed control systems.

B. Web survey

The literature review was complemented by a web survey, in which 54 international experts from R&D (67%), academia (24%), automotive or supplier industry (15%) and other fields such as public authorities, consultants etc. (13%) responded to specific questions concerning ADS. Those included the role of road infrastructure, market readiness as well as to which extent certain factors influence the performance of the three respective ADS groups. For the latter one, respondents gave a rating from very low to very high regarding the influence. Based on the number of responses, each factor could be assigned a lowest and highest possible score. Within this range, results are indicated as 'low', 'med' and 'high' in Fig. 1-3.

III. Results and Derived Requirements

The role of the road infrastructure was rated as "very important" by 76% of the respondents. Only about 2% believes that the infrastructure is not important at all. The following subsections summarize the responses on the question "To which extent do the following factors influence the performance of [the respective ADS group]?" and give recommendations on infrastructure design and planning.

A. Lane assistance systems

This group of systems assists drivers by keeping or changing the lane automatically. It is also capable of performing overtaking or lane merging maneuvers through automatic steering and/or braking. The systems are primarily based on vision sensors, but also employ accurate 3D road map data as well as radar technology.

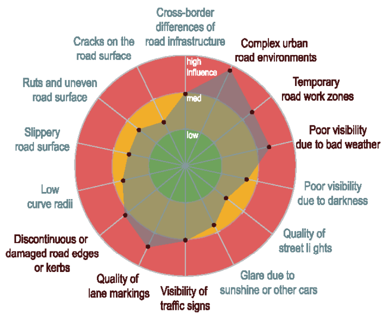


Fig. 1. Factors influencing the performance of lane assistance systems

Based on the ratings in Fig. 1, the most important road infrastructure requirements were derived as follows:

- Clear, unobstructed and functional traffic speed signs
- Consistent placement of traffic signs
- Cross-border harmonized characteristics of traffic signs in terms of color, shape, font
- Infrastructure-based warning systems for bad weather and poor visibility

- Roadside V2I/I2V communication infrastructure (via wireless G5 or cellular networks)
- Lane marking requirements as above, in case of automated braking/swerving maneuvers

B. Collision avoidance systems

This bundle of systems helps avoid collisions with other traffic participants and obstacles on the road by means of automated steering and/or braking. Radar, Lidar and/or vision sensors in the vehicles are employed for obstacle detection. In addition, cooperative traffic systems (wireless communication between vehicles, infrastructure or vulnerable road users) enable collision warnings in complex traffic environments or obscured intersections.

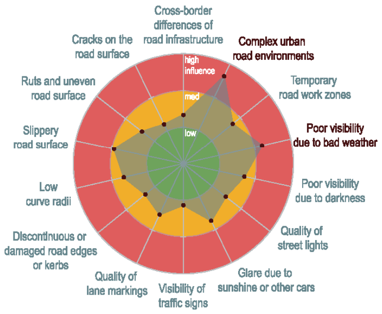


Fig. 2. Factors influencing the performance of collision avoidance systems

The following requirements were derived from the influences depicted in Fig. 2:

- Pedestrian and bicyclist protection and shielding at (urban) intersections
- Infrastructure-based warning systems for bad weather and poor visibility
- Road surface with a sufficient friction coefficient to allow emergency maneuvers
- Presence of wireless communication beacons at certain locations, if required
- Lane marking requirements as above, in case of automated maneuvers, i.e. swerving, braking

C. Speed control systems

This group of systems adapts the vehicle's speed dynamically, based on the legal speed limit or external factors such as road alignment, weather conditions, congestion, road works etc. Cooperative infrastructure to vehicle (I2V) technology is employed as well as vision sensors capable of identifying and processing traffic speed signs. Furthermore, accurate map data with speed limits are crucial. The results of the rating for speed control systems are given in Fig. 3. Hence, infrastructure requirements were identified as follows:

- Clear, unobstructed and functional traffic speed signs
- Consistent placement of traffic signs
- Cross-border harmonized characteristics of traffic signs in terms of color, shape and font
- Infrastructure-based warning systems for bad weather and poor visibility

- Roadside V2I/I2V communication infrastructure (via wireless G5 or cellular networks)
- Lane marking requirements as above, in case of automated braking/swerving maneuvers

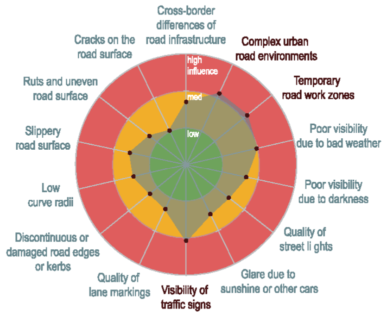


Fig. 3. Factors influencing the performance of speed control systems

IV. Conclusion and Discussion

According to the study results, the main infrastructure challenges for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather condition. For lane assistance, the visibility and quality of lane markings are of particular importance, while speed control systems require clearly visible traffic speed signs. Road surface deficiencies such as uneven surface or cracks are rated as the least influencing factors. These findings help to plan and design (future) road networks to be prepared for automated road vehicles. Furthermore, ADS testing procedures can be focused on the factors identified in this study.

It is to be mentioned that the three groups analyzed in this study do not cover the whole range of subsystems in automated transport. For example, automated parking systems, vehicle platooning, routing systems or car2car communication systems are also available or under development, respectively. Besides legal prerequisites, precise geolocation/map data and robust driver state monitoring to hand over control, connectivity between vehicles, road users and infrastructure as well as optimal interplay between automated and non-automated vehicles are major enablers for a safe and efficient operation of automated transport.

References

[1] T. M. Gasser, 'Herausforderungen automatischen Fahrens und Forschungsschwerpunkte', presented at the 6. Tagung Fahrassistenz, Munich, Germany, 2013.

[2] P. Larsson et al., 'Test and evaluation plans', Ford Research & Advanced Engineering, Germany, interactIVe deliverable 7.4, Nov. 2012.

[3] E. Jakobsson et al., 'Function description and requirements', Continental Automotive GmbH, Germany, HAVEit deliverable 11.1, Sep. 2008.

[4] C. Lipski, B. Scholz, K. Berger, C. Linz, T. Stich, and M. Magnor, 'A Fast and Robust Approach to Lane Marking Detection and Lane Tracking', in IEEE Southwest Symposium on Image Analysis and Interpretation, 2008. SSIAI 2008, 2008, pp. 57–60.

[5] C. G. Keller et al., 'Active Pedestrian Safety by Automatic Braking and Evasive Steering', IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1292–1304, Dec. 2011.

[6] On-road Automated Vehicle Standards Committee, 'SAE J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems'. SAE International, 16-Jan-2014.

HIGHLY AUTOMATED DRIVING AND ITS REQUIREMENTS ON ROAD PLANNING AND DESIGN

Isabela Mocanu, Philippe Nitsche, Peter Saleh
Department of Mobility, AIT Austrian Institute of Technology, Austria
Isabela.mocanu@ait.ac.at, Philippe.nitsche@ait.ac.at, Peter.saleh@ait.ac.at

ABSTRACT

This paper presents the results of a study on road infrastructure requirements in the context of increased use of highly automated vehicles. An investigation of the interplay between road infrastructure and automated vehicles was performed, with the aim of defining requirements for future road design and planning towards safe and efficient road traffic. Based on a comprehensive review of literature, three subgroups of automated driving systems have been selected for further analysis, which differ in their purpose of use, technologies applied and market readiness: 1) Lane assistance, 2) Collision avoidance and 3) Speed control systems. Results from an online expert survey conducted for this work indicate that the main challenges for the introduction of highly automated vehicles are complex urban environments, temporary work zones and poor visibility due to bad weather condition. Following a list of requirements for road engineering derived from those results, the paper is concluded with a discussion about future needs for amendments in road planning and design regulations.

1. INTRODUCTION

Current developments and ambitions for the future of the road transport system are now working towards autonomous driving. Over the past few years automation of road vehicles has gained an increasing presence on the agendas of companies and public authorities. OEMs and their suppliers, as well as the EC [1] have started to push Automated Driving Systems (ADS) into the forefront of research.

Assisted driving means driving enhanced by dedicated control and consists of in-vehicle advanced driving assistance systems (ADAS) ranging from emergency brake assistance to stability and traction control that support the driver, while still being in control. In contrast, fully autonomous driving means the vehicle controls itself completely, without the need for a human driver. These two extremes represent level 1 and level 5 respectively of the five levels of automation, as introduced by the Society of Automotive Engineers (SAE) [2].

Although autonomous vehicle technology is complex, its elements can be divided into four basic component categories: sensors, mapping, perception, and communication [3]. Sensor technology includes a variety of such hardware as multiple video cameras for daylight conditions, forward-looking infrared sensors for night conditions and detection of humans and animals, radar for measuring range and velocity, global positioning systems (GPS) to determine location, accelerometers and gyroscopes to detect changes in speed and direction, and light detection and ranging (LIDAR) that employs spinning lasers and photoreceptors to create a three-dimensional model of the immediate environment.

Multiple factors are involved when discussing the successful implementation of Automated Driving Systems (ADS) at different automation levels but a major factor is technical reliability. The reliability of these systems strongly depends on their optimal functionality under varying road infrastructure and transnational differences.
This paper investigates the relation between road infrastructure and automated vehicles, with the aim of defining the requirements for the safe planning and design of future roads. The study attempts to answer two main research questions: 1) What are the most influencing factors to be considered in future road design for optimum functionality of Automated Driving Systems (ADS)? 2) What are the features concerning road infrastructure to enable an accelerated deployment of self-driving vehicles?

To this end, a literature review of current automated driving systems available or in research at the moment was performed, which yielded a list of systems that were later grouped based on their similarity in purpose. The review then looked at each system, in terms of technology and sensors employed. Based on each system's specifications, a list of factors that could influence the performance of the systems was compiled. The review was complemented by performing a web-survey, in which 54 experts responded to specific questions regarding ADS. The questions included the role of infrastructure for ADS performance, technological readiness of ADS, as well as the level of influence that certain infrastructure factors can have on the performance of ADS.

The paper is organized as follows: Section 2 describes the current state of play in the field of automated driving systems. In Section 3 the results of the literature review are presented along with descriptions of the automated driving subsystems selected for further investigation within the survey presented in Section 4. The requirements on road planning and design are described in Section 5, while overall conclusions are drawn in Section 6.

2. BACKGROUND/ STATE OF PLAY

Automated Driving Systems can be seen as a further stage of driving assistance systems. Automated parking, automatic emergency braking, adaptive cruise control and lane keeping assist are examples of technologies that have leapt into the market in the past few years. The American National Highway Traffic Safety Administration (NHTSA), Germany's Bundesanstalt für Straßenwesen (BASt) and the Society of Automotive Engineers (SAE) have introduced certain levels of automation, which differ in the extent of human driver involvement. The European industry has agreed to use the SAE levels from zero to five as common understanding of automated driving (see Figure 1).

SAE Level	SAE Name	SAE Narrative Definition	Execution of Steering/ Acceleration/ Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System capability (driving modes)	BASt Level	NTHSA Level
	Human Driver monitors the driving environment							
0	No Automation	the full-time performance by *the human driver* of all aspects of the *dynamic driving task*	Human Driver	Human Driver	Human Driver	N/A	Driver only	0
1	Driver Assistance	the *driving mode-specific* execution by a driver assistance system of either steering or acceleration/deceleration	Human Driver and Systems	Human Driver	Human Driver	Some Driving Modes	Assisted	1
2	Partial Automation	Part-time or driving mode-dependent execution by **one or more driver assistance systems** of both steering and acceleration/deceleration. Human driver performs all other aspects of the *dynamic driving task*.	**System**	Human Driver	Human Driver	Some Driving Modes	Partially Automated	2
	Automated driving system ("system") monitors the driving environment							
3	Conditional Automation	*driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task* - *human driver* **does** respond appropriately to a *request to intervene*	System	**System**	Human Driver	Some Driving Modes	Highly Automated	3
4	High Automation	*driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task* - *human driver* **does not** respond appropriately to a *request to intervene*	System	System	**System**	Some Driving Modes	Fully Automated	3/4
5	Full Automation	**full-time** performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	**System**	Some Driving Modes		

Figure 1: Levels of vehicle automation

There is a clear shift from level 2 (partial automation) to level 3 (conditional automation), namely that the vehicle system takes over the monitoring of the driving environment. While in level 3 the driver still has to take over vehicle control appropriately if requested, level 4 (high automation) does not require the driver to respond in time, at least for a certain amount of time or travel route. Promising near-future systems of level 4 are the highway pilot or parking garage pilot. Level 5 (full automation) is the only level where the human driver is fully kept out of the loop from source to destination, or might not even present in the car. The latter level is often called "robot car", e.g. particularly interesting for car sharing companies to relocate their fleet.

While level 1 and 2 systems are already on the market (e.g. adaptive cruise control, stop &go functionality, lane keeping assist), little market research can be found about the timescale of introducing level 3 to 5 on European roads. However, experts estimate fully automated cars to be ready not before 2025, although the area of use must be considered. For example on highways, highly automated vehicles are expected earlier than on urban roads.

Clearly, the technical reliability of ADS depends on ensuring full functionality under varying road infrastructure and transnational differences, as well as guaranteeing a safe interplay with non-autonomous vehicles and vulnerable road users. Therefore, standardized procedures for evaluating ADS are highly relevant. At the moment, there is no set of European regulations on validating the reliability of highly automated road vehicles. While some states in the US have laid the foundation for ADS testing on public roads and already allowed it for a certain amount of vehicles [4]–[10] the EU still struggles with harmonization and regulatory issues. The Vienna Convention [11] constituted a deployment barrier for level 3 to 5, because its Article 8 had originally specified that the driver must maintain permanent control of the vehicle.

In 2014, this limitation was amended in order to allow automated vehicles on public roads as long as they can be overridden or deactivated [12], which established the legal foundation for the automation levels 3 and 4. The UNECE regulation No. 79 [13] states, that whenever automated steering becomes operational, this shall be automatically disabled, if the vehicle speed exceeds 10 km/h by more than 20 percent. This is certainly a hurdle which would require another amendment.

Many prototypes of automated vehicles have been developed and demonstrated in recent years, revealing technical challenges for automotive and sensor developers. To tackle those challenges, there are technical development trends regarding sensing, perception and trajectory planning, as summarized in [14]. They include vision-based algorithms for object recognition, Simultaneous Localization and Mapping (SLAM) [15] and sophisticated path and motion planning. Keeping these technologies in mind, the next sections focus on the challenges from a road engineering point of view.

3. SCOPING CURRENT ADS DEVELOPMENTS

In the scope of this paper, a comprehensive literature search was conducted to identify current ADS developments on the market, still under research or in the prototyping phase. The focus was on partial, conditional and highly automated driving, where the driver can be out of the loop, but may take back control. By reviewing advancements of the automotive and sensor industry (e.g. [14]–[19]), as well as research projects (e.g. [18], [21]–[23]), conference and journal papers (e.g. [24]–[28]), an extensive list of ADS was compiled. The underlying assumption of this paper is that ADS comprise a set of subsystems for different tasks of vehicle control and environment sensing. Ultimately, the decision was to focus on three distinct groups of automated driving systems, which differ in their purpose of use, technologies applied and market readiness, namely 1) lane assistance, 2) collision avoidance and 3) speed control systems.
It is significant to mention that the three groups analysed in this study do not cover the whole range of subsystems in automated transport. For example, automated parking systems, truck platooning, re-routing or automated docking systems for public transport are also available or under development.

Lane assistance systems aid drivers by performing actions such as keeping or changing the lane automatically. These systems have also the capability of performing overtaking or lane merging, through automatic steering and/or braking. This group is composed of:

- Lane keeping assistant
- Lane changing assistant
- Lane merging assistant
- Overtaking assistant

Table 1 presents the technologies and sensors that this group of systems can employ. These systems are primarily based on vision sensors (i.e. cameras) which detect various road markings such as for lanes, edges of the road, road works, etc. [28], [29]. However, short range or long range radar can also be used to detect the environment around the vehicle and has a higher accuracy for measurements [24], [30], [31]. Other technologies employed include inertial sensors, GNSS and V2V (vehicle to vehicle) communication [21]. From these findings, infrastructure-related factors that could influence the performance of the sensors and therefore of lane assistance systems were derived, as shown in Table 2:

Table 1: Technologies employed

Technologies employed
Vision sensors
Short/Long range radar
Inertial sensors
Automated braking and steering
I2V communication
Digital map

Table 2 Infrastructure factors

Infrastructure-related factors
Visibility of lane markings (size, colour, age, retro-reflectivity, etc. [30]
Road surface conditions (i.e. existence of cracks, potholes, etc.)
Road edges and road delineation
Road works zones markings

Collision avoidance systems aid drivers to avoid collisions with other traffic participants (e.g. vehicles, pedestrians, cyclists) or other obstacles by means of automated steering and /or braking. The group is composed of:
- Obstacle collision avoidance including vehicles and other objects
- Vulnerable road users collision avoidance

Table 3 presents the technologies and sensors employed by this ADS group. Cooperative communication, either vehicle to infrastructure (V2I) or infrastructure to vehicle (I2V) plays an important role, as for example, in the detection and warning of approaching road users [25]. Vision sensors, as well as radar or lidar (i.e. long range radar) can also detect obstacles on the road [31], [32]. Automated steering and braking are crucial in this ADS group, as these technologies can help reduce the severity of an accident or even prevent it [16], [18]. Based on these considerations, several infrastructure elements can influence the performance of the employed technologies, as shown in Table 4.

Table 3 Technologies employed

Technologies employed
Radar/Lidar
V2I and I2V communication
Inertial sensors
Automated braking and steering
Vision sensors
Digital map

Table 4 Infrastructure factors

Infrastructure-related factors
Beacons (for infrastructure to vehicle or vehicle to infrastructure communication)
Road surface conditions (i.e. existence of cracks, potholes, etc.)
Road edges and road delineation
Visibility of lane markings (size, colour, age, retro-reflectivity, etc.

Speed control systems aid drivers by adapting the vehicle’s speed automatically, based on legal speed limits or other external factors, such as road alignment, weather conditions, congestion, road works, etc. [33], [34]. The group is consisted of:
- Intelligent Speed Adaptation (ISA) – active
- Dynamic speed adaptation
- Curve speed control

Table 5 presents the technologies and sensors employed by speed control systems. Cooperative I2V technology is employed, as well as vision sensors capable of identifying and processing traffic speed signs. Moreover, accurate map data with speed limits are crucial [20]. Once more, automated steering and braking are relevant for this ADS group. Based on these findings, Table 6 presents the infrastructure elements that can influence the performance of this ADS group.

Table 5 Technologies employed

Technologies employed
Digital map
V2I and I2V communication
Vision sensors
Automated braking and steering

Table 6 Infrastructure factors

Infrastructure-related factors
Beacons (for infrastructure to vehicle /vehicle to infrastructure communication)
Visibility of traffic signs (colour, age, size, etc.)
Road edges and road delineation
Visibility of lane markings (size, colour, age, retro-reflectivity, etc.

4. ONLINE EXPERT SURVEY ON ADS PERFORMANCE

The literature review described above was complemented by a web survey, in which 54 international experts from R&D (67%), academia (24%), automotive or supplier industry (15%) and other fields such as public authorities, consultants etc. (13%) responded to specific questions concerning ADS. Those included the role of road infrastructure, market readiness as well as to which extent certain factors influence the performance of selected ADS groups. The role of the road infrastructure was rated as “very important” by 76% of the respondents. Only about 2% believes that the infrastructure is not important at all.

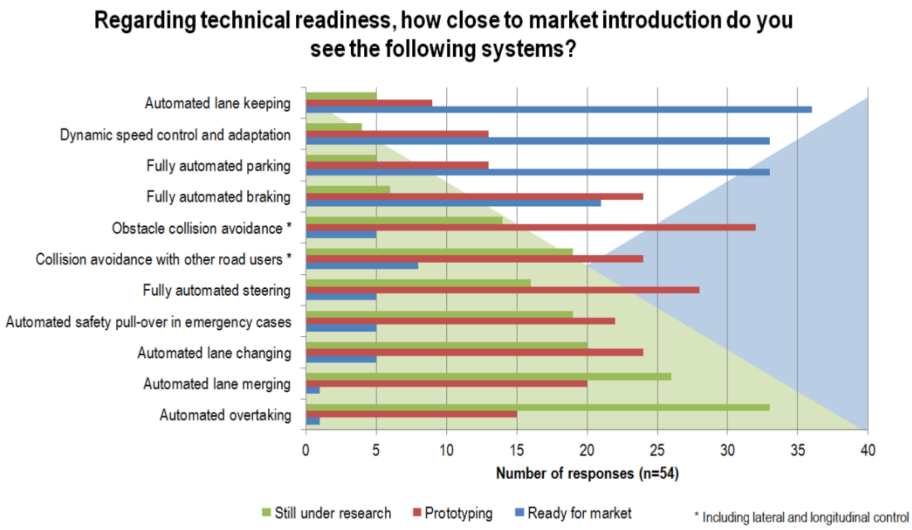


Figure 2 Number of online survey responses regarding the technical readiness of ADS

In terms of technological readiness for market (see Figure 2**Fehler! Verweisquelle konnte nicht gefunden werden.**), the systems were rated as still under research, prototyping and ready for market. Responses indicate that lane keeping, speed adaptation and parking systems are in an advanced stage of deployment, while assistance systems that must handle more complex maneuvers are still under research. It is important to note that the rating is based on technological readiness in the year 2014, disregarding legal, liability and privacy issues.

To the question regarding to which extent certain factors influence the performance of specific ADS groups, the respondents gave a rating from very high to very low regarding the influence of a factor, e.g. poor visibility. In order to present the rating on a linear scale, the following weighting function has been applied:

$$Weighted\ influence\ of\ a\ factor = \frac{I - N}{3N}$$

with N as the total number of responses for each factor, and $I = \sum_{i=1}^{4}(n_i \cdot i)$, with n as the number of responses for each rating category i from 1 (very low) to 4 (very high influence).

Resulting in a scale from 0 to 1, the influence weights were simply divided into three equal ranges indicating 'low', 'medium' and 'high influence' for each automated driving subsystem (see Figure 3 - Figure 5).

Figure 3: Factors influencing the performance of lane assistance systems

Considering the fact that lane assistance systems (see Figure 3) are mainly based on vision sensors, it is not surprising that poor visibility and low quality of lane markings and road edges are among the factors with a high influence. Complex urban environments were rated with the highest influence, meaning that lane assistance systems might struggle with obstructed intersections, parked vehicles or crossing pedestrians and cyclists. Temporary work zones may constitute a problem for lane assistance of ADS, since lane markings are often replaced by temporary markings.

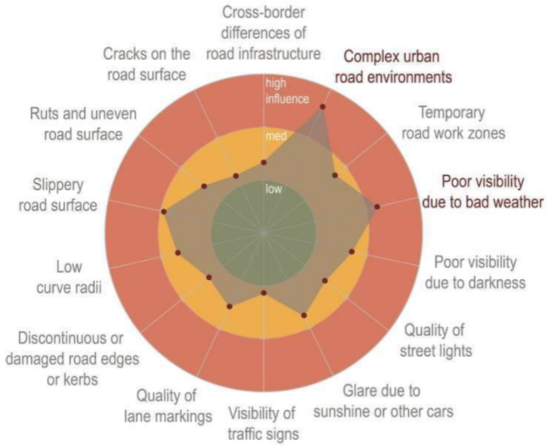


Figure 4: Factors influencing the performance of collision avoidance systems

As seen in Figure 4, the ratings for collision avoidance systems indicate that visibility due to bad weather is still an issue, but not the visibility and quality of road infrastructure and equipment. For this subsystem of ADS, the interaction between road users comes into play, which could obviously be a challenge in complex urban environments.

Figure 5: Factors influencing the performance of speed control systems

Speed control systems are mainly based on traffic sign detection and/or digital maps, which are influenced the most by changing environments, such as it happens on temporary work zones or urban roads and intersections (see Figure 5).

Summarizing the web survey results, the main challenges for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather conditions. Road surface characteristics, road alignment and lighting were rated as minor influencing factors.

5. REQUIREMENTS ON ROAD PLANNING AND DESIGN

Requirements for road specifications (road quality, road condition, safety equipment, forgiving roadsides, road signs, markings, etc.) are changing due to the shift from manual driven vehicles to autonomous cars. Accommodating self-driving vehicles will pose many challenges for those who plan and design transportation facilities. Based on the results of the literature research, as well as on the results of the online survey, a comprehensive list of infrastructure requirements for automated driving systems, with respect to road planning and design was derived as follows. Note that some of the following requirements could be solved by the preference of using vehicle localization and motion planning assisted by enhanced digital road maps. However, it is assumed that not all road networks will be fully covered in geographical databases.

- Clear visibility of lane markings, irrespective of weather and road conditions;

Lane markings need to be clearly identifiable by automated driving systems and their vision sensors. For lane assistance systems, the quality of lane markings is crucial. Lane markings should be detected by the employed sensors irrespective of weather conditions (e.g. wet road, etc.) or road conditions (e.g. presence of cracks, potholes).

- Harmonised characteristics of lane markings at a transnational level, in terms of colour, luminance, shape, etc.

As ADAS becomes state of the art and ADS increases in market presence, there is a clear need for harmonisation at transnational level. For example, different vision sensors are tailored to scan and to process different characteristics of lane markings. Therefore, a common design across Europe – machine readable – would satisfy this requirement.

- Proper road works markings, distinguishable but in correspondence with normal lane markings

Road works represent an every-day operation of road transport and influence traffic. However temporary lane markings at road work sites may pose a challenge for automated driving systems, therefore it is important that these temporary markings are harmonious and in correspondence with normal lane markings.

- Well removed old lane marking remnants

In correspondence with the requirement above, ADS need to detect and process the correct lane markings in order to make the appropriate decisions. Hence, old lane marking remnants should be well removed, so they do not create any confusion.

- Defined and visible road delineation and continuous road edges or kerbs

Road delineation and edges are an important influencing factor that can influence the reliability of ADS. For example, in case of an obstacle on the road, the vehicle needs to make either an emergency brake or swerving and therefore needs to optimally detect the edges of the road, in order to make a safe-conscious decision.

- Roadside V2I/I2V communication infrastructure (via wireless G5 or cellular networks)

Cooperative communication between vehicles and infrastructure is becoming more and more important for the successful deployment of automated driving systems. Installation of relevant ITS-technologies to deploy connected driving will become part of a state-of-the-art road planning and design.

- Infrastructure based warning systems for poor weather conditions and visibility

Poor weather conditions such as fog, rain, ice can disrupt every-day traffic, as well as influence the sensors' functionality that ADS employ. Therefore, warning systems based on roadside weather stations that communicate though I2V cooperative technologies represent a requirement towards deployment of ADS.

- Pedestrian and bicyclist protection and shielding at (urban) intersections

Complex urban environments, such as intersections with multiple crossing paths and sight obstructions due to parked vehicles and buildings represent a challenge for current ADS. The detection of all possible contextual environments surrounding an intersection is still in the future, therefore pedestrian and bicyclist protection and shielding in intersections is highly relevant.

- Clear, unobstructed, well placed and functional traffic speed signs

As with lane markings, traffic signs need to be clearly identifiable by automated driving systems and their sensors. For speed control systems, the quality and visibility of traffic signs is crucial, if speed information is not available through I2V communication. Therefore, special attention should be paid not only on the quality and age of traffic signs, but also to their clear visibility, irrespective of obtrusive elements (e.g. trees, bushes, etc.).

- Cross-border harmonized characteristics of traffic signs in terms of color, shape and font

In order to enable cross-border functionality of ADS, transnational differences constitute a challenge for ADS sensors. E.g. similar to lane markings, there is a clear need for harmonisation of traffic signs and a common design in terms of colour, shape, font, etc.

6. CONCLUSIONS AND DISCUSSION ON FURTHER IMPLICATIONS FOR THE TRANSPORTATION ENGINEERING PROFESSION

This study investigated the interplay between automated vehicles and road infrastructure, with the aim of defining the requirements for the safe planning and design of future roads.
It is significant to mention that the ADS groups analysed in this study do not cover the whole range of sub-systems in automated transport. However, automated steering and braking technology is essential for the majority of systems, as it provides the

main functionality of automated control. The optimal performance of the sensors that they employ highly depends on the infrastructure factors identified in this study.

Apart from these considerations, there is the need to look beyond the current situation. According to oncoming vehicle trends, road engineers need to be aware about potential changes of common used road planning rules. Listed below are a series of examples that could not only pose challenges to road planners but also significantly influence revisions and adaptations of road planning directives and norms.

The formation of "road trains" of vehicles, commonly named vehicle platoons, electronically linked to a lead vehicle with a professional driver is currently being tested in Europe and offers the potential for safe hands-off driving on limited-access highways [35]. Special lanes may be needed for platooning, and additional space may be required for vehicles joining and separating from platoons.

Parking-assist systems are already being offered as options for some cars. This technology is likely to be enhanced to allow drivers and passengers to be dropped off and picked up while the vehicles self-park and un-park. Curb frontage loading areas may need to be expanded, and parking areas can be redesigned to be more compact and perhaps located farther away from the buildings they serve.

Regarding lane widths and pavement design, automated vehicles are expected to track more precisely within lanes, which could allow lanes to be narrowed. More precise tracking also could lead to wear patterns that may require changes to pavement design and maintenance.

Future research efforts should concentrate on cost-benefit ratios as baseline for investment decisions with regard to road engineering towards automated driving. As ADS will increase their presence in road transport, they will pose economic challenges for road operators and planners concerning the costs of investments, deployment and maintenance. However, new business models need to be investigated as well as their effect on the traditional road planning and design.

REFERENCES

[1] J. Jääskeläinen, "Automated Driving from the European Perspective," presented at the Workshop on Automation in Road Transport, Brüssel, 03-Jul-2013.

[2] On-road Automated Vehicle Standards Committee, "SAE J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems." SAE International, 16-Jan-2014.

[3] J. M. Lutin, A. L. Kornhauser, and E. Lerner-Lam, "The Revolutionary Development of Self-Driving Vehicles and Implications for the Transportation Engineering Profession," *ITE Journal*, vol. 83, no. 7, pp. 28–32, Jul. 2013.

[4] B. W. Smith, "Automated vehicles are probably legal in the United States," *1 Tex. A&M L. Rev. 411*, 2014.

[5] T. H. Kean Jr. and N. J. Sacco, *SENATE, No. 734 - State of New Jersey - 216th Legislature*. 2014.

[6] Council of the District of Columbia, *B19-0931 - Autonomous Vehicle Act of 2012*. 2013.

[7] J. P. Brandes and R. Corcoran, *CS/HB 1207: Vehicles with Autonomous Technology*. 2012.

[8] M. Denis, *Nevada Legislature SB-313*. 2013.

[9] M. Kowall, *SB-0663 Public Act 251 of 2013*. 2013.

[10] A. Padilla, *SB-1298 Vehicles: autonomous vehicles: safety and performance requirements, Chapter 570*. 2012.

[11] "Chapter XI. Transport and Communications, B. Road Traffic - 19. Convention on Road Traffic. Vienna, 8 November 1968." United Nations Conference on Road Traffic, Nov-1968.

[12] "Report of the sixty-eighth session of the Working Party on Road Traffic Safety," Economic Commission for Europe, Inland Transport Committee, Working Party on Road Traffic Safety, Apr. 2014.

[13] "UNECE Regulation No. 79, Uniform Provisions Concerning the Approval of Vehicles with regard to Steering Equipment."

[14] R. Okuda, Y. Kajiwara, and K. Terashima, "A survey of technical trend of ADAS and autonomous driving," in *Proceedings of Technical Program - 2014 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)*, 2014, pp. 1–4.

[15] H. Lategahn, A. Geiger, and B. Kitt, "Visual SLAM for autonomous ground vehicles," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1732–1737.

[16] TRW Automotive, "Automatic Emergency Braking Collision Mitigation Braking." TRW Automotive, 2013.

[17] Bosch, "Automotive Electronics - Parking made easy - Parking assistance systems from Bosch." Robert Bosch GmbH, 2009.

[18] D. McConnell, "ContiGuard® - Innovations during Normal Driving, Hazard, Pre-Crash, and Post Crash Phases."

[19] Delphi, "Delphi - Collision Mitigation System." Delphi Electronics&Safety, 2009.

[20] C. Grover, I. Knight, F. Okoro, I. Simmons, G. Couper, P. Massie, and B. Smith, "Automated Emergency Brake Systems: Technical requirements, costs and benefits," TRL, Project report PPR 227.

[21] J. Sanchez, J. P. Tecnalia, D. Jeffery, M. McDonald, N. Ford, H. Mosebach, A. Schieben, H. Zwijnenberg, E. van Nunen, S. Maas, T. van den Broek, B. Krosse, and M. van Schijndel-de Nooij, "Definition of necessary vehicle and infrastructure systems for Automated Driving," European Commission, Brüssel, Final Report SMART 2010/0064, 2011.

[22] P. Larsson, I. Esberg, M. van Noort, D. Willemsen, E. García, F. Fahrenkrog, A. Zlocki, J. Scholliers, S. Koskinen, A. Varhelyi, and S. Schönebeck, "InteratIVE - Test and evaluation plans," Ford Research & Advanced Engineering, Germany, interactIVe Deliverable 7.4, Nov. 2012.

[23] E. Jakobsson, A. Beutner, S. Pettersson, H. Stratil, A. Bartels, F. Ahlers, K. Fuerstenberg, A. Amditis, G. Thomaidis, K. Pagle, F. Flemisch, G. Temme, H. Mosebach, A. Schieben, J. Schomerus, E. Frey, M. Sabatier, J. Nemenyi, N. Rauch, M. Armbruster, A. de la Fortelle, M. Netto, L. Nouveliere, S. Mammar, S. Glaser, S. Boverie, A. Giralt, M. Strauss, A. Abele, H. Zeng, B. Chretien, M. Fiedler, and A. Hoess, "Function description and requirements," Continental Automotive GmbH, Germany, HAVEit Deliverable 11.1, Sep. 2008.

[24] M. Klotz and H. Rohling, "24 GHz radar sensors for automotive applications," in *13th International Conference on Microwaves, Radar and Wireless Communications. 2000. MIKON-2000*, 2001, vol. 1, pp. 359–362 vol.1.

[25] C. G. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe, and D. M. Gavrila, "Active Pedestrian Safety by Automatic Braking and Evasive Steering," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1292–1304, Dec. 2011.

[26] C. Lipski, B. Scholz, K. Berger, C. Linz, T. Stich, and M. Magnor, "A Fast and Robust Approach to Lane Marking Detection and Lane Tracking," in *IEEE Southwest Symposium on Image Analysis and Interpretation, 2008. SSIAI 2008*, 2008, pp. 57–60.

[27] G. S. Zanardo, T. Stanger, D. Lang, and L. del Re, "An integrated evaluation approach for performance and safety of autonomous vehicles," in *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, 2013, pp. 933–934.

[28] S. Li and Y. Shimomura, "Lane marking detection by side Fisheye Camera," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008. IROS 2008*, Frankreich, 2008, pp.

606–611.

[29] D. Nienhuser, T. Gumpp, J. M. Zollner, and R. Dillmann, "Recognition and attribution of variable message signs and lanes," in *2008 IEEE Intelligent Vehicles Symposium*, 2008, pp. 55–60.

[30] H. Rohling and E. Lissel, "77 GHz radar sensor for car application," in *Radar Conference, 1995., Record of the IEEE 1995 International*, 1995, pp. 373–379.

[31] Continental, "Industrial Sensors - Made for Automotive - Usable for Industry." Continental, 2010.

[32] Freescale Semiconductor Inc., "Automotive Radar - High-resolution 77 GHz radar." freescale Xtrinsic, 2011.

[33] D. Partrouche, "Intelligent Speed Adaptation in curves for autonomous vehicles," Magister Arbeit, University of Joseph Fourier,Grenoble, Grenoble, 2006.

[34] S. G. Charlton and J. J. de Pont, "Curve speed management," Land Transport New Zealand, New Zealand, Research Report 323, 2007.

[35] S. Solyom and E. Coelingh, "Performance limitations in vehicle platoon control," in *2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2012, pp. 1–6.

Accident Analysis and Prevention 107 (2017) 137–151

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Pre-crash scenarios at road junctions: A clustering method for car crash data

Philippe Nitsche[a,*], Pete Thomas[b], Rainer Stuetz[a], Ruth Welsh[b]

[a] AIT Austrian Institute of Technology, Giefinggasse 2, 1210 Vienna, Austria
[b] Loughborough University, Epinal Way, Loughborough LE11 3TU, UK

ARTICLE INFO

Keywords:
Automated cars
Road safety
Intersections
Clustering
Car crashes
Pre-crash scenarios

ABSTRACT

Given the recent advancements in autonomous driving functions, one of the main challenges is safe and efficient operation in complex traffic situations such as road junctions. There is a need for comprehensive testing, either in virtual simulation environments or on real-world test tracks. This paper presents a novel data analysis method including the preparation, analysis and visualization of car crash data, to identify the critical pre-crash scenarios at T- and four-legged junctions as a basis for testing the safety of automated driving systems. The presented method employs *k*-medoids to cluster historical junction crash data into distinct partitions and then applies the association rules algorithm to each cluster to specify the driving scenarios in more detail. The dataset used consists of 1056 junction crashes in the UK, which were exported from the in-depth "On-the-Spot" database. The study resulted in thirteen crash clusters for T-junctions, and six crash clusters for crossroads. Association rules revealed common crash characteristics, which were the basis for the scenario descriptions. The results support existing findings on road junction accidents and provide benchmark situations for safety performance tests in order to reduce the possible number parameter combinations.

1. Introduction

Over the past few years, automation of road vehicles has gained an increasing presence on the agendas of companies and public authorities, which have started to push Automated Driving Systems (ADS) into the forefront of research. On spots in a road network, where traffic conflicts are likely to occur, e.g. intersections, it must be ensured that automated vehicles can operate safely and efficiently, and even more important, that conventional vehicles driven by humans will have at least the same safety level as they have now. The technical reliability of ADS depends on the functionality under varying road infrastructure and transnational differences as well as on a safe interplay with traditional vehicles and vulnerable road users. Consequently, testing and validation procedures for those systems are paramount. There is a need for comprehensive testing, either in virtual simulation environments or on real-world test tracks. This leads to a challenge, namely to find the key driving situations to be evaluated. Since it is unrealistic to cover all possible combinations of traffic situations and environment conditions, the most representative "benchmark" scenarios must be known.

As road intersections are locations, where the paths of multiple traffic participants are crossed, they are considered high-risk spots for safety researchers. For automated vehicles, road intersections of whatever type constitute a major point of interest along their routes due to the increased likelihood of conflicts with other road users. This paper presents a method to identify such conflict scenarios for the case of road junctions in the UK. It is important to note that the study excludes roundabouts and focuses on three-legged and four-legged intersections, both signalized and unsignalized. The study is based on 1056 junction crashes in the UK, which are initially partitioned by applying the *k*-medoids clustering method (Kaufman and Rousseeuw, 1990). As a second step, association rules (Agrawal et al., 1993) are computed to find associated crash attributes that ultimately build the scenario definition.

This paper is structured as follows: Section 2 gives an overview on relevant literature on road junction safety as well as clustering techniques. The proposed methodology is explained in Section 3, followed by a description of the crash data used in this study (see Section 4). The cluster algorithm and association rule technique are given in Sections 5 and 6, respectively. Section 7 comprises the results, before they are compared to existing findings and related to limitations and future work in Section 8. Finally, the paper is concluded in Section 9.

2. Background

2.1. Motivation and research objectives

Concerning road safety, it is still not clear what impact automated vehicles will have on crash risk, and what kinds of (new) risks they

* Corresponding author.
E-mail address: philippe.nitsche@ait.ac.at (P. Nitsche).

http://dx.doi.org/10.1016/j.aap.2017.07.011
Received 6 March 2017; Received in revised form 24 May 2017; Accepted 10 July 2017

might cause. In particular, the safety risks coming with a mixed vehicle population, namely traffic with both driver-less and driver-operated vehicles are still subject to research. Although automated cars use sophisticated on-board sensors to recognize their environment, they have limitations, e.g. in challenging urban traffic situations, inclement weather conditions or when facing unexpected behaviour of traffic participants.

In Nitsche et al. (2014), an expert survey was conducted including questions on the role of road infrastructure, market readiness as well as to which extent certain factors influence the performance of selected automated driving functions on public roads. In summary, the main challenges found for ADS are complex urban environments, temporary work zones and poor visibility due to bad weather conditions. Road surface characteristics, road alignment and lighting were rated as minor influencing factors.

Three-legged and four-legged junctions are high-risk areas, which future automated cars should be capable to pass safely. Therefore, intersections play a particularly important role in testing assisted and automated driving. Automated vehicles should be capable of safely manoeuvring through an intersection and of avoiding or mitigating a collision. Intersection crash avoidance and mitigation systems (ICAMS) can be categorized into (1) infrastructure-only systems, such as active warning signs for drivers based on detected vehicles, (2) vehicle-based systems, including algorithms to predict and avoid collisions based on in-vehicle sensor data, (3) car-to-car systems based on vehicular communication and (4) cooperative infrastructure-to-vehicle communication systems (Mages, 2008). While the first system group is primarily made for human drivers, automated vehicles mainly rely on vehicle-based systems, but may be assisted by cooperative systems.

The main research gap addressed by this work is that there are no standardized procedures for evaluating automated driving systems in junction environments. To this end, the research objective is to provide a set of pre-crash scenarios to understand typical high-risk situations at junctions. Due to a lack of accident data involving automated vehicles, a reasonable starting point is to analyse historical accidents with human drivers, assuming there is a certain overlap of crash risk. The study is preparatory research to a sub-microscopic simulation study, where virtual test drives will be conducted and ICAMS will be evaluated under varying conditions. The scenarios obtained in the underlying study will help to reduce the possible number of model parameter variations, such as vehicle trajectories, velocities, road and junction parameters etc.

2.2. Safety at road junctions

A query from the CARE crash database (ETSC, 2001) for the years 2003 to 2013 was analysed to get a picture about the intersection accident situation in the European Union. In general, it was found that every third road accident occurs at a junction. Four-legged intersections have the highest amount of both fatal and serious injuries with 43.9 and 43.2 percent, respectively. However, it must be noted that those percentages also depend on the exposure of different junction types, which has not been further analysed in this review. Due to the higher number of conflict points, four-legged junctions are generally unsafer than three-legged junctions (e.g. Bauer and Harwood, 1996; Harwood, 1995; David and Norman, 1975; Hanna et al., 1976). In this paper, safety-critical scenarios are obtained for three- and four-legged junctions, respectively, to further analyse this safety difference.

According to the CARE analysis, persons on pedal cycles and motorcycles were more often fatally injured at junctions than persons using other modes of transport. Every fourth fatally injured bicyclist was killed at a junction, while only every tenth fatally injured car occupant died due to a junction crash.

Van Maren (1980) reported that (multi-lane) unsignalized intersections have a lower number of crashes per million conflicts than signalized intersections. For signalized intersections, it was found that the dominant crash types are rear-end and head-on collisions (Polders et al., 2015; Obeng, 2007), however, Abdel-Aty et al. (2006) states that this also depends on the number of lanes and traffic volumes. In comparison to that, the majority of unsignalized intersection accidents are angle collisions (e.g. Molinero Martinez et al., 2008; Arndt, 2003; Layfield et al., 1996; Pickering and Hall, 1985). The most important variables affecting the safety of unsignalized intersections were studied by Haleem et al. (2010). Accordingly, these include the traffic volume on the major road and the existence of stop signs, and among the geometric characteristics, the configuration of the intersection, number of right and/or left turn lanes, median type on the major road, and left and right shoulder widths. In particular for angle crashes at unsignalized intersections, the factors were found to be traffic volume on the major road, the upstream distance to the nearest signalized intersection, the distance between successive unsignalized intersections, median type on the major approach, percentage of trucks on the major approach, size of the intersection and the geographic location within the state (Abdel-Aty and Haleem, 2011).

Several accident studies (Molinero Martinez et al., 2008; Lee et al., 2004; Najm et al., 2001) show that failure to yield right-of-way is the most dominant violation in crossing path scenarios. This is followed by running a traffic signal or sign as one of the most frequent violations. Sandin (2009) concluded that the most common causation patterns include missed observation due to distraction or sight obstructions, which then led to no, late or premature action. Furthermore, a common causation was found to be incorrect prediction or faulty diagnosis, e.g. the drivers did not expect another vehicle to cross their path. Automated driving systems are expected to mostly solve the safety problems caused by those factors, e.g. through sensing and perception technologies. However, factors such as sight obstructions, unexpected road user behaviour and human error by other drivers still pose problems.

The method presented in this paper analyses historical accident data to understand the critical situations and factors at road junctions. Similar research has been conducted (e.g. Polders et al., 2015; Plavsic, 2010; Molinero Martinez et al., 2008; INTERSAFE, 2005; Wiltschko, 2004), however, the usage of k-medoids clustering and association rules in this context is novel.

2.3. Clustering accident data

In most cases, accident data as used in this study is of categorical nature, i.e. described by qualitative attributes (also called nominal attributes) of mainly arbitrary order. Although the categories can be coded as numbers, e.g. 1: female, 2: male, those numbers would not have mathematical meaning (e.g. Han et al., 2011; Lourenco et al., 2004). Therefore, dedicated statistical methods are necessary to analyse categorical data. Common clustering methods for categorical data are SQEEZER (He et al., 2002), ROCK (Guha et al., 1999), LIMBO (Andritsos et al., 2004), STIRR (Gibson et al., 1998), Link Clustering (Zengyou et al., 2005) or CACTUS (Ganti et al., 1999). Also, conventional clustering algorithms were modified to deal with categorical data, such as *k*-modes (Huang and Ng, 1999; Huang, 1997), *k*-histograms (Zengyou et al., 2003), *k*-medoids (Kaufman and Rousseeuw, 1990) or Generalized Self-Organizing Maps (Hsu, 2006), all of which have their advantages for different applications. Basically not a clustering method, but a popular classification algorithm for categorical data is Latent Class Analysis (Goodman, 1974), which is a model-based approach, assuming that a mixture of underlying probability distributions generates the data. Another approach is to use Multiple Correspondence Analysis (MCA, Lê et al., 2008) as a preprocessing step to transform the categorical variables to a continuous scale. Afterwards standard hierarchical or partitional clustering methods can be applied, usually only on the first principal components to reduce the dimensionality and stabilize the clustering by deleting the noise from the data.

As a popular and simple data mining technique, various researchers used association rules to discover patterns in their data (e.g. Weng et al., 2016; Kumar and Toshniwal, 2015; Montella, 2011; Mirabadi and Sharifian, 2010; Pande and Abdel-Aty, 2009). In this study, association rules are applied to clusters discovered by the *k*-medoids method to get more information on the underlying patterns of accident attributes, as explained in the following section.

3. Overall methodology

The methodology for evaluating the safety performance of assisted and automated driving systems is depicted in Fig. 1. Depending on the objectives and contents of the test study, the target crash population and the safety performance indicators can be defined. This paper is devoted to the left half of the flow chart, with the objective to derive pre-crash scenarios for cars at road junctions. Inspired by a study from Kumar and Toshniwal (2015), the idea was to initially partition the data by a clustering technique for categorical data, and then apply the association rule method on the data subsets to identify further parameters for the respective clusters.

The follow-up study will cover the right half of the chart, by evaluating the safety performance in a virtual simulation environment. The simulation models can be structured into (1) road environment models (including pavement, roadside and environmental conditions such as weather), (2) vehicle models (including sensor and control systems) and (3) driving (behaviour) models. Each of these model groups has numerous parameters to set, leading to a high number of possible combinations in the simulation runs. The method presented can aid engineers in parametrizing the models and to select the parameters that were found to be critical.

The crash data used and its processing steps are explained in Section 4, including the procedure of attribute selection, attribute coding and grouping into two levels. Level 1 is a reduced set of attributes describing the main collision parameters, for better partitioning and easier interpretation of the results, while Level 2 adds additional attributes describing the environment and causation factors. Level-1 data is used as input for the *k*-medoids clustering algorithm and level-2 data for finding association rules. The main reasons why this two-level approach was chosen are the following:

1. The *k*-medoids method achieved good clustering results on a smaller set of attributes. No clear partitioning was achieved when using all available attributes.
2. The results from applying the association rules on the whole dataset (without prior clustering) would be hard to interpret due to the high number of obtained rules. It must be noted that depending on the sample size and attribute dimensionality, millions of rules might be computed. This requires post-processing by applying dedicated algorithms or pruning techniques.

4. Data collection and processing

4.1. Background on the crash data used

The data used for this study stems from a project called OTS (On-The-Spot), which was commissioned by the UK Department for Transport and the Highways Agency (HA). It aimed to establish an in-depth research database of a representative sample of road accidents in the UK, to better understand the cause of accidents and injuries (Hill et al., 2001). Two crash investigation teams collected data from the years 1999 to 2010. One team was located at Loughborough University covering the South Nottinghamshire area in the East Midlands, and the other at the Transport Research Laboratories (TRL) covering the Thames Valley region.

The teams were responsible for collecting information at the scene of the accidents or, when the accidents already occurred, by liaison with emergency services, hospitals and local authorities. To arrive at the accident scene as quickly as possible, the teams had a direct link with the local police, and response vehicles driven by an OTS police officer were used (Cuerden et al., 2008). Data from both teams were collated into a single database that contains more than 2,000 variables.

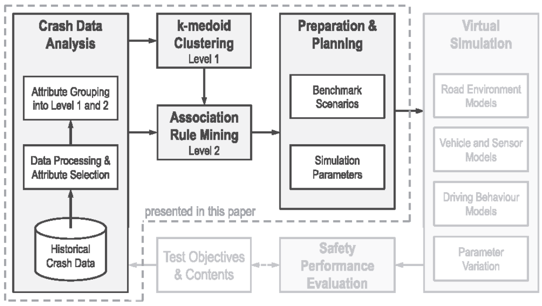


Fig. 1. Overall methodology for evaluating the safety performance of assisted and automated driving systems.

4.2. Data collection

OTS is part of the RAIDS (Road accident in-depth studies) project, whose data query and export tool was used to download all necessary data elements including collisions with the following prerequisites:

- Junction type = "T or staggered junction", "Crossroads", "Multiple junction", "Other junction" or "Using private drive or entrance"
- Police Accident Severity = "Fatal", "Serious" or "Slight"[1]

As mentioned before, roundabouts were excluded from this study. The junction types included comprise signalized and unsignalized junctions of different shapes. This query resulted in 1056 crash cases from the OTS database, including more than 400 variables. However, it was decided to analyze the data on the car driver level, i.e. every sample corresponds to one driver involved in a crash, regardless if he/she was injured or not. This also means that every sample contains a car driven by the respective driver. Consequently, if two or more vehicles are involved in the same crash, the underlying crash and environment data is simply duplicated. Furthermore, there should be at least one car (including car-derived VANs, minibuses and SUVs) involved. This required a second query from the exported database as follows:

- Seating position of occupant = "Driver/Rider"
- At least 1 vehicle = "Car"
- Total number of vehicles ≥ 2

This additional query resulted in an increased sample size of 1540, i.e. car drivers. The requirement of more than one vehicle means that single-car accidents were intended to be excluded, because speeding, fatigue or other human causation for single vehicle accidents are not relevant for the study. Another reason for having one record per driver is given by the background of the analysis, which focuses on safety risks involving automated vehicles instead of drivers. To this end, it is necessary to know the critical situations to be handled by drivers nowadays, as they are likely to happen to automated vehicles as well. Each sample is thus associated with an ego car, later denoted as car A, which collides with a secondary vehicle or road user, later denoted as B.

4.3. Attributes selection and coding

The number of variables was further reduced according to the

[1] It is important to mention that although the police reported a certain injury level, this might have been adapted by the crash investigation team based on more precise evidence.

following steps:

1. Include only variables that fit the scope of the study (see next section), e.g. not relevant were weekday or time of the crash, occupant data such as age or gender, vehicle damage or detailed injury data of different body parts.
2. Exclude variables with low variance, because they would fail to make a positive impact on model performance. In this study, all observations with more than 95 percent same values were excluded.
3. Group or combine highly correlated variables, e.g. OTS injury severity and police injury severity.
4. Exclude variables having unknown values in more than 30 percent of all samples.

Following this reduction process, the number of variables has been reduced to 41, which were grouped according to the original OTS data hierarchies "scene", "vehicle" and "path". The "scene" variables include general attributes about the crash, such as collision type and maximum injury of all involved persons. The "vehicle" variables are related to the pre-crash and collision circumstances from the perspective of the individual vehicle, i.e. driver, and includes for example the precipitating factor attributed to the vehicle, driver injury level or the pre-impact manoeuvre. The "path" variables describe the road environment, e.g. junction type, weather, traffic density or speed limit.

The original data contains variables in the following format: "Maximum injury level = Serious" from the four possible values uninjured, slight, serious and fatal. For the further calculations, all variables were converted to the binary-coded format. Consequently, this resulted in many more attributes, as each possible value was assigned to its own column, but it is a necessary step for applying most clustering algorithms.

The high number of attributes of the pre-processed OTS dataset made it necessary to further prepare the data for clustering. Usually, fewer attributes make it easier to interpret the clusters. Initial experiments with a varying number of attributes as input showed that the performance of the *k*-medoid method suffers from a higher dimensionality. Therefore, all attributes were divided into two levels as follows:

1. First level (5 variables, 25 attributes, see Table 1): This level of attributes was used as input for the *k*-medoids clustering. The idea is to derive clusters based on a set of main collision attributes first, before association rule mining is applied to each cluster with the second level attributes.
2. Second level (15 variables, 86 attributes, see Table 2): This level adds more detailed attributes on road infrastructure and accident causation to the level-1 attributes. They are intended to help tell a "story" describing each cluster by association rule mining.

As described above, the second-level attributes deliver more information on the accident environment and causation. Most of the additional attribute groups in Table 2 are related to the vehicle's path describing the road layout, e.g. road type, speed limit or curvature. The attribute groups "collision code", "precipitating factor" and "driver injury" were added to the list to better understand the accident circumstances.

4.4. Further removal of unknowns

Samples with at least one unknown attribute value were removed as part of the data processing steps. This happened at two instances, namely (1) before computing the cluster with level-1 data and (2) before computing the rules with level-2 attributes for the data in each cluster. The first removal of unknowns resulted in a final sample size of $n = 1325$ for clustering, including $n = 930$ for T-junctions, $n = 368$ for crossroads and $n = 27$ for other or no junctions. The frequencies of the attributes are given on the right-hand side in Table 1. The second removal of unknowns was done on the extended level-2 dataset. Therefore, the final overall sample size ($n = 1070$) of the dataset used for the association rules is different to the clustering dataset (see Table 2).

5. Clustering of junction crashes

Due to different principles of clustering algorithms, one method might produce different clusters to another method. Hence, one has to choose the most appropriate method for the underlying dataset, taking into account the sample size, the number of attributes, the attribute types as well as the desired output of the study. The following sections address the clustering method chosen, which parameters were chosen and how it was applied to the OTS dataset.

5.1. The k-medoids method

The *k*-medoids method was chosen for the clustering, because it can cope with categorical data and is robust against outliers. It uses objects called medoids instead of centroids, as the popular *k*-means method does. Instead of using the mean as centre of the cluster, a member of the cluster is chosen as centre, whose average dissimilarity to all the objects in the cluster is minimal. In other words, the medoid is the most centrally located point in the cluster. Thus it is more robust to outliers, because it does not minimize a sum of squared Euclidean distances, as *k*-means does. Furthermore, *k*-medoids allows clustering categorical data, where a mean is impossible to define. For this reason, alternative dissimilarity measures can be applied, such as the "Hamming distance" (Hamming, 1950; Wegner, 1960) or the "Jaccard coefficient" (Jaccard, 1901).

One of the most powerful and commonly used algorithm for *k*-medoids is PAM (Partitioning Around Medoids) proposed by Kaufman and Rousseeuw (1990). It proceeds in two steps as follows: Build step:

1. Choose k objects to become the medoids, or in case these objects were provided use them as the medoids
2. Calculate the dissimilarity matrix if it was not informed
3. Assign every object to its closest medoid

Swap step:

1. Within each cluster, each object is tested as a potential medoid by checking if the sum of within-cluster distances gets smaller using that object as the medoid. If so, the object is defined as a new medoid.
2. If at least one medoid has changed, go to (3), else end the algorithm.

The PAM algorithm works effectively for relatively small datasets such as the underlying OTS dataset. For larger datasets, alternative *k*-medoids algorithms should be used, such as CLARA (Clustering Large Applications, Kaufman and Rousseeuw, 1990).

5.2. Parameters used

The PAM algorithm was used, because it is most appropriate for the given sample size. The algorithm can produce better solutions than other *k*-medoids algorithms in some situations, but the computation times can be longer. The Hamming distance, originally used for the detection of errors in information transmission, was chosen as distance measure. It simply gives the number of mismatches between two vectors, thus it does not prefer 1's over 0's.

To study the separation of the resulting clusters, silhouette analysis (Rousseeuw, 1987) was used. Each cluster is represented by silhouette coefficients, which provide a measure of how close each point in one cluster is to points in the neighbouring clusters. Observations with silhouette coefficients near 1 are very well clustered. Small values

Table 1
Crash attributes used for *k*-medoid clustering (level 1).

Category	Short name	Description	Count	Rel. frequency
Max. injury (of all persons involved in the crash)	MaxInj = Uninjured	No person injured (OTS injury level)	196	14.8%
	MaxInj = Slight	At least one person slightly injured (OTS injury level)	919	69.4%
	MaxInj = SeriousFatal	At least one person seriously or fatally injured (OTS injury level)	210	15.8%
Junction shape (attributed to the vehicle's path)	JctShp = X-minJoin	Road continues straight on with (minor) road joining from the left and right (crossroad)	224	16.9%
	JctShp = X-brkMaj	Road is temporarily broken by a (major) road passing across the vehicles path (Crossroad)	144	10.9%
	JctShp = NoJct	No junction present	20	1.5%
	JctShp = Other	Private drive, entrance or other junction type	7	0.5%
	JctShp = T-minLeft	Road continues straight on with (minor) road joining from the left	350	26.4%
	JctShp = T-minRight	Road continues straight on with an additional (minor) road joining from the right (T-Junction)	309	23.3%
	JctShp = T-termMaj	Road terminates with a (major) road passing across the vehicles path (T-Junction or accel. lane)	271	20.5%
First interaction (Road user type or object which the vehicle first interacted with)	1stIntAct = Car	Driver interacted with another car	987	74.5%
	1stIntAct = LGV-HGV	Driver interacted with a large or heavy goods vehicle	97	7.3%
	1stIntAct = PTW	Driver interacted with a powered two-wheeler (motorcycle or moped)	115	8.7%
	1stIntAct = Other	Driver interacted with another type of vehicle or object	37	2.8%
	1stIntAct = Cycle	Driver interacted with a bicyclist	50	3.8%
	1stIntAct = Pedestrian	Driver interacted with a pedestrian	39	2.9%
Manoeuvre (Action of the vehicle immediately before crash)	Manvr = GoingAheadOther	Driver was going straight ahead	781	58.9%
	Manvr = TurnL	Driver was turning left	59	4.5%
	Manvr = TurnR	Driver was turning right	79	6.0%
	Manvr = WaitTurnR	Driver was waiting to turn right	353	26.6%
	Manvr = Other	Driver was reversing, doing a u-turn, overtaking, undertaking, held up or waiting to turn left	53	4.0%
First point of impact (First point to come into contact with another vehicle, pedestrian or other object)	1stImpact = Back	First point of the impact was the car's back	126	9.5%
	1stImpact = Front	First point of the impact was the car's front	674	50.9%
	1stImpact = Nearside	First point of the impact was the car's nearside	218	16.5%
	1stImpact = Offside	First point of the impact was the car's offside	307	23.2%

indicate that the observation is close to the decision boundary between to neighbouring clusters and observations with negative values are probably placed in the wrong cluster. The average silhouette width provides a measure for clustering validity, and is used to choose the most appropriate number of clusters.

The best number of clusters k was achieved by iteratively stepping from $k_{min} = 2$ to $k_{max} = 15$ clusters. Experiments with the dataset showed that a k_{max} greater than 15 does not result in any more change of the error function, as the curve flattens. The results from each k were compared to find the best k, i.e. the one with the lowest average silhouette value. Actually, finding the best k is one of the most debated problems in cluster analysis. In literature, various validity metrics can be found to compute the performance in partitioning, among which are the Akaike's Information Criterion (Akaike, 1974), the Bayesian Information Criterion (Schwarz, 1978), Calinski-Harabasz (Calinski and Harabasz, 1974) or Davies Bouldin index (Davies and Bouldin, 1979). For the scope of this study, it was sufficient to compare the silhouette values for graphical display for validating clusters. The entire clustering is displayed by combining the silhouettes into a single plot, as seen in Fig. 2 (right) and Fig. 3 (right) in a later section. The height of the silhouette represents the cluster size. For evaluating the best k, the average silhouette value of all objects within a cluster is calculated and compared to the others.

6. Specifying crash scenarios

As explained in the methodology section, the obtained clusters are further analysed by association rule mining, which was implemented in R by using the arules package (Hahsler et al., 2017, 2005). This section gives an overview on the principle of association rules and how the rules help to derive scenario parameters.

6.1. The association rules method

Association rule mining is a method to discover associations between attributes, also called "frequent itemset mining". A popular example of association rules is the market basket analysis, where retailers can get insights into which items are frequently purchased together so that marketing strategies and product shelving can be optimized. For example, if a customer buys "beer", then he/she often buys "crisps". This would be expressed as "beer → crisps", where the item "beer" is called the antecedent and the item "crisps" the consequent. One itemset I can contain multiple items. Applying the association rules terminology to the OTS dataset, then each sample is called a transaction $\{t_1, t_2, \ldots, t_n\} \in T$, and each attribute is an item $\{i_1, i_2, \ldots, i_m\} \in I$. An association rule can be written in the following mathematical form: $X \to Y$ where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. Each rule is characterised by its support (see Eq. (1)) and its confidence (see Eq. (2)).

$$\mathrm{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{n} = P(X) \tag{1}$$

For itemsets, the support value gives the proportion of transactions t in the dataset, which contains the itemset X. For rules, the support is defined as the support of all items in the rule, i.e. $\mathrm{supp}(X \to Y) = \mathrm{supp}(X \cup Y) = P(X \wedge Y)$.

$$\mathrm{conf}(X \to Y) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X)} = P(Y|X) \tag{2}$$

Equivalently, the confidence measures the strength of the rules and gives the conditional probability of the consequent Y given the antecedent X. In other words, it is the proportion of the transactions that contains X, which also contains Y. To explain the difference between the two measures, it is important to mention that two rules with flipped

Table 2
Additional crash attributes used for association rule mining (level 2).

Category	Short name	Description	Count	Rel. frequency
Collision type (The category letter of the UK STATS-19 collision code)	Coll = D-Cornering	Cornering (D)	16	1.5%
	Coll = H-CrossingNoTurns	Crossing (no turns) (H)	202	18.9%
	Coll = J-CrossingVehTurning	Crossing (vehicle turning) (J)	236	22.1%
	Coll = M-Manoeuvring	Manoeuvring (M)	104	9.7%
	Coll = Other	Other collision code	11	1.0%
	Coll = A-OvertakingLaneChange	Overtaking and lane change (A)	30	2.8%
	Coll = P-PedestrOther	Pedestrians Other (P)	25	2.3%
	Coll = F-RearEnd	Rear end (F)	188	17.6%
	Coll = L-RightTurnAgainst	Right turn against (L)	204	19.1%
	Coll = G-TurningVsSameDir	Turning versus same direction (G)	54	5.0%
Precipitating factor (The main cause of the crash, attributed to the respective occupant)	Prec = FailAvoidDriver	Driver failed to avoid object or vehicle on carriageway	64	6.0%
	Prec = FailAvoidOther	Other road user failed to avoid object or vehicle on carriageway	58	5.4%
	Prec = FailGiveWayDriver	Driver failed to give way	266	24.9%
	Prec = FailGiveWayOther	Other road user failed to give way	217	20.3%
	Prec = FailStopDriver	Driver failed to stop	84	7.9%
	Prec = FailStopOther	Other road user failed to stop	95	8.9%
	Prec = LossCntrDriver	Driver lost control of vehicle	23	2.1%
	Prec = LossCntrOther	Other road user lost control of vehicle	17	1.6%
	Prec = OtherDriver	Other precipitation by driver	27	2.5%
	Prec = OtherOther	Other precipitation by another road user	29	2.7%
	Prec = PedEnter	Pedestrian entered road without due care (driver not to blame)	17	1.6%
	Prec = PoorOvtkDriver	Inappropriate overtake by driver	7	0.7%
	Prec = PoorOvtkOther	Inappropriate overtake by other driver or rider	23	2.1%
	Prec = PoorMnvrDriver	Inappropriate turn or manoeuvre by driver	80	7.5%
	Prec = PoorMnvrOther	Inappropriate turn or manoeuvre by other driver or rider	63	5.9%
Driver injury (OTS injury level of the respective driver)	DrvInj = Uninjured	Driver suffered no injury	576	53.8%
	DrvInj = Slight	Driver was slightly injured	445	41.6%
	DrvInj = Serious	Driver was seriously injured	42	3.9%
	DrvInj = Fatal	Driver was fatally injured	7	0.7%
Area (around the crash location)	Area = Rural	Rural area (countryside, fields and only sparse housing)	368	34.4%
	Area = Urban	Urban area (at least one side of the road built up)	702	65.6%
Horizontal geometry (Qualitative assessment of curvature of road)	HorizGeom = Left	Left curve	22	2.1%
	HorizGeom = LeftSharp	Left sharp curve	4	0.4%
	HorizGeom = LeftSlight	Left slight curve	51	4.8%
	HorizGeom = Right	Right curve	25	2.3%
	HorizGeom = RightSharp	Right sharp curve	9	0.8%
	HorizGeom = RightSlight	Right slight curve	77	7.2%
	HorizGeom = Straight	Straight (no curve)	882	82.4%
Lighting (Light conditions at the time of the crash)	Light = DarkNSL	Darkness: no street lighting	50	4.7%
	Light = DarkSLUnk	Darkness: street lighting unknown	11	1.0%
	Light = DarkSL	Darkness: street lights lit	188	17.6%
	Light = DayNSL	Daylight: no streetlighting present	571	53.4%
	Light = DaySLUnk	Daylight: streetlighting unknown	243	22.7%
	Light = DaySL	Daylight: streetlights present	7	0.7%
Road type (on which the crash occurred)	RdType = DualCgw	Dual carriageway	161	15.0%
	RdType = OneWayStr	One way street	26	2.4%
	RdType = SingCgw	Single carriageway	883	82.5%
Speed limit (posted at the crash location)	SpdLim ≤ 20 mph	20 mph and less	1	0.1%
	SpdLim = 30 mph	30 mph	584	54.6%
	SpdLim = 40–50 mph	40 or 50 mph	270	25.2%
	SpdLim = 60 mph	60 mph	159	14.9%
	SpdLim = 70 mph	70 mph	56	5.2%
Surface (Road surface condition due to weather at the crash location)	Surf = Dry	Dry surface	673	62.9%
	Surf = Flood	Flooded surface	9	0.8%
	Surf = Icy	Icy surface	6	0.6%
	Surf = Snowy	Snowy surface	3	0.3%
	Surf = Wet	Wet surface	379	35.4%
Traffic control (Type of traffic control at the location of the crash)	TrfCtrl = None	No active or static yield instruction	582	54.4%
	TrfCtrl = GW	Static give-way instruction	245	22.9%
	TrfCtrl = Stop	Static stop instruction	14	1.3%
	TrfCtrl = Light	Traffic light control	229	21.4%

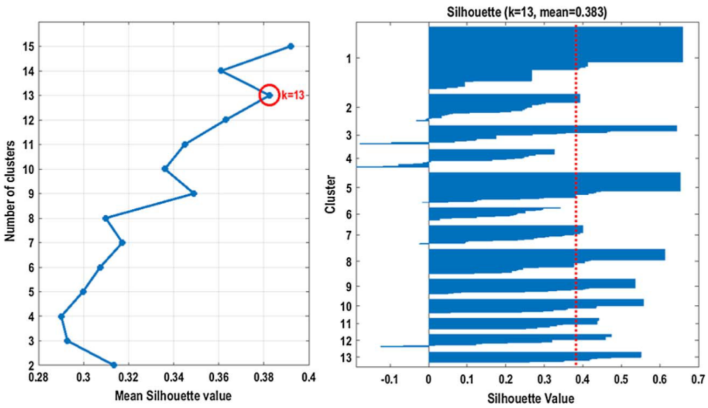


Fig. 2. Mean silhouette values for all k's (left) and silhouette plot for $k = 13$ (right) for T-junction clusters.

antecedent and consequent would both have the same support value. However, they would not have the same confidence, because the direction is taken into account.

The most common implementation was proposed by Agrawal et al. (1993), who called their method the Apriori algorithm. Accordingly, finding association rules involves two steps: (1) Find all frequent itemsets and (2) generate association rules from the frequent itemsets. The algorithm necessitates two parameters, namely a minimum support threshold, and a minimum confidence. By definition, if an itemset is below the minimum support threshold, then it is not frequent. If so, all its subsets must also be infrequent and can be pruned. In contrary, any subset of a frequent itemset must be frequent. By following this principle iteratively, the number of possible itemset configurations can be reduced tremendously with a simple algorithm.

The second step is to generate rules from the frequent itemsets found in Step 1. Here, the minimum confidence threshold comes into play: For each frequent itemset I, all nonempty subsets are generated. For every non-empty subset s of I, create the rule $s \rightarrow (I - s)$ if the minimum confidence for this rule is given. Since the rules are generated from frequent itemsets, each one also satisfies the minimum support. In this way, strong association rules can be found.

Depending on the data dimensionality, and on how low the minimum support and confidence thresholds have been set, the algorithm might produce millions of rules. Dedicated rule pruning and post-processing methods have been developed to find the rules of most interest. It was found that the confidence measure is a rather poor measure to discover the dependence of the consequent with respect to the antecedent (Guillaume et al., 1998; Silverstein et al., 1998). This paper uses a metric called lift (see Eq. (3)), also known as "interestingness".

$$\mathrm{lift}(X \rightarrow Y) = \mathrm{lift}(Y \rightarrow X) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X) \cdot \mathrm{supp}(Y)} = \frac{P(X \wedge Y)}{P(X)P(Y)} \tag{3}$$

If the lift value is less than 1, then the occurrence of X is negatively correlated with the occurrence of Y, meaning that the occurrence of one likely leads to the absence of the other one. If the resulting value is greater than 1, then X and Y are positively correlated, meaning that the occurrence of one implies the occurrence of the other. If the lift equals 1, then X and Y are independent (Han et al., 2011). By setting an appropriate minimum lift value greater than 1, only high-lift rules can be extracted for interpretation.

6.2. Parameters used

The choice of the minimum support and confidence depends on the application and the expected outcome of the study. In theory, it is desirable to obtain rules with high support, high confidence and a lift value much greater than 1. The idea of this paper implies the analysis of certain accident situations and characteristics, which can be very rare (Montella et al., 2012). After experimenting with different values, a minimum support of 0.03 was chosen, so that all itemsets occurring in less than 3 percent of the samples are disregarded. Choosing a lower threshold results in an increase of computation time and rules, which would all have to be interpreted. Choosing a higher support value might disregard relevant information about the clusters. There are different approaches in literature on the choice of a minimum confidence value. For example, Montella (2011) chose a threshold with conf = 0.1 for

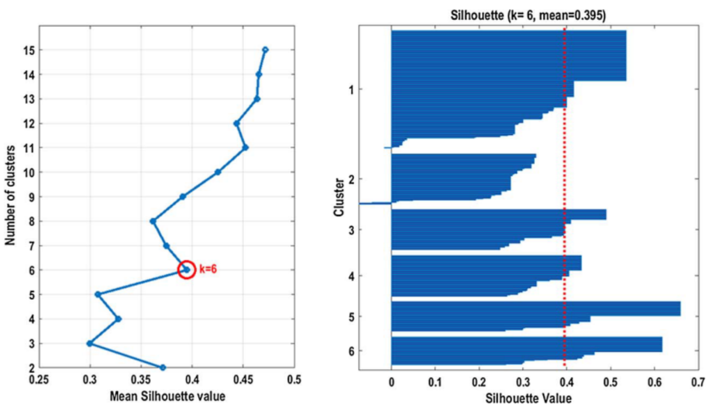


Fig. 3. Mean silhouette values for all k's (left) and silhouette plot for $k = 6$ (right) for four-legged junction clusters.

their powered two-wheeler (PTW) study, which is much lower than usual. However, in this paper it is preferred to obtain rules, where the probability of the consequent given the antecedent is higher than 75 percent. Additionally, only rules with a lift > 1.25 are considered for the results.

To further reduce the number of rules obtained, redundant rules were excluded according to the following procedure: A rule is redundant if a more general rule with the same or a higher lift exists. That is, a more specific rule is redundant if it is only equally or even less correlated than a more general rule. A rule is more general if it has the same consequent but one or more items removed from the antecedents. Formally, a rule $X \rightarrow Y$ is redundant if for $X' \subset X : \text{lift}(X' \rightarrow Y) \geq \text{lift}(X \rightarrow Y)$ (Hahsler et al., 2017).

7. Results

The crash dataset was divided into the two main junction types: (1) Three-legged T-junctions and (2) four-legged crossroads. For other types of junctions (e.g. private drives, pedestrian crossings), the sample size was too small ($n = 27$) to compute clusters. This partitioning prior to clustering was done due to the scope of the study, namely to provide targeted scenarios and parameter variations for virtual vehicle simulations. The goal was not to find clusters characterized by junction types, but by driving situations, manoeuvres and injury outcome (see level-1 attributes). Furthermore, the number of intersection legs was found to be a significant variable to model intersection crashes (Abdel-Aty et al., 2006) and was used to group intersection crashes in various studies (e.g. Abdel-Aty and Haleem, 2011; Arndt, 2003; Persaud and Nguyen, 1998; Vogt and Bared, 1998).

7.1. Clusters found for T-junctions

The silhouette plot in Fig. 2 (left) shows the average silhouette values (cluster validity) for all *k*s. In general, the higher the number of clusters the higher the silhouette values get. A higher number of clusters might be over-fitting and a lower number of clusters might be under-fitting. To find the best *k*, a compromise between cluster size and cluster validity had to be found. Association rules, which are computed for each cluster in the next step, were originally made for large-scale data. Hence, the goal was to avoid very small clusters, i.e. results with clusters containing less than 30 samples are disregarded ($k = 14$ and $k = 15$). Since $k = 13$ has the highest average silhouette value with 0.383, the lowest number of samples that were allocated to the wrong cluster, and overall, the lowest percentage of clusters with negative silhouette values, it was chosen as most valid *k*.

Fig. 2 (right) depicts the silhouette plot for each of the thirteen clusters, with one horizontal bar per sample within the cluster. Samples with a negative silhouette value might be assigned to the wrong cluster. However, the number of those samples is considerably low, expect for cluster 4, where the average silhouette value suffers compared to the other clusters. Cluster 4 must therefore be treated carefully when interpreting the results.

The frequencies of each attribute within each cluster were compiled in a table to present the results at a glance (see Table 3). Cells shaded in grey indicate that the distribution of numbers for the given field is significantly different from the distribution in the whole population (χ^2-test with significance $\alpha = 0.05$) and that the particular number highlighted is over-represented. Due to values lower than 5 in the expected frequency table, the χ^2-test could not be applied to all observations.

Cluster T-C1 is the largest cluster with a size of 212 crashes, from which all resulted in slight injury. More than 90 percent of the accidents occurred at T-junctions with a minor road joining from the left. “1stImpact = Front” and “1stImpact = Back” are over-represented as well as “Manvr = GoingAheadOther”. There is no clear indication on the collision type of this cluster, thus association rules are used for further analyses. The third largest **cluster T-C2** clearly groups collisions while turning, with a highly significant representativeness of frontal and nearside impacts, all of which occurring at roads terminated by a major road. Powered two-wheelers (PTW) and bicyclists have relatively high frequencies, but the car is still the dominant crash partner. **Cluster T-C3** with 62 samples represents car-to-car collisions at roads with minor roads joining from the right, mainly resulting in slight injury. Since there are mainly impacts on the back of the car, this cluster can be seen as rear-end crash group. **Cluster T-C4** occurred on a road with a minor road joining from the right, with nearside impacts in 77 percent of the cases and high frequencies for “Manvr = TurnR” and “1stIntAct = Car”. The second largest cluster **Cluster T-C5** indicates rectangular collisions with another car crossing the car's trajectory from the right, although this assumption will be validated by association rule mining. **Cluster T-C6** is characterized by a left turn into a major road, which results in a collision mainly with another car. This cluster has a relatively high number of bicycle crashes (10). All 63 accidents in **Cluster T-C7** resulted in no injury for any of the participants. This is clearly a minor risk cluster mainly with cars and goods vehicles involved, with “Manvr = GoingAheadOther” having a high frequency. **Cluster T-C8** represents slight injury collisions with mainly other cars or PTW. Offside impacts were found over-represented, while turning right into a major road. **Cluster T-C9** involves nearside collisions only, which happened on a T junction with a minor road joining from the left, while the car was going straight. **Cluster T-C10** represents a group of high-risk collisions with serious or fatal injuries in all 46 cases. Front impacts

Table 3
Cluster results for T-junctions ($k = 13$, $n = 930$).

Level-1 Attributes	T-C1	T-C2	T-C3	T-C4	T-C5	T-C6	T-C7	T-C8	T-C9	T-C10	T-C11	T-C12	T-C13
Sample size	**212**	**90**	**62**	**62**	**102**	**43**	**63**	**83**	**52**	**46**	**38**	**42**	**35**
MaxInj=Uninjured	0	11	7	8	15	9	63	7	0	0	4	2	5
MaxInj=Slight	212	69	52	42	78	30	0	68	45	0	29	0	25
MaxInj=SeriousFatal	0	10	3	12	9	4	0	8	7	46	5	40	5
JctShp=T-minLeft	195	0	1	0	0	0	58	3	51	41	0	1	0
JctShp=T-minRight	0	0	58	62	102	0	0	0	0	0	38	14	35
JctShp=T-termMaj	17	90	3	0	0	43	5	80	1	5	0	27	0
1stIntAct=Car	183	60	53	40	81	24	53	60	37	33	27	0	23
1stIntAct=LGV-HGV	18	6	5	4	5	2	7	2	5	5	4	3	2
1stIntAct=PTW	3	10	3	10	6	4	0	14	3	2	3	35	6
1stIntAct=Other	4	4	1	3	4	2	2	2	4	2	1	1	0
1stIntAct=Cycle	1	8	0	5	2	10	0	5	2	1	1	2	1
1stIntAct=Pedestrian	3	2	0	0	4	1	1	0	1	3	2	1	3
Manvr=GoingAheadOther	201	0	17	6	97	0	50	7	45	43	27	1	0
Manvr=Other	8	11	1	4	5	0	4	2	2	3	11	2	0
Manvr=TurnL	2	0	0	1	0	43	7	0	4	0	0	2	0
Manvr=TurnR	1	75	5	51	0	0	1	69	0	0	0	36	35
Manvr=WaitTurnR	0	4	39	0	0	0	1	5	1	0	0	1	0
1stImpact=Back	25	9	55	0	0	5	10	0	0	4	0	1	0
1stImpact=Front	162	68	1	0	102	18	35	0	0	37	0	11	35
1stImpact=Nearside	0	13	1	48	0	6	10	0	52	0	0	1	0
1stImpact=Offside	25	0	5	14	0	14	8	83	0	5	38	29	0

Table 4
Cluster results for four-legged junctions (k = 6, n = 368).

Level-1 Attribute	X-C1	X-C2	X-C3	X-C4	X-C5	X-C6
Sample size	**142**	**60**	**48**	**49**	**35**	**34**
MaxInj=Uninjured	22	13	8	10	4	4
MaxInj=Slight	98	39	35	29	28	24
MaxInj=SeriousFatal	22	8	5	10	3	6
JctShp=X-minJoin	142	0	48	0	0	34
JctShp=X-brkMaj	0	60	0	49	35	0
1stIntAct=Car	118	44	38	39	30	28
1stIntAct=LGV-HGV	9	4	4	6	2	4
1stIntAct=PTW	3	7	1	3	1	0
1stIntAct=Other	3	1	1	0	0	2
1stIntAct=Cycle	2	2	4	1	1	0
1stIntAct=Pedestrian	7	2	0	0	1	0
Manvr=GoingAheadOther	116	32	25	35	25	29
Manvr=Other	4	0	0	0	0	1
Manvr=TurnL	5	9	2	2	1	1
Manvr=TurnR	15	19	21	12	9	3
Manvr=WaitTurnR	2	0	0	0	0	0
1stImpact=Back	12	5	0	0	0	0
1stImpact=Front	130	55	0	0	0	0
1stImpact=Nearside	0	0	48	0	35	0
1stImpact=Offside	0	0	0	49	0	34

are over-represented and "Manvr = GoingAheadOther" and "1stIntAct = Car" have high frequencies. Association rules will be used to analyse this cluster in more detail. In comparison to T-C9, **cluster T-C11** involves offside collisions only, which happened on a T junction with a minor road joining from the right, while the car was going straight or made another manoeuvre. Five of the 38 cases resulted in serious or fatal injury. **Cluster T-C12** is a PTW cluster, with 40 out of 42 collisions resulting in serious or fatal injury. In 85 percent of the cases, the car was turning right. Association rules will be used to analyse this cluster in more detail. The smallest **cluster T-C13** is characterized by right-turns into a minor road, with "1stImpact = Front" in all cases. Five of the 35 cases resulted in serious or fatal injury, which is most likely due to the six cases involving PTW. Association rules will be used to analyse this cluster in more detail.

7.2. Clusters found for four-legged junctions

For the crossroads dataset with 368 samples, k = 6 was found to be most valid for separating the clusters, because it has a high average mean silhouette value of 0.395. The silhouette plot in Fig. 3 (left) shows the average silhouette values for all k's. Although larger values were computed for higher k's (10–15), they were disregarded due to their small cluster sizes (< 30) and possible overfitting. Fig. 3 (right) depicts the silhouette plot for each of the six clusters, with one horizontal bar per sample within the cluster. The total mean silhouette value is higher and the number of samples with a negative value is lower compared to the T-junction dataset. This means that for the attributes and for the k chosen, the crossroads dataset seems to be better separated.

As for the T-junction dataset, the frequencies of each attribute within each cluster were compiled in a table to present the results at a glance (see Table 4). Cells shaded in grey indicate that the distribution of numbers for the given field is significantly different from the distribution in the whole population (χ^2-test with significance α = 0.05) and that the particular number highlighted is over-represented.

Table 4 shows that the four-legged junction dataset is mainly separated by the type of junction and first point of impact. Experiments with varying parameters, such as initial medoid configuration or including the missing values did not result in different partitions. Including more attribute groups resulted in a decrease of the average silhouette value. For all clusters, the χ^2-test was not applied to the attribute groups "1stIntAct" and "Manvr" due to expected frequency values lower than 5. For the attribute group "1stImpact", only cluster X-C1 had sufficient frequency values for a χ^2-test. The distributions for injury level ("MaxInj") do not significantly differ in any cluster from the total population in their attribute group.

Cluster X-C1 is the largest cluster with 142 samples, which seems to mainly include rear-end collisions, as the clusters X-C3 to X-C6 have no samples for "1stImpact = Back" and cluster 2 has only 5. **Cluster X-C2** groups situations on crossroads broken by a major road, with high numbers for turning left or right as well as "1stImpact = Front". Cars and PTWs were mostly involved. All situations in **Cluster X-C3** occurred on a road with minor roads joining from the left and right, and in all situations the car was hit on its nearside. All situations in **Cluster X-C4** occurred on a road broken by a major road passing the car's path, and in all situations the car was hit on its offside. All situations in **Cluster X-C5** occurred on a road broken by a major road passing the car's path, and in all situations the car was hit on its nearside, mainly by another car. As for the previous clusters, there is no statistical significance given for the manoeuvre, interaction or injury level distribution. The smallest **Cluster X-C6** represents collisions at roads with minor roads joining from left and right, where the car was hit on its offside, while going straight over the junction.

7.3. High-injury scenarios derived from association rules

For each identified cluster, association rules were computed using the parameters given in Section 6.2. In total, the analysis of each cluster resulted in 35 different crash scenarios comprising various parameters. Due to the high number obtained, not all of the rules for each cluster can be given in this paper. Therefore, only high-risk scenarios, which resulted in serious or fatal injury, are presented in this section, as they provide a set of safety-critical situations. More precisely, the further scenarios include crash situations from the T-junction clusters T-C4, T-C10, T-C12 and T-C13, and from the crossroads clusters X-C1, X-C2, X-C4 and X-C6. All rules obtained for each cluster are available as supplementary material to this paper.

As an example, Cluster T-C10 is selected for further explanation. Given the distributions in Table 3, the cluster can be described as follows: "The car hits another car with its front resulting in serious or fatal injury, while going straight on a road with a minor road joining from the left."

A useful attribute to give a clearer indication about the crash circumstances is the collision type (indicated by letters A to Q in the OTS data specification, see Appendix A). For cluster T-C10, the collision types L ("Right Turn Against") and J ("Crossing with Vehicle Turning") were found to be the most frequent. Therefore, all rules containing those attributes within their items were further analysed to see which other attributes are associated with them.

Table 5 gives the 2-item and 3-item rules for T-C10 and collision type L, sorted by the five highest support values. The rules are sorted by the support to obtain the attributes that are often associated with each other. It can be seen that this collision type is associated with single carriageways (rule nr. 1) as well as with no traffic control ("TrfCtrl = None", see rule nr. 2, 4 and 11) and going straight ("Manvr = GoingAheadOther", see rule nr. 3). Another car as collision partner has already been defined by the cluster, but the rules reveal that "Coll = L_RightTurnAgainst" and "FirstIntAct = Car" are associated with dry surface (see rule nr. 5), uninjured driver of the ego car (see rule nr. 10), a fail to give way by the other car driver (see rule nr. 12), daylight (see rule nr. 13), 40–50 mph speed limit (see rule nr. 9) and urban area (see rule nr. 22).

Table 6 gives the 2-item and 3-item rules for T-C10 and collision type J, sorted by the five highest support values. It can be seen that this collision type is associated with a fail to give way by the other driver (see rule nr. 1). This combination is further associated with another car as collision partner (see rule nr. 5), no traffic control (see rule nr. 6), wet surface (see rule nr. 10), single carriageway (see rule nr. 11), rural area (see rule nr. 12), serious driver injury (see rule nr. 20) and 40–50 mph speed limit (see rule nr. 35). Taking a deeper look into the serious driver injuries, it can be noted that they are further associated

Table 5
Rules obtained for T-C10 with collision type L, sorted by the five highest support values.

Nr.	Antecedent	Consequent	Supp	Conf	Lift
1	Coll = L_RightTurnAgainst	RdType = SingCgw	0.237	0.818	1.413
2	Coll = L_RightTurnAgainst & TrfCtrl = None	RdType = SingCgw	0.237	1.000	1.727
3	Coll = L_RightTurnAgainst & Manvr = GoingAheadOther	RdType = SingCgw	0.237	0.900	1.555
4	Coll = L_RightTurnAgainst & RdType = SingCgw	TrfCtrl = None	0.237	1.000	1.357
5	Coll = L_RightTurnAgainst & Surf = Dry	FirstIntAct = Car	0.184	1.000	1.357
6	Coll = L_RightTurnAgainst & Surf = Dry	RdType = SingCgw	0.158	0.857	1.481
7	Coll = L_RightTurnAgainst & Area = Rural	RdType = SingCgw	0.158	0.857	1.481
8	Coll = L_RightTurnAgainst & DrvInj = Uninjured	RdType = SingCgw	0.132	1.000	1.727
9	Coll = L_RightTurnAgainst & SpdLim = 40–50 mph	FirstIntAct = Car	0.132	1.000	1.357
10	Coll = L_RightTurnAgainst & DrvInj = Uninjured	FirstIntAct = Car	0.132	1.000	1.357
11	Coll = L_RightTurnAgainst & DrvInj = Uninjured	TrfCtrl = None	0.132	1.000	1.357
12	Coll = L_RightTurnAgainst & Prec = FailGiveWayOther	FirstIntAct = Car	0.132	1.000	1.357
13	Coll = L_RightTurnAgainst & Light = DayNSL	FirstIntAct = Car	0.132	1.000	1.357
14	Coll = L_RightTurnAgainst & Light = DaySLUnk	RdType = SingCgw	0.105	1.000	1.727
15	Coll = L_RightTurnAgainst & SpdLim = 40-50 mph	Surf = Dry	0.105	0.800	1.448
16	Coll = L_RightTurnAgainst & DrvInj = Uninjured	Surf = Dry	0.105	0.800	1.448
17	Coll = L_RightTurnAgainst & Prec = FailGiveWayOther	Surf = Dry	0.105	0.800	1.448
18	Coll = L_RightTurnAgainst & Light = DayNSL	Surf = Dry	0.105	0.800	1.448
19	Coll = L_RightTurnAgainst & Light = DaySLUnk	TrfCtrl = None	0.105	1.000	1.357
20	Coll = L_RightTurnAgainst & Area = Urban	FirstIntAct = Car	0.105	1.000	1.357
21	Coll = L_RightTurnAgainst & Light = DaySLUnk	HorizGeom = Straight	0.105	1.000	1.267
22	Coll = L_RightTurnAgainst & Area = Urban	HorizGeom = Straight	0.105	1.000	1.267
23	Coll = L_RightTurnAgainst & Surf = Wet	HorizGeom = Straight	0.105	1.000	1.267

with 40–50 mph speed limit (see rule nr. 27/28), wet surface (rule nr. 29) and single carriageway (rule nr. 30). However, this set of rules show that there is no clear indication on some attributes, such as the road type, as “RdType = DualCgw” is among the frequent items (see rules 42–45). Also, the driver can be uninjured or seriously injured or the area can be urban or rural. Those varying attributes could be used as varying parameter in the virtual simulation, while the others constitute the “static” environment and situation.

While the rules in the tables are relatively easy to interpret, this is no more the case with 4-, 5- or 6-item rules, also due to the high number of obtained rules. Therefore, each set of rules (comprising 2- to 6-item rules) was further visualized by directed graphs that were created from adjacency matrices of the associations found between all attributes. The graph was then reduced to the edges that direct to a certain consequent, represented by edge tables including source, target and weight of the edges. In this case, the targets (or consequents) were the collision types L (see Fig. 4) and J (see Fig. 5) and the sources were all remaining attributes. The weight or thickness of each edge represents the amount of associations identified between the respective antecedent node and the given consequent in the centre. In other words, nodes with thick edges indicate dominant crash attributes and thus define the scenario. For antecedent nodes that are not present in the graph, there were no associations found in the rules, thus they can be considered negligible for the respective scenario. Note that the graph does not reflect support, confidence or lift.

By visually inspecting the graphs and rules tables, the scenarios for this cluster can be described as follows (note that all crashes in the data occurred on UK roads with left-hand traffic):

Scenario T-10.1 (related to collision type L): Car *A* goes straight on a major road and hits another car *B* with its front, which is coming from the opposing direction and is turning right into a minor road. This happens on a single carriageway with a speed limit of 40 mph or 50 mph at an unsignalized junction, and is caused by *B* failing to give way. The surface is dry and *B* suffers serious or fatal injury.

Scenario T-10.2 (related to collision type J): Car *A* goes straight on a major road and hits another car *B*, which is emerging from a minor road on the left with the intention to turn right. This happens on a single carriageway in a rural area with a speed limit of 40 mph or 50 mph at an unsignalized junction, and is caused by *B* failing to give way. The surface is wet and *A* suffers serious injury.

The same procedure was applied to the other clusters and their collision types. Figs. 6 and 7 illustrate all high-injury scenarios identified in a simplified manner to better understand the descriptions in the text. The red dots in the figures are the points of impact (i.e. front, offside or nearside). Surface conditions, area (rural,urban), speed limits, vehicle types and injury levels are not shown, but described in the following from the perspective of car *A*, i.e. the ego car associated with each sample.

Scenario T-4.1: Car *A* turns into a minor road and is hit by a PTW *B* on its nearside, which is going straight in the opposing direction. This happens on a single carriageway with 40–50 mph speed limit without active or static yield instruction and is caused by *A* failing to give way or manoeuvring inappropriately.

Scenario T-12.1: Car *A* turns right into a major road and is hit by a PTW *B* on the offside, which is going straight on the crossing path. This happens on a rural single carriageway controlled by a static give-way sign and is caused by *A* failing to give way. The surface is wet and *B* suffers serious or fatal injury.

Scenario T-12.2: Car *A* turns right into a minor road and is hit on the offside by a PTW *B*, which is overtaking. This happens on an urban single carriageway with 30 mph speed limit without active or static yield instruction and is caused by an inappropriate overtake from *B*.

Scenario T-12.3: Car *A* turns left into a major road and is hit by a PTW *B* on its offside, which is going straight on the major road from the right. This happens on an urban single carriageway with 30 mph speed limit controlled by give-way signs and is caused by *A* failing to give way. *B* suffers serious or fatal injury.

Scenario T-13.1: Car *A* turns into a minor road and hits a PTW *B* with its front, which is going straight in the opposing direction. This happens on a rural single carriageway with 30 to 50 mph speed limit without active or static yield instruction and is caused by *A* failing to give way or manoeuvring inappropriately. The surface is wet and *B* suffers serious or fatal injury.

Scenario X-1.1: Car *A* goes straight on a major road and hits another car *B* with its front, which is crossing the path from the left. This happens on a rural single carriageway with 60 mph speed limit without active or static yield instruction and is caused by *B* failing to give way.

Scenario X-2.1: Car *A* comes from a minor road and goes straight over a four-legged junction and hits another car or PTW *B* with its front, which crosses the path from the right. This happens on a rural road with 40–50 mph speed limit controlled by static give-way signs and is caused by *A* failing to give way.

Table 6
Rules obtained for T-C10 with collision type J, sorted by the five highest support values.

Nr.	Antecedent	Consequent	Supp	Conf	Lift
1	Coll = J_CrossingVehTurning	Prec = FailGiveWayOther	0.211	0.800	2.338
2	Coll = J_CrossingVehTurning	Light = DayNSL	0.211	0.800	1.520
3	Coll = J_CrossingVehTurning & Light = DayNSL	HorizGeom = Straight	0.211	1.000	1.267
4	Coll = J_CrossingVehTurning & HorizGeom = Straight	Light = DayNSL	0.211	1.000	1.900
5	Coll = J_CrossingVehTurning & FirstIntAct = Car	Prec = FailGiveWayOther	0.184	0.875	2.558
6	Coll = J_CrossingVehTurning & TrfCtrl = None	Prec = FailGiveWayOther	0.184	0.875	2.558
7	Coll = J_CrossingVehTurning & Area = Rural	Light = DayNSL	0.184	1.000	1.900
8	Coll = J_CrossingVehTurning & Area = Rural	HorizGeom = Straight	0.184	1.000	1.267
9	Coll = J_CrossingVehTurning & Prec = FailGiveWayOther	Surf = Wet	0.158	0.750	1.781
10	Coll = J_CrossingVehTurning & Surf = Wet	Prec = FailGiveWayOther	0.158	1.000	2.923
11	Coll = J_CrossingVehTurning & RdType = SingCgw	Prec = FailGiveWayOther	0.158	0.857	2.505
12	Coll = J_CrossingVehTurning & Area = Rural	Prec = FailGiveWayOther	0.158	0.857	2.505
13	Coll = J_CrossingVehTurning & Surf = Wet	Light = DayNSL	0.158	1.000	1.900
14	Coll = J_CrossingVehTurning & Light = DayNSL	Surf = Wet	0.158	0.750	1.781
15	Coll = J_CrossingVehTurning & Surf = Wet	Area = Rural	0.158	1.000	1.407
16	Coll = J_CrossingVehTurning & Area = Rural	Surf = Wet	0.158	0.857	2.036
17	Coll = J_CrossingVehTurning & Surf = Wet	HorizGeom = Straight	0.158	1.000	1.267
18	Coll = J_CrossingVehTurning & HorizGeom = Straight	Surf = Wet	0.158	0.750	1.781
19	Coll = J_CrossingVehTurning & TrfCtrl = None	RdType = SingCgw	0.158	0.750	1.295
20	Coll = J_CrossingVehTurning & DrvInj = Serious	Prec = FailGiveWayOther	0.105	1.000	2.923
21	Coll = J_CrossingVehTurning & SpdLim = 30 mph	Area = Urban	0.079	1.000	3.455
22	Coll = J_CrossingVehTurning & Area = Urban	SpdLim = 30 mph	0.079	1.000	3.800
23	Coll = J_CrossingVehTurning & SpdLim = 30 mph	Surf = Dry	0.079	1.000	1.810
24	Coll = J_CrossingVehTurning & Surf = Dry	SpdLim = 30 mph	0.079	0.750	2.850
25	Coll = J_CrossingVehTurning & SpdLim = 30 mph	RdType = SingCgw	0.079	1.000	1.727
26	Coll = J_CrossingVehTurning & SpdLim = 30 mph	TrfCtrl = None	0.079	1.000	1.357
27	Coll = J_CrossingVehTurning & DrvInj = Serious	SpdLim = 40–50 mph	0.079	0.750	2.375
28	Coll = J_CrossingVehTurning & SpdLim = 40–50 mph	DrvInj = Serious	0.079	1.000	3.455
29	Coll = J_CrossingVehTurning & DrvInj = Serious	Surf = Wet	0.079	0.750	1.781
30	Coll = J_CrossingVehTurning & DrvInj = Serious	RdType = SingCgw	0.079	0.750	1.295
31	Coll = J_CrossingVehTurning & Area = Urban	Surf = Dry	0.079	1.000	1.810
32	Coll = J_CrossingVehTurning & Surf = Dry	Area = Urban	0.079	0.750	2.591
33	Coll = J_CrossingVehTurning & Area = Urban	RdType = SingCgw	0.079	1.000	1.727
34	Coll = J_CrossingVehTurning & Area = Urban	TrfCtrl = None	0.079	1.000	1.357
35	Coll = J_CrossingVehTurning & SpdLim = 40–50 mph	Prec = FailGiveWayOther	0.079	1.000	2.923
36	Coll = J_CrossingVehTurning & SpdLim = 40–50 mph	Surf = Wet	0.079	1.000	2.375
37	Coll = J_CrossingVehTurning & SpdLim = 40–50 mph	Light = DayNSL	0.079	1.000	1.900
38	Coll = J_CrossingVehTurning & SpdLim = 40–50 mph	Area = Rural	0.079	1.000	1.407
39	Coll = J_CrossingVehTurning & SpdLim = 40–50 mph	HorizGeom = Straight	0.079	1.000	1.267
40	Coll = J_CrossingVehTurning & DrvInj = Uninjured	Light = DayNSL	0.079	1.000	1.900
41	Coll = J_CrossingVehTurning & DrvInj = Uninjured	HorizGeom = Straight	0.079	1.000	1.267
42	Coll = J_CrossingVehTurning & RdType = DualCgw	Light = DayNSL	0.079	1.000	1.900
43	Coll = J_CrossingVehTurning & RdType = DualCgw	Area = Rural	0.079	1.000	1.407
44	Coll = J_CrossingVehTurning & RdType = DualCgw	FirstIntAct = Car	0.079	1.000	1.357
45	Coll = J_CrossingVehTurning & RdType = DualCgw	HorizGeom = Straight	0.079	1.000	1.267
46	Coll = J_CrossingVehTurning & Surf = Dry	RdType = SingCgw	0.079	0.750	1.295

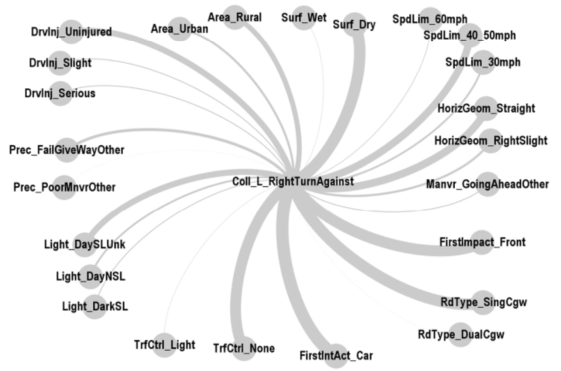


Fig. 4. Weighted, directed graph obtained from all association rules for cluster T-C10 having collision type L as consequent.

Scenario X-4.1: Car *A* turns right into a major road and is hit by a car or LGV *B* on the offside, which is going straight on the major road from the right. This happens on a rural dual carriageway with 40–50 mph speed limit controlled by static give-way signs and is caused by *A* failing to give way. The surface is wet and *A* suffers serious or fatal injuries.

Scenario X-6.1: Car *A* goes straight on a major road and is hit by car *B* on the offside, which comes from a minor road and crosses the path from the right. This happens on a single carriageway road with 30 mph speed limit controlled by traffic lights and is caused by *B* failing to give way. The surface is wet and *B* suffers serious or fatal injuries.

Scenario X-6.2: Car *A* goes straight on a major road and is hit by car *B* on its offside, which turns right from the opposing direction. This happens on a road with 60 mph speed limit controlled by traffic lights and is caused by *B* loosing control of the vehicle. *B* suffers serious or fatal injuries.

7.4. Comparison with high-frequency scenarios

This section compares the high-injury scenarios to the most frequent scenarios identified. Figs. 8 and 9 depict the top five high-frequency scenarios for three-legged and four-legged junctions, i.e. the scenarios with the highest number of crashes included. Table 7 shows the crash counts for each scenario including a short description. Some of the three-legged junction scenarios were combined due to their similarities.

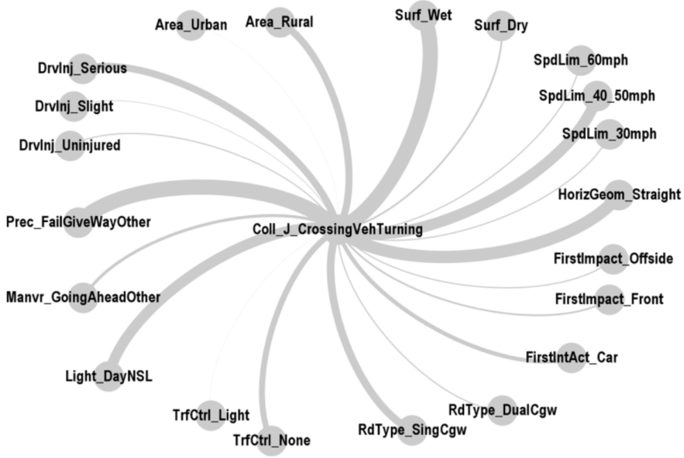


Fig. 5. Weighted, directed graph obtained from all association rules for cluster T-C10 having collision type J as consequent.

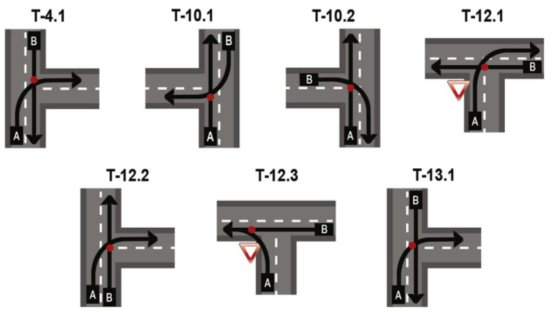


Fig. 6. Simplified illustrations of all high-injury scenarios identified for three-legged junctions.

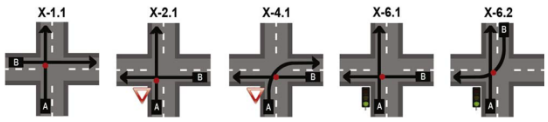


Fig. 7. Simplified illustrations of all high-injury scenarios identified for four-legged junctions.

Fig. 8. Simplified illustrations of the five most frequent scenarios identified for three-legged junctions.

Fig. 9. Simplified illustrations of the five most frequent scenarios identified for four-legged junctions.

For example, T-2.1 and T-8.1, which were derived from two different clusters, were grouped. This was also done for the second and third most frequent scenarios for three-legged junctions. The count column in the table gives the number of crashes within the respective cluster that are allocated to the particular collision type. For example, the 44 samples for T-1.1 are the collisions of type F (rear-end) within cluster T-C1.

It can be observed that the top five most frequent scenarios at four-legged junctions do not include rear-end collisions. This finding corresponds to the crossing-path scenarios identified by Najm et al. (2001), which are primarily angle crashes. Furthermore, there is no particular scenario involving car-pedestrian or car-bicycle collisions only. This can be explained by the low number of pedestrians (2.4%) and cyclists (3.6%) as collision partners, compared to other cars (71.8%), motorcycles (9.7%) or goods vehicles (7.8%).

The low number of vulnerable road users is also a reason why the high-frequency scenarios at three-legged junctions do not include any of the high-injury scenarios. However, the three-legged junction scenarios include two rear-end collisions (T-5.1/5.2 and T-1.1), which are not included in the high-injury scenarios. This is due to the fact that the injury outcome was found to be lower for rear-end collisions than for angle collisions, which was also reported by Beck (2015).

8. Discussion

8.1. Relation to existing findings

The pre-crash scenarios described above build the foundation for further research on testing assisted and automated vehicle technologies. This paper focussed on the scenarios with serious or fatal injury outcome, which were compared to the high-frequency scenarios. Although there is no doubt about the importance of vulnerable road user safety, neither the cluster analysis nor the association rule method resulted in a distinct pedestrian or cyclist scenario. Considering the frequency of certain crash types at junctions, car-pedestrian and car-cyclist collisions are discounted, which might not be true if injury frequencies were taken into account.

The method of clustering intersection crashes into distinct groups, including such a high number of variables as used in this study, is novel. Abdel-Aty et al. (2006) analysed numerous parameters to identify crash profiles for 45 different intersection configurations in Florida, however, this was made for different AADT values and numbers of lanes, which were not included in this study. Also, the objective of this study is different, because it aims at extracting relevant combinations of junction situations for simulation, while Abdel-Aty et al. (2006) provided crash profiles that assist in identifying intersections with specific problems. Therefore, the results cannot be directly compared.

dataset. In summary, the results support existing findings about junction safety and add further definition to the clusters identified. For example, as indicated in literature, higher injury levels coincide with powered two-wheelers involved as well as higher speed limits. The study is preparatory research to a sub-microscopic simulation study, where virtual test drives will be conducted and automated collision avoidance and mitigation systems will be evaluated under varying conditions. The scenarios obtained will help to reduce the possible number of model parameter variations, such as vehicle trajectories, velocities as well as road and junction parameters.

Acknowledgements

The research work for this paper was conducted in the scope of a joint PhD project between Loughborough University and the AIT Austrian Institute of Technology. The accident data in this study was acquired in cooperation with the UK Department for Transport that provided access to the RAIDS database. The Road Accident In Depth Studies (RAIDS) programme and associated database were commissioned by the United Kingdom Department for Transport in 2012 to consolidate data gathered from historic in depth collision investigation programmes dating back to the year 2000. Data collection is ongoing and since 2012, 1200 new cases have been investigated, the data is made available free of charge over the internet however conditional access is limited to those with a defined research need. For further information please contact RAIDS@dft.gov.uk.

Appendix A. Collision codes from STATS-19

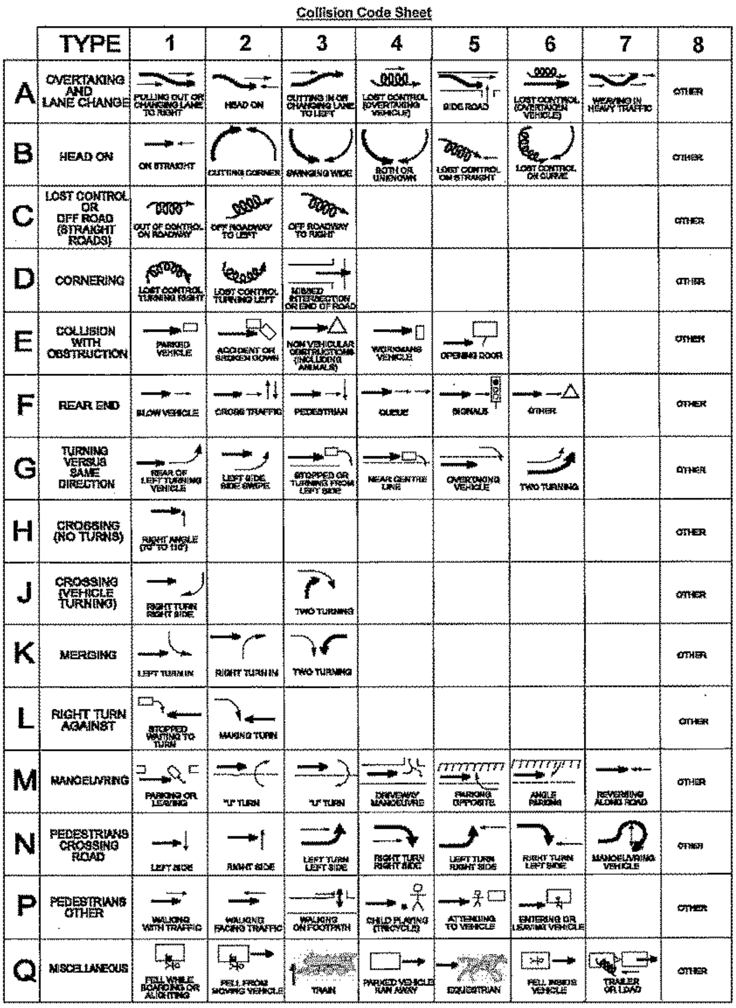


Fig. A10. Collision codes from STATS-19.