

Research Article

Leveraging the GIDAS Database for the Criticality Analysis of Automated Driving Systems

Stefan Babisch,¹ Christian Neurohr ,² Lukas Westhofen ,² Stefan Schoenawa,³ and Henrik Liers¹

¹VUFO GmbH, Semperstrasse 2A, 01069 Dresden, Germany

²German Aerospace Center (DLR) e.V., Institute of Systems Engineering for Future Mobility, Escherweg 2, 26129 Oldenburg, Germany

³Volkswagen Group, Berliner Ring 2, 38440 Wolfsburg, Germany

Correspondence should be addressed to Christian Neurohr; christian.neurohr@dlr.de

Received 10 August 2022; Revised 10 March 2023; Accepted 28 March 2023; Published 8 May 2023

Academic Editor: Long Truong

Copyright © 2023 Stefan Babisch et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A safe introduction of automated driving systems on urban roads requires a thorough understanding of the traffic conflicts and accidents. This understanding is paramount to constructively safeguard these systems, i.e., to design a system that exhibits an adequate performance even in critical situations. In this work, we present an approach to gather knowledge by analyzing the German In-Depth Accident Study (GIDAS) database, which is representative of all German traffic accidents, along with the influencing factors that are hypothesized to be associated with increased criticality in relation to automated driving. In order to gain an insight into the risk associated with these factors in real-world accidents, we determine their presence in the database's accident cases within a selected operational domain, enabled by translation from a natural language description to the database scheme employed by GIDAS. This initial catalog as well as the subsequent statistical considerations is motivated by analyzing the criticality for automated driving systems in urban areas. Based on this catalog, our work delineates a method for quantification of risk associated with such influencing factors in a given operational domain based on real-world accident data. This quantification can subsequently be used in decompositional, scenario-based risk assessment before system design and for the embedding safety argumentation. This paper, therefore, provides a blueprint of how the matured field of traffic accident research studies and its results, in particular accident databases, can be leveraged for risk assessment of the operational domain of automated driving systems.

1. Introduction

Demonstrating the safety of vehicles operated by an automated driving system (ADS), i.e., driving automation at SAE Level 3 or higher [1], is a challenging endeavor. For these levels of driving automation, ADS-equipped vehicles are safety-critical complex systems operating in and interacting with the open context provided by the operational domain (OD) [2], for which distance-based statistical approaches to safety can no longer be considered adequate [3, 4]. Mainly, this is due to an uncountably infinite amount of potentially safety-critical situations and scenarios [5], each of which is

slightly different and has a very low probability of occurrence. Scenario-based approaches to verification and validation of ADSs aim at eliciting a manageable set of scenario classes from which suitable representatives are derived in order to evaluate the ADS with finite effort [6]. However, the tasks of obtaining such a set of scenario classes, demonstrating its completeness, and assessing the risks within these classes, both for the ADS and for human traffic as reference values, are generally understood as key obstacles for the homologation of ADSs. The importance of these challenges led to the definition of the ISO 21448 [7], which is concerned with the Safety of the Intended Functionality (SOTIF). For

this, it prescribes the usage of methods to reduce both the number of known and unknown hazardous scenarios.

The ISO 21448 suggests the systematic identification of triggering conditions so as to reduce the set of unknown hazardous scenarios. Among the other methods [8–10], on a more abstract level, Neurohr et al. [11] proposed a criticality analysis as an approach. It decomposes the OD according to influencing factors that are associated with increased criticality, that is, the so-called criticality phenomena (CP). In order to elicit a manageable set of CP, the criticality analysis validates whether each factor is actually associated with criticality. Such a validated list can then be passed to subsequent steps in the safety case, such as risk assessment, causal analysis, or testing. For the CP that can be identified within GIDAS [12], these statistical results can be used to estimate the associated accident risk and as a starting point for subsequent causal analysis [13]. Moreover, this validated list shall be ordered by relevance so as to guide the efforts downstream into the most critical scenario classes. Of course, a well-founded validation of CP necessitates the consideration of a wide variety of traffic data.

In this regard, traffic accident databases present a valuable source of data to analyze. We would like to know which turn maneuvers occur more frequently in traffic accidents,

for example, turning left or right, at a T-junction or at an intersection. Based on this, we can investigate further and analyze which factors are associated with these turn maneuvers, for example, an “occluded bicyclist” for right turns. From this, we may derive a scenario class called “occluded bicyclist at right turn” based on the evidence from accident data. A safety case for the homologation of an ADS can benefit from using risk assessment within the OD as a reference value for the relative acceptance criteria such as a positive risk balance [14].

In particular, given an operational domain OD and a criticality phenomenon CP, we are interested in a formalization of a reference value for relative acceptance criteria. Besides this, such a value can also be used as a starting point in system design, for example, to exclude phenomena from the operational design domain (ODD) [1] or to guide resource allocation for design, implementation, and testing efforts. To achieve this formalization, in accordance with the ISO 26262 [15] and 21448 [7], we factorize the risk of an accident with passenger car involvement and damage to persons into a product of exposure, controllability, and severity followed by an application of Bayes’ theorem to the first two factors to obtain the following equation:

$$\begin{aligned} & \text{Risk (CP, Accident, Severity|OD)} \\ &= \frac{P(\text{CP|OD})}{\text{Exposure}} \frac{P(\text{Accident|CP,OD})}{\text{Controllability}} \frac{P(\text{Severity|Accident,CP,OD})}{\text{Severity}} \\ &= P(\text{CP|Accident, OD})P(\text{Accident|OD})P(\text{Severity|Accident, CP, OD}), \end{aligned} \quad (1)$$

which can be assessed more easily compared to the standard risk decomposition in the second line of equation (1), as follows:

- (i) $P(\text{CP|Accident, OD})$ can be estimated by analyzing the CP in GIDAS
- (ii) $P(\text{Accident|OD})$ is obtainable from the national accident statistics
- (iii) $P(\text{Severity|Accident, CP, OD})$ can be fetched from severity entries in GIDAS

In contrast, for the standard risk decomposition into exposure, controllability, and severity, two problems arise:

- (i) Estimating the quantity $P(\text{CP|OD})$, which requires large-scale, representative, naturalistic driving data from a nonbiased measurement source to estimate frequencies of CP
- (ii) Estimating the quantity of $P(\text{Accident|CP, OD})$, which moreover requires this naturalistic data to include a representative sample of accidents

Hence, the proposed decomposition of equation (1) enables the risk estimation for CP in practice by solely relying on accident databases and accident statistics from the OD. In this publication, based on this decomposition, we instantiate a risk assessment for a catalog of CP in urban areas using real-world accident data from the German In-Depth Accident Study (GIDAS) database. As a preparatory step, the CP-catalog is translated into the query language of the GIDAS database scheme. Based on the results of filtering the database according to the translated CP, we are able to estimate the quantities of equation (1), which can be used two-fold: (a) for decompositional risk assessment of CP-associated scenario classes in a safety case and (b) for ordering CP according to their relevance within a criticality analysis.

As ADS-equipped vehicles (AVs) were not yet represented in the GIDAS database in the year of analysis (2020), the results can be used within a criticality analysis for AVs only with caution. Therefore, we focus on the CP that are relevant for AVs on a vehicle-level, as determined by an expert judgment and leave the CP that are specific to the machine-perception of ADSs for analyses based on more

suitable data sources. However, the performed analysis provides a blueprint for future analyses of accident databases for AVs. For this, requirements on the collection of accident data involving AVs can be derived from our analysis. The main contributions of this work are as follows:

- (i) A method for risk assessment of scenario classes spanned by CP based on accident databases and (OD-level) accident statistics
- (ii) Identification of CP within a representative accident database
- (iii) A blueprint for translating a linguistic specification of CP to a format that enables their recognition in data
- (iv) A large dataset relating accident cases to the presence of CP as a direct result of translating the CP-catalog into the GIDAS database query language
- (v) Estimation of quantities required for the proposed assessment of CP-associated accident risk
- (vi) Accident risk assessment for combinations of CP

The introduction of Section 1 is followed by the preliminaries in Section 2, which contains brief descriptions of the criticality analysis of the GIDAS database and a discussion of related work, respectively. Section 3 presents the preparatory steps for the analysis, that is, obtaining a catalog of CP, selecting the relevant subset of GIDAS accidents, and translation of the CP to the database scheme. In Section 4, we evaluate the corresponding database queries to obtain a large binary dataset linking CP to GIDAS accident cases. Based on this relation matrix, we delineate how to estimate the risk-related quantities of equation (1) for a single CP and for combinations of CP. After a discussion of the results in Section 5, we conclude the work at hand and provide an outlook to future work in Section 6.

2. Preliminaries

The previously sketched risk decomposition serves as our main motivation, but it obviously relies on a given catalog of CP. For this work, such a catalog is given by conducting a criticality analysis, a method developed in the project “verification and validation methods” (VVM) (<https://www.vvm-projekt.de/en>). In this section, so as to provide the adequate context, we briefly introduce the concept of the criticality analysis and the GIDAS database, and consider relevant related works.

2.1. Criticality Analysis for the Verification and Validation of Automated Driving Systems. The VVM project aims at developing methods and processes for the verification and validation of AVs at SAE Levels 4 and 5 [1]. In this regard, previous works by the authors proposed a methodical criticality analysis to structure the open context in which AVs are supposed to operate by eliciting a finite and manageable set of artifacts [11]. Neurohr et al. define criticality of a traffic situation as the combined risk of the

involved actors when the traffic situation is continued ([11], Definition 1). Note that this actor-agnostic definition of criticality can be extended from situations to scenarios in multiple ways by aggregating the criticality of a time sequence, for example, by choosing the maximum or some quantile (see Section 5.2 in [16]).

A fundamental concept of the criticality analysis is the observable influencing factors in the traffic world that are associated with increased criticality, called the criticality phenomena (CP) (Definition 2 in [11]). In terms of the terminology used for hazard analysis and risk assessment, CP represents abstract classes of danger (Remark 2 in [11]) and can thus be used as a template in subsequent hazard analyses. Moreover, as motivated in the introduction, their associated risk may serve as a reference value for accepting or rejecting the assessed system-dependent risk for homologation. As a concept, CP does encompass more than just direct causes of accidents, cf. ([17], pp. 13–15) as, by definition, only a statistical association with criticality is required. Therefore, these influencing factors may directly increase criticality, appear at the beginning of a causal chain of events, or even represent a spurious criticality association.

A quite prominent set of examples of CP is given by the abstract spatial phenomenon of “occlusion” of an object by another for a given subject and its numerous concretizations such as “occluded vehicle” or “occluded traffic sign.” Also, let us already mention the two phenomena “nonego-traffic participant (TP) violating right of way” and “intersecting planned trajectories of TPs” which, together with “Occlusion,” will serve as a running example for a set of CP throughout this work.

As shown in Figure 1, the associative part of the method branch consists of the steps “identification and formalization of criticality phenomenon” and “estimation of criticality association,” followed by a potential “abstraction, refinement, or discardment” step, depending on the established associative criticality. The first step requires input from the knowledge basis. It can be structured as follows (Section V.A.1 in [11]):

- (1) Acquire and structure knowledge
- (2) Search the available knowledge basis for observations associated with criticality, including knowledge from laws, guidelines, and experiences
- (3) Describe each CP using an ontological basis, including abstractions and concretizations, an ontological classification, and relations to other phenomena

The second step, “estimation of criticality association,” generates evidence for the associational validity of a CP from (4). If evidence is not sufficient, one can readjust the level of abstraction or discard the phenomenon. For each identified and relevant CP, the causal part of the method branch starts an extensive causal inquiry into the established criticality association [13]. Therefore, before spending resources on causal analysis, the relevance estimation shall be backed by strong empirical pieces of evidence. We highlight that the results of this

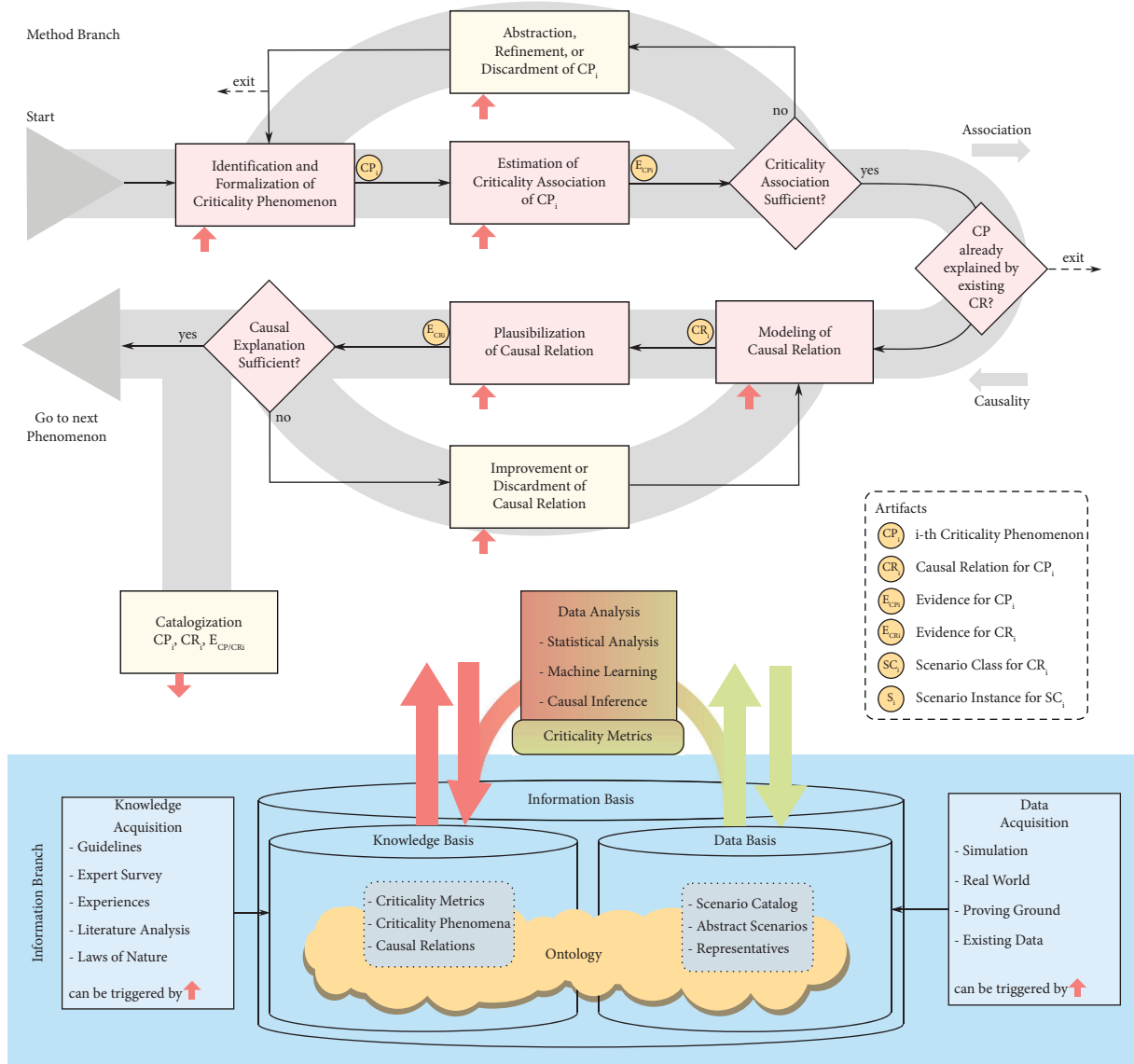


FIGURE 1: The method and information branches of the criticality analysis.

work can be used for such an estimation, specifically by ranking CP based on the results of equation (1).

Establishing confidence in the criticality association of a considered phenomenon prerequisites criticality to be measurable. To this end, criticality metrics can be employed, with the goal to evaluate the criticality of a situation or scenario numerically (Section V.A.6 in [11]). Since the 1970s, such conflict indicators were developed to allow not only to observe accidents but also to observe other serious conflicts leading to near misses [18] (Section 2 in [16]). Reversely, observing an actual accident can also be seen as an exceptional traffic conflict, namely, the most serious one. Accidents are, therefore, a subset of all critical situations identified by criticality metrics.

The accident metric (<https://criticality-metrics.readthedocs.io/en/latest/index-scale/AM.html>) is a qualitative, binary-valued

scenario-level criticality metric that evaluates to 1, if an accident did happen in a scenario, and 0 otherwise. Accident databases can thus be interpreted as collections of scenarios for which the corresponding accident metric equals to 1. Hence, the statistical analysis of an accident database regarding the presence of a CP yields empirical evidence for the relevance of the said phenomenon, that is, when the criticality is measured using a variation of the accident metric. While the accident metric has a perfect specificity for identifying scenarios as critical, its sensitivity is low, as all critical nonaccident scenarios are excluded by definition. In order to draw reliable statistical statements from analyses of accident databases alone, large and qualitative datasets are thus required. Fortunately, as explained in Section 2.2, the GIDAS database provides us a set of high-quality accident data that is considered to be representative for the total German accident statistics.

2.1.1. Limitations. In the VVM project, the criticality analysis is conducted for a generic class of AVs at SAE Levels 4 and 5 in an urban environment [11]. As suggested by Damm and Galbas [19], the process step “identification and formalization of criticality phenomenon” has been divided into two parts, namely, identification of CP that

- (i) Are relevant to human traffic, for example, not dependent on the perception of the environment through a sensor setup
- (ii) Become relevant for AVs that rely on sensor technology for perception

As the GIDAS database did not (yet) contain accidents involving AVs at the time of analysis, it has been restricted to CP identified in part (i). It does, however, provide a blueprint for future analyses for part (ii).

Another limitation concerns the legal dimension of traffic accidents. The criticality analysis is not concerned with questions of guilt or liability, so we do not require such information within the analyzed accident cases.

2.2. The GIDAS Database. GIDAS (<https://www.gidas.org/start-en.html>) is the largest in-depth accident study in Germany with its adjacent database being among the world’s leading traffic accident databases. Since 1999 the GIDAS project has collected on-site accident cases in the areas of Hanover and Dresden. Due to a well-defined statistical sampling plan, cf. Section 2.2.2, representative statements about the German national accident statistics are possible. GIDAS collects data from all types and kinds of traffic accidents with personal injury. For each accident case, or simply case, GIDAS collects on an average about 3 500 pieces of information that are coded in the database, including data on the involved vehicles and persons, occurred injuries, surrounding infrastructure, and environmental conditions. Moreover, every accident is reconstructed by experienced reconstruction engineers so as to obtain knowledge about speeds, steering or braking maneuvers, and collision parameters. As the project is funded by the German Federal Highway Research Institute (BAST) and the German Research Association of Automotive Technology (FAT), access to the data is restricted to GIDAS consortium members. While Figure 2 gives an overview of the total content of the GIDAS database, for the analysis presented in the following sections, a subset of the whole content has been selected.

2.2.1. The GIDAS Codebook. In order to provide the members and partners of GIDAS with an adequate overview of all the information present in the database, there exists a corresponding codebook containing various pieces of information for each variable that is coded for in GIDAS. For every variable used by GIDAS, the codebook contains an entry containing, among others, the name and a description of the variable and a range of possible labels. As of June 30, 2020, more than 2 400 individual variables are encoded in

GIDAS and documented in the codebook. An example is shown in the following dashed box for the variable “ORTSL,” which provides information about the location of the accident:

Variable: ORTSL (accident site, ORTSLAGE in German)

Record: UMWELT (environment)

Label: accident site

Valid date period: since 1999-07-01

Mandatory variable: yes

Description: the accident site relates to the official record of the location of the accident, in particular whether the accident is located inside or outside of a built-up area. If the accident scene is on the boundary line, the location of the collision point should be indicated. This information must always correspond with police records.

Defined labels:

3: urban

4: rural without highway

5: highway

This example highlights that, due to the intensive memory requirements of textual entries, many specific characteristics of cases are stored as numerical codes in the database. Therefore, an interpretation of the cases is only possible in conjunction with this codebook (or with extensive experience working with the GIDAS database).

Moreover, let us mention a special set of variables inspired from the catalog of accident causes from DESTATIS ([17], p. 50) which has been used frequently in this work. There are many variables in GIDAS relating to accident causes such as “HURSU,” which not only encodes the hypothesized main cause for the whole accident but also lists several (partial) causes for each involved traffic participant, named “URSWIS1,” . . . , “URSWIS4.” The variables “HURSU” and “URSWIS” have, among others, been applied to identify CP using SQL queries. For most of the CP, however, a combination of multiple GIDAS variables was required in order to identify them in the database, cf. Appendix A.

2.2.2. Weighting and Representativeness. In order to account for the bias within the database and to ensure the representative results, the GIDAS database is weighted towards the German national accident statistics DESTATIS of 2019 [17]. Bias within the data is due to multiple reasons, including investigation teams not being informed about all accidents within the investigation area, information about injuries not always being immediately available, or noneliminable differences in the investigation areas (cf. Section 2.2.3). By weighting the data, GIDAS results can be used for statements over the total German accident statistics. The weighting of the GIDAS data is based on the following three variables:

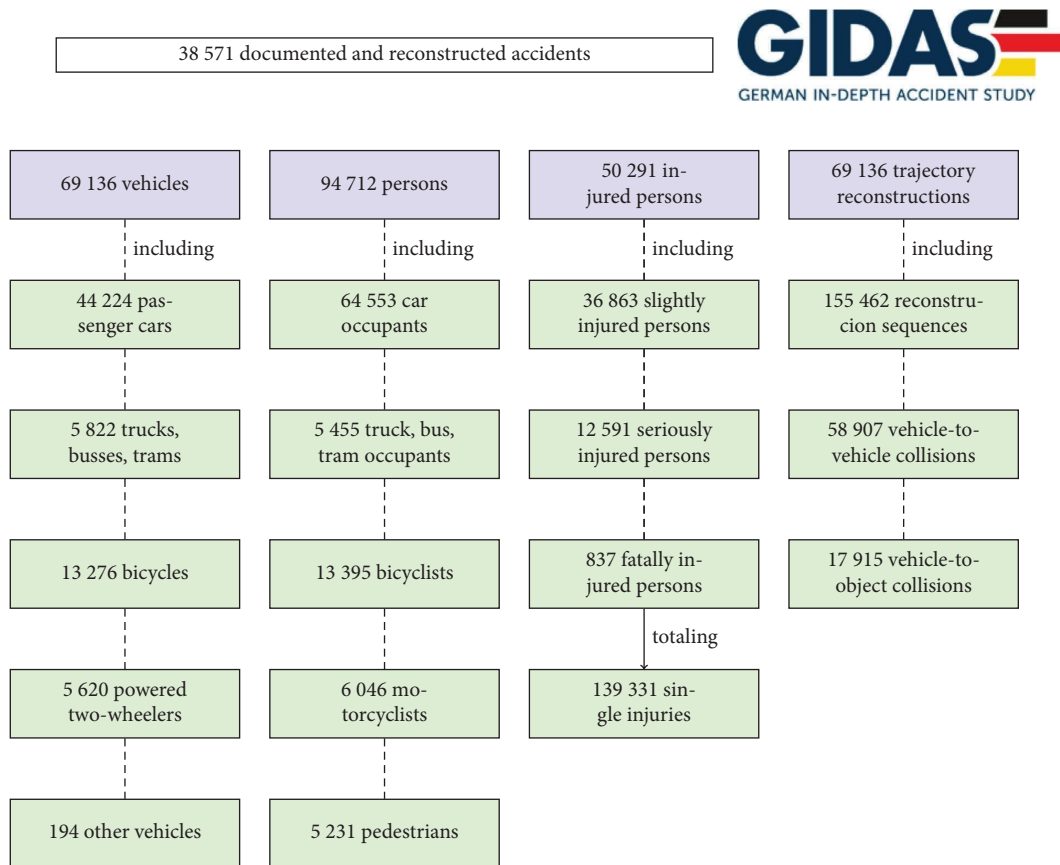


FIGURE 2: An overview of the contents of the GIDAS database effective from June 30, 2020.

- (i) Accident site (“ORTSL”), i.e., urban, rural, or highway
- (ii) Accident severity (“PVERL”) according to DESTATIS ([17], p. 40), i.e., accidents with slightly injured persons, seriously injured persons, or fatally injured persons
- (iii) Type of accident (“UTYP”), which distinguishes seven different categories according to DESTATIS ([17], p. 36)

After the application of the weighting procedure, the database and the results are representative for these variables. In addition, other aspects are indirectly corrected by the weighting procedure, for example, the distributions of road users or time of the day. Depending on these variables, for each accident case i , the weighting factor w_i and the extrapolation factor e_i are calculated as

$$w_i = \frac{\# \text{Acc. in Germany with UTYP}_i, \text{PVERL}_i, \text{ORTSL}_i \cdot \# \text{Acc. in GIDAS}}{\# \text{Acc. in Germany} \cdot \# \text{Acc. in GIDAS with UTYP}_i, \text{PVERL}_i, \text{ORTSL}_i}, \quad (2)$$

$$e_i = \frac{\# \text{Acc. in Germany with UTYP}_i, \text{PVERL}_i, \text{ORTSL}_i}{\# \text{Acc. in GIDAS with UTYP}_i, \text{PVERL}_i, \text{ORTSL}_i}.$$

Since the weighting factors w_i range from 0.279 to 2.084, the weighted sum of accidents (or persons, injuries etc.) may amount to nonintegers. Therefore, the frequencies that rely on these factors are rounded to the closest integer, for example, in Figure 2, the introduced rounding errors being negligible.

2.2.3. Measurement Bias. As with any dataset, we have to be cautious about the possible biases introduced by the measurement principle. As the data rely on the reports of the onsite team and police created after the accident, it is only possible to approximate the exact happenings of the accident scenario. This may introduce biases by (a) limiting which CP

can be identified in GIDAS and (b) underestimating the measured frequencies of those CP that can be identified. As an example of the first case, CP such as emotional states or lines of the gaze of traffic participants may be nonidentifiable within the GIDAS, and this can be partially mitigated by questioning the traffic participants. However, they may (consciously or unconsciously) incorrectly reconstruct the course of events. For the second case, the presence of defective vehicle parts (for example, headlights) may be underestimated if some collisions alter the physical state of these parts significantly (for example, headlights are destroyed in frontal collisions). However, by relying on experienced onsite teams, such biases can be limited, especially when compared to other data sources. For example, data from camera-equipped drones may be completely biased against adverse weather conditions, which is not the case for GIDAS.

2.3. Related Work. As the work at hand lies between the areas of classical traffic accident research studies and the engineering of ADSs, the related work is analyzed on both the aspects: (a) accident databases and their use for accident research studies and prevention as well as (b) incident databases of ADSs and their exploitation in a corresponding safety case.

2.3.1. Human Road Traffic. At this point, let us mention the vast amount of naturalistic driving study (NDS) databases, such as the SHRP2 [20] and UDRIVE [21]. Those, albeit delivering fruitful insights into the nature of human road traffic, are only slightly related to the scope of this work which focuses on the extreme traffic situations in the form of traffic accidents with damage to persons. Note that, those can also be extracted from sufficiently-sized NDS databases by means of identifying collisions or near-collisions [22].

(1) Accident Databases. Many countries collect and aggregate accident census data such as age and type of involved participants. For example, while Japan maintains a huge and all-inclusive accident database (ITARDA) (<https://www.itarda.or.jp/English>), in the United States of America the National Highway Traffic Safety Administration (NHTSA) operates a variety of accident data collection systems (<https://www.nhtsa.gov/data/crash-data-systems>). In particular, let us mention the Fatality Analysis Reporting Systems (FARS) (<https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>), which collect data on fatal traffic accidents, and the National Automotive Sampling System (NASS) (<https://www.nhtsa.gov/crash-data-systems/national-automotive-sampling-system>), from which representative samples of accident data can be sourced. The International Traffic Safety Data and Analysis Group (IRTAD) aggregates such data internationally [23, 24].

Ziegler et al. present an overview on the various national data sources, such as the police accident investigations, in-depth studies, and detailed case recordings for scientific applications, so as to assess their quality [25]. Within

Germany, GIDAS [12], examined in this work, falls into the second category, i.e., it contains detailed data from a representative sample of accidents collected by onsite investigation teams. To the best of the authors' knowledge, there are no other published studies which systematically identify safety-critical influencing factors for automated driving within GIDAS.

The previous work of Esenturk et al. [26, 27] analyzed the UK accident data collection STATS19 (<https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>) with the goal of generating valuable test scenarios for ADSs. In particular, the authors relate a list of 17 variables, each with a very limited set of values, to the severity of accident data [26]. They learn a logistic regression model on this data that is able to predict the accident severity on a binary scale, i.e., slight or severe, quite reliably and argue, that this model can be used for systematic test case generation. The work at hand is similar in that we also aim at leveraging accident data for ADSs safety. However, the scope of our work is vastly more encompassing concerning the methodical background for elicitation of safety-relevant factors given by the criticality analysis. In particular, after establishing associational relevance, the criticality analysis suggests causal inference techniques to avoid deriving scenarios with spurious criticality correlations [13]. Moreover, the broader view also concerns the amount of variables, the source of data, and the variety of performed statistical analyses. Instead of solely targeting the derivation of test scenarios, we leverage the accident data for the estimation of quantities of interest generally relevant within the verification and validation processes.

(2) Analysis for Accident Research Studies and Prevention. The close examination of accident cases can lead to a deepened understanding of their influencing factors [28], which in turn provides guidance to the road traffic safety measures [29]. In particular, accident databases have been leveraged for this purpose, without a specific focus on ADSs, for example, for accidents involving vulnerable road users [30].

2.3.2. Automated Driving Systems

(1) Incident Databases. Testing automated vehicles on public roads can be accompanied by the responsible companies disclosing details on potential incidents, such as disengagements (the safety driver having to take over control) or collisions (with static objects or traffic participants). These incidents can be analyzed to infer statements on causes, what led to the incident?, and outcome, what was the observable effect of the incident?

The state of California requires the reporting of such disengagements, which therefore have been subjected to various analyses. One of the first studies, a descriptive analysis of the disengagements between 2014 and 2015, was performed by Vinayak et al. [31]. For the twelve analyzed collisions, they found that in eleven cases, the other traffic participants were at fault. Regarding the reasons behind

disengagements, system failures made up 56%. Factors from the environment of the vehicle, such as weather and other traffic participants, accumulated to 17%. Another descriptive analysis of the reports from 2014 to 2017 found, similar, system failures to be mainly responsible for disengagements (52%). Again, factors from the vehicle's environment were contributing to 11% of the cases, containing issues such as poor lane markings (43%), construction zones (21%), and heavy pedestrian traffic (19%) [32]. Further comparisons of the recent work have been aggregated by Sinha et al. [33]. Based on these data, statistical analyses have been executed in order to estimate the relevance of factors behind disengagements, including decision tree learning [34] and regression models [35].

These approaches differ from the work at hand in that they operate only on small sample sizes, i.e., they are not representative for all critical situations within the OD. This concerns, for example, certain weather conditions such as snow and ice in the winter months. Adding to this issue, the subjects are not "skilled" driving functions but rather early implementations of being actively under development. This implies that many issues of functional safety (in the sense of ISO 26262 [15]) are likely to be present. For example, software faults, underspecification of requirements, or hardware faults are likely to cause misbehavior that leads to disengagements in situations in which a skilled driver would have performed flawlessly. The work at hand is concerned with factors influencing the safety of the intended functionality ("triggering conditions" in the sense of ISO 21448 [7]), and is therefore focused on the situations in which even a skilled driver will be challenged. Albeit analyzed within the incident studies, that these only make up for a smaller part, i.e., 11% to 48%, depending on the definition [31, 32, 34]) of the total disengagements, leading to the aforementioned issues in the sample size. Furthermore, disengagement reports are highly dependent on the manufacturers' willingness to disclose all relevant information as well as to analyze the incidents to a sufficient depth. As shown by Favaro et al., as of 2017, this has often led to quality issues when extracting information from those reports [32], for example, weather information was missing in roughly three-quarters of the cases. Moreover, basing an analysis of the safety-relevant factors within the OD on disengagement reports from a diverse set of companies still leaves open comparability issues including unaligned semantics of the factors, for example, what exactly constitutes a "bicyclist." For such factors, we suggest that by analyzing large-scale accident data with a unified semantical basis and an in-depth reconstruction performed by professionals can fill the aforementioned gaps of representativeness, comparability, and completeness.

(2) *Usage in Safety Cases.* The performance of driving automation can be compared to human performance in order to assess the performance of the technology that is to be introduced. For advanced driver assistant systems, this has, for example, been performed by means of NDS data [36]. Such a comparison has to be made against a representative database of human performance, for example, GIDAS or SHRP2. For more complex ADSs, a first comparative

analysis has been conducted by Goodall [37], where ADS rear-crash accidents according to the aforementioned incident reports were compared to human performance as measured in SHRP2. As risk assessment of such coarse scenario classes can lead to the so-called approval trap [3], so our work extends this comparative idea towards generic scenario classes. Thus, we lay the foundation for a fine-grained scenario-based comparison of the risk induced by human drivers with the risk induced by ADSs.

3. Preparation of Analysis

In this section, we describe the preparatory steps required for conducting an analysis of the GIDAS database. In particular, these are the following:

- (1) Create a catalog of CP in relation to ADSs within a selected OD, cf. Section 3.1, as a part of the criticality analysis
- (2) Select a subset of the database that corresponds to the class of systems and the OD for which risk assessment is to be performed, cf. Section 3.2
- (3) In order to search the GIDAS database for accident cases with CP present, translate them into the query language of the database (SQL), cf. Section 3.3

3.1. Evolution of the Catalog of Criticality Phenomena. Initially, the CP-catalog was created to organize the results of the process step "identification and formalization of criticality phenomenon" of a criticality analysis, as briefly introduced in Section 2.1.

For the initial version of the catalog, the authors searched several sources of information for phenomena hypothesized to be associated with the increased criticality in urban traffic, namely,

- (i) A comprehensive analysis of the official German driving license catalog of questions by Verkehrsblatt Fragenkatalog (<https://www.verkehrsblatt.de/docs/fragenkatalog>)
- (ii) Consideration of various results from the PEGASUS project, such as the ontological framework by Bagschik et al. [38] and its implementation
- (iii) The work conducted on automation risks by Kramer et al. [39] and the corresponding technical report [40]
- (iv) A nonpublic list of risk factors identified within the PEGASUS project (<https://www.pegasusprojekt.de/en>)
- (v) The accident causes from DESTATIS ([17], p. 50), and finally
- (vi) Consolidation via expert knowledge provided by the authors and the VVM consortium.

The process of translating CP, as described in Section 3.3, and intense discussion among peers led to an iterative refinement of the catalog. The latest version of the catalog before the analysis contained the following:

- (i) 166 CP formulated in natural language on varying levels of abstraction, uniquely identified by an ID $\in \{1, \dots, 166\}$
- (ii) A German translation and a textual description of each phenomenon as well as an ontological classification based on the urban 6-layer model [42]
- (iii) Existing relations between phenomena such as abstractions, concretizations, synergies with other phenomena, and synergies with specific contexts
- (iv) A collection of tags associated with the phenomena
- (v) A preliminary estimate of the criticality that is associated with each phenomenon on a 3-point scale (high, medium, and low)

3.2. Selection of the Master Dataset. The first step towards analyzing the GIDAS database regarding the CP-catalog is the selection of a subset of interest from all available GIDAS cases, in the following referred to as the master dataset. It was selected in such a way that it is representative of the OD under consideration within the VVM project, i.e., urban scenarios with passenger car involvement. The version of the GIDAS database which has been utilized (effective from June 30, 2020) contains 41 381 real-world accidents of which 38 571 are completely documented and reconstructed. To account for the aforementioned OD, several filter criteria have been applied to the GIDAS database, restricting the master dataset to cases that

- (i) Happened between the years 2007 and 2020 (many interesting variables were not available in GIDAS before 2007.)
- (ii) Were located in urban areas
- (iii) Involved at least one passenger car
- (iv) For which a reconstruction of the accident is available

Applying these filter criteria to all GIDAS cases yields a total of $n = 15\,417$ relevant accident cases. If we denote by w_i the weight factor for accident case i , as explained in Section 2.2.2, then the weighted number of accident cases used for the analysis is given by $[\sum_{i=1}^n w_i] = 12\,997$. Note that, when projected to the total German traffic accident statistics in 2019, the GIDAS master dataset represents $[\sum_{i=1}^n e_i] = 160\,385$ accident cases with e_i denoting the extrapolation factor for each case.

3.3. Translating Criticality Phenomena to the GIDAS

3.3.1. Database Scheme. Each CP represents an abstract class of danger that could potentially interfere with an AV's driving task. For the identification of instances of such abstract classes of danger within GIDAS, it is necessary to translate this description into characteristic codes out of the GIDAS codebook, cf. Section 2.2.1. Therefore, it is required to analyze every single CP in terms of its semantics as well as its potential identifiability within the GIDAS. While a general process for such a translation has been examined

extensively by Westhofen et al. [43], its partially instantiated steps are sketched in Figure 3.

Intuitively, we start by expressing a phenomenon using natural language, such as "occluded pedestrian." This expression is based on the implicit world model of the creator, for example, by using a set of assumptions on what the definition of a pedestrian and an occlusion encompasses.

It is conceptually advantageous to then convert this implicit world model to an explicit and formal ontology and use this ontology to represent the resulting phenomena. For the scope of this work, that is, limited and scientific, a unified ontology was directly negotiated over all involved parties during the translation process. However, a formal representation of this unified ontology was not made explicit. Note that, not explicating this step does not invalidate the scientific results, but when applied in an industrial large-scale setting, a rigorous formal ontology will be mandatory to both mitigate complexity issues and for aligning multistakeholder views.

The natural language definitions of the CP are then directly mapped onto the database scheme and its query language, if possible. In some cases, CP is not translatable to the GIDAS codebook; for example, if the variable under consideration is not recorded in the postaccident analysis, then this concerns the activation of indicator signals of the involved TPs. Often, such variables can be estimated or derived from the official accident report. Due to those approximations of the intended semantics of the CP, it becomes necessary to register such modeling assumptions in a formal ontology mapping.

At this point, it should be mentioned that GIDAS is a traffic accident database and thus represents an exceptional part of the road traffic. Under these conditions, not all CP are identifiable, mainly because the described circumstances are temporary, subjective, cannot be determined in retrospect, or are too general or negligible regarding the course of the accident. There exist some CP which can only be identified through an individual case analysis. However, if no suitable keywords could be determined, the translation was dispensed due to the immense effort required. As to analyze GIDAS in terms of a specific CP, an individual user analysis and detailed expert knowledge on the GIDAS codebook is required.

By assigning the associated GIDAS parameters to the description of the CP, queries can be executed on the database to retrieve binary results on their presence in accident cases. Exemplary database queries realized as SQL code for the three running examples CP "intersecting planned trajectories of TPs" (#17), "nonego-TP violating right of way" (#31) and "occlusion" (#131) can be found in Appendix A.

We now illustrate this process by means of the CP "nonego-TP violating right of way" (#31), as shown in Table 1. Its natural language description directly contains the following two pieces of information: first, the phenomenon concerns a nonego-TP, so there are no restrictions for the ego; second, a nonego-TP violates the right of way. From the point of view of the traffic accident database, this means that this traffic participant bears, at least partially, the guilt of causing the accident. Whether the ego is to blame here as

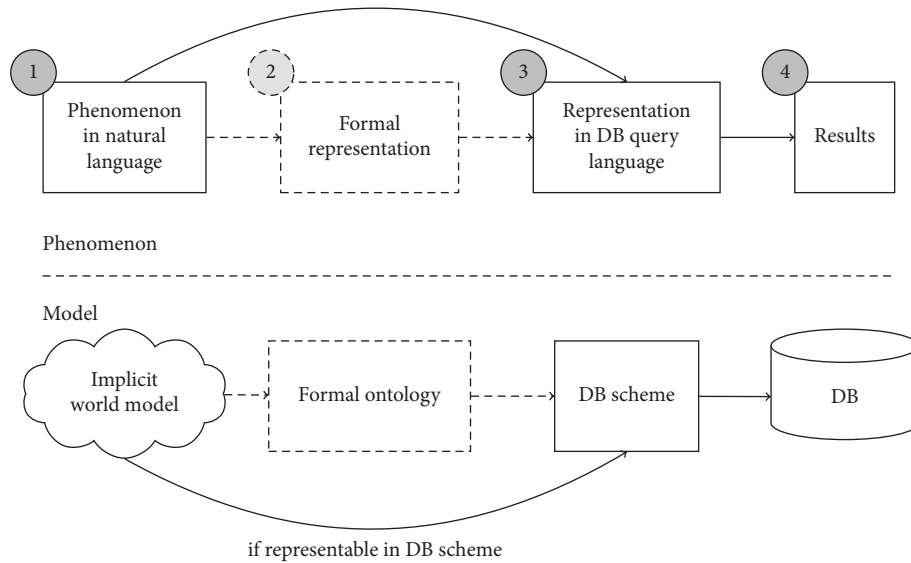


FIGURE 3: The general workflow of identifying a machine-comprehensible representation of a criticality phenomenon using database (DB) queries. The work at hand skips step two and translates criticality phenomena directly to the query language of the GIDAS database.

well is not relevant for the identification of this CP. From the several parameters in GIDAS relating to the cause(s) of the accident, for this CP, we used the parameter “URSWIS” which encodes for actor-specific accident causes, as mentioned in Section 2.2.1. For each TP involved in the accident, up to four causal factors are recorded. So, for the phenomenon at hand (#31), the URSWIS-code of the nonego-TP is relevant. Examples of causal factors encoded by URSWIS include the nonadherence to several right-of-way regulations such as “left yields to right,” traffic signs and lights, manual traffic control via police officers, or the misbehavior of pedestrians, for example, not paying attention to the traffic or suddenly emerging from occlusions. Accordingly, this CP also includes priority violations, specifically the realization of “nonego-TP violating right of way” (#31) as SQL query can be found in Section A.2.

4. Results of Analysis

This section describes the analysis of the GIDAS database and its results, based on the CP-catalog, the selected master dataset, and the translation of the CP to the database scheme. In particular, we delineate how the quantities required for the risk decomposition of equation (1) are estimated for $P(\text{CP}|\text{Accident}, \text{OD})$ and $P(\text{Severity}|\text{Accident}, \text{CP}, \text{OD})$, and we use our analysis of the GIDAS database and for $P(\text{Accident}|\text{OD})$ we rely on the German national traffic statistics. Furthermore, we generalize this risk estimation from a single CP to conjunctions of CP. The resulting data are included as supplementary material [41].

4.1. Results of Translating Criticality Phenomena. In the initial version of the CP-catalog, as described in Section 3.1, all CP were formulated in the natural language. For their identification within the GIDAS database, the authors analyzed each of the 166 CP regarding their semantics and

translated them, if possible, into SQL queries using suitable variables and parameter ranges from the GIDAS codebook as described in Section 3.3.

In total, we were able to translate 116 out of the 166 CP ($\approx 70\%$) in the CP-catalog into SQL queries, which, in turn, implies that 50 out of the 166 CP ($\approx 30\%$) could not be identified within the GIDAS. This means that, for the remaining CP, either (i) other data sources need to be consulted, (ii) manual analysis of the GIDAS case report needs to be conducted, or (iii) that some of the CP needs to be discarded. From the point of view of the criticality analysis, it is clear that other sources besides the GIDAS database are needed and that some of the influencing factors hypothesized to be associated with criticality will have to be discarded. Let us remark that it (ii) was not an option for the authors, as this would require immense personal efforts.

As to give some examples, we distinguish the three classes of nonidentifiable CP as follows:

- (1) *Missing Data.* The first class consists of CP that could not be analyzed due to a lack of appropriate information in GIDAS. Examples include “missing indicator signal,” “small distance to lane marking,” “object near ego” (think of a can, ball, or branch), and “right-of-way dead-lock.” this is by far the largest class of nonidentifiable CP.
- (2) *Infeasible Translation.* For the second class, a manual analysis of the individual accident cases would have been necessary, as the authors could not find an adequate translation into the query language. Let us mention the examples “construction site remains” and “overriding right of way” for which none of the available parameters in GIDAS enable reliable automated identification of accident cases. Although manual analysis of single cases by an expert could, in principle, identify such CP, but this was omitted due to effort constraints.

TABLE 1: A small excerpt from the CP-catalog containing the ID, name, ontological classification, and description of the three criticality phenomena from the running example.

ID	Criticality phenomenon	Ontological classification	Description
#17	Intersecting planned trajectories of TPs	Related \longrightarrow spatial	The intended trajectories of at least two TPs intersect nonparallelly. In addition, the cases of parking in/out and backwards driving are excluded here
#31	Nonego-TP violating right of way	Related \longrightarrow policy \longrightarrow rules	Any violation of a right of way rule (can be indicated through lights, signs, or not at all) as defined by the traffic rules
#131	Occlusion	Related \longrightarrow perception \longrightarrow characteristics	For a perceiving TP, occlusion is the visual obstruction of some entity through some opaque object such that the obstructed entity cannot or only partially be perceived by the TP

Of course, most of the criticality phenomena (rows) and most of the labels (columns) are omitted here, but the entire CP-catalog is available as supplementary material [41].

- (3) *Irrelevant for Human Traffic.* The third (and smallest) class is comprised of CP which are not yet relevant to traffic accidents involving human-operated passenger cars but could become relevant for ADSs, for example “dead spot” or “balloon near ego.”

Note, however, that these classes are not disjoint as the classes (2) and (3) can be considered subclasses of (1).

4.2. The Case-Phenomenon Relation Matrix. The application of the SQL queries to the GIDAS database generates, for each of the 116 identifiable CP, a list of accident cases where the respective phenomenon was present. For evaluation, we created a case-phenomenon relation matrix of size $15\,417 \times 116$ where each row represents a GIDAS accident case from the master dataset. Each accident case i is identified via an (anonymized) case number and comes with a weighting factor w_i , an extrapolation factor e_i which projects to the overall German accident statistics, cf. Section 2.2.2 and an associated severity s_i on a three-point ordinal scale. The columns correspond to the 116 CP for which SQL queries could be realized. Therefore, an entry $a_{i,j}$ of the $15\,417 \times 116$ case-phenomenon relation matrix is defined as

$$a_{i,j} = \begin{cases} 1, & \text{if CP}_j \text{ was present in accident case } i, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where $i = 1, \dots, 15\,417$, $j = 1, \dots, 116$, and CP_j denote the CP identified with ID $\#j$. A small excerpt from the case-phenomenon relation matrix can be seen in the lower right part of Table 2.

From this matrix, together with the factors w_i and e_i , the absolute, relative, and projected frequencies can easily be computed for all of the 116 identifiable CP as

$$\begin{aligned} \text{abs_freq}(\text{CP}_j) &= \sum_{i=1}^n w_i a_{i,j}, \\ \text{rel_freq}(\text{CP}_j) &= \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i a_{i,j} = \frac{\text{abs_freq}}{\sum_{i=1}^n w_i}, \end{aligned} \quad (4)$$

$$\text{proj_freq}(\text{CP}_j) = \sum_{i=1}^n e_i a_{i,j},$$

where $n = 15\,417$, cf. Section 2.2.2. An evaluation of the quantities of equation (4) for the CP from the running example is given by Table 3. Besides these simple frequentist quantities, the case-phenomenon relation matrix enables, among other statistical analyses, estimating the risk associated with CP according to equation (1), as will be demonstrated in Section 4.3.

4.2.1. Nonexistent and Rare CP in the Master Dataset.

Note that, five CP from the CP-catalog did not occur at all in the master dataset, i.e., $\text{abs_freq}(\text{CP}) = 0$. As, in this work, the source of the data consists of urban traffic accidents with passenger car involvement and damage to persons, and CP being rare or nonexistent in this dataset shall not yet lead to their discardment neither in a criticality analysis nor in a safety case. However, we may need to consult other datasets before we can reject the hypothesis that the phenomenon is associated with criticality.

Let us consider some examples. While “violation of zip merging” may be a CP in general, it is not an associated risk with personal injury accidents in urban areas. This may be due to the usually low-speed level during inner-city merging that prevents such accidents. Another example is “manual traffic control” which rarely occurs nowadays in Germany. Thus, the probability of identifying such an accident in GIDAS is very small. However, this phenomenon might become more relevant for ADSs due to the communication barrier. Similarly, the weather-related phenomenon “hail” only occurs with an absolute frequency of two in the master dataset. Hail is a rare phenomenon in general and often of short duration but of high intensity. Therefore, human drivers tend to quickly adapt their driving behavior to this critical environmental condition, thereby mitigating the risk of a personal injury accident.

4.3. Bayesian Approach for the Assessment of Risk Associated with Criticality Phenomena.

Picking up the Bayesian approach for assessing the accident risk associated with CP from Section 1 and, in particular, equation (5), we now elaborate on how the involved quantities can be estimated empirically for the 116 CP that were identifiable within the GIDAS. We recall that we factorized the risk of an accident with passenger car involvement and damage to persons associated with a CP within an OD as

$$\begin{aligned} &\text{Risk}(\text{CP}, \text{Accident}, \text{Severity}|\text{OD}) \\ &= \underbrace{P(\text{CP}|\text{OD})}_{\text{Exposure}} \underbrace{P(\text{Accident}|\text{CP}, \text{OD})}_{\text{Controllability}} \underbrace{P(\text{Severity}|\text{Accident}, \text{CP}, \text{OD})}_{\text{Severity}} \quad (5) \\ &= P(\text{CP}|\text{Accident}, |\text{OD})P(\text{Accident}|\text{OD})P(\text{Severity}|\text{Accident}, \text{CP}, \text{OD}), \end{aligned}$$

TABLE 2: An excerpt of the $15\,417 \times 116$ case-phenomenon relation matrix together with an anonymized GIDAS case number, weighting factor w_i , extrapolation factor e_i , and accident severity s_i for each case.

GIDAS case number (anon.)	Weighting factor	Extrapolation factor	Severity	Criticality phenomenon ID											
				#17	#19	#20	#21	#22	#23	#24	#25	#26			
47415	1.089	13.433	1	0	0	0	0	0	0	0	0	0	0	0	0
24580	0.362	4.472	2	1	0	0	0	0	0	0	0	0	0	0	0
32474	0.743	9.171	1	1	0	1	0	0	0	0	0	0	0	0	0
45433	0.468	5.777	2	1	1	0	0	0	0	0	0	0	0	0	0
40614	1.089	13.433	1	0	0	0	0	0	0	0	0	0	0	1	1
93567	0.743	9.171	1	1	0	1	0	0	0	0	0	0	0	0	1

A value of 1 means that the corresponding criticality phenomenon was present in the respective GIDAS accident case, while a value of 0 indicates its absence.

TABLE 3: Frequentist quantities for the CP from the running example.

ID	Criticality phenomenon	Absolute frequency	Relative frequency (%)	Projected frequency
#17	Intersecting planned trajectories of TPs	7 156	55.1	88 305
#31	Nonego-TP violating right of way	2 644	20.3	32 628
#131	Occlusion	2 978	22.9	36 746

Absolute and relative frequencies describe their occurrence in the master dataset, cf. Section 3.2, while the projection provides an estimate for the total German traffic accident statistics in 2019. While absolute and projected frequency is rounded to integers, relative frequency is rounded to three digits.

where in the following calculations, the OD is assumed to be urban areas in Germany, as discussed in Section 3.2, in accordance with the VVM project's criticality analysis.

4.3.1. Estimation of Risk-Related Quantities

(1) *Estimation of the First Factor.* For a phenomenon CP_j , the first quantity can be estimated directly from the case-phenomenon relation matrix $(a_{i,j})$ of Section 4.2 and the weights w_i from Section 2.2 as

$$P(CP_j | \text{Accident}, \text{OD}) \approx \text{rel_freq}(CP_j) = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i a_{i,j}, \quad (6)$$

with $n = 15\,417$. Note that the validity of this estimate rests on the representativeness of GIDAS for all German traffic accidents and, in particular, on the characteristics of the master dataset, as introduced in Section 3.2.

(2) *Estimation of the Second Factor.* The second quantity, $P(\text{Accident} | \text{OD})$, is a factor that is independent of CP and can hence be estimated from the national German traffic statistics for the OD. In particular, were only on

- (i) DESTATIS to retrieve $r_1^{(2019)} = 127\,856$ as the number of accidents with damage to persons and passenger car involvement in urban areas in 2019 [17]
- (ii) The German Federal Institute BAST for the total amount of kilometers driven by passenger cars in urban areas of Germany in the year 2014 [44], which amounts to $r_2^{(2014)} = 154\,580\,000\,000$ km

Under the assumption that the yearly amount of total kilometers has not changed significantly between 2014 and 2019, we estimate the probability of an accident with passenger car involvement and damage to persons in urban areas in Germany in 2019 as

$$\begin{aligned} P(\text{Accident} | \text{OD}) &\approx \frac{r_1^{(2019)}}{r_2^{(2014)}} \approx \frac{r_1^{(2019)}}{r_2^{(2019)}} = \frac{127\,856 \text{ accidents}}{154\,580\,000\,000 \text{ km}} \\ &= 8.27 \cdot 10^{-7} \frac{\text{accidents}}{\text{km}} = 827 \frac{\text{accidents}}{10^9 \text{ km}} =: \mathcal{C}_{\text{Accident,OD}}. \end{aligned} \quad (7)$$

(3) *Estimation of the Third Factor.* Within GIDAS, the severity of a personal injury p is encoded on a three-point ordinal scale according to the official definition as

- (i) $p = 3$, if the person succumbs to the inflicted injury within 30 days (“fatal injury”)
- (ii) $p = 2$, if the person is admitted, at least for 24 hours, to a hospital for stationary treatment (“serious injury”)

- (iii) $p = 1$, for all other injuries (“slight injury”)

The severity $s_i \in \{1, 2, 3\}$ of a given accident case i is then defined as the maximum over the individual personal injuries that occurred in the accident. Based on this, for each CP_j that was identifiable and occurred at least once in the dataset, we estimate the probability of reaching a given minimal severity level s as

$$P(\text{Severity} \geq s | \text{Accident}, CP_j, \text{OD}) \approx \frac{\sum_{i=1}^n \mathbf{1}_{s_i \geq s} w_i a_{i,j}}{\text{abs_freq}(CP_j)} \frac{\text{acc. of severity} \geq s}{\text{accidents}}, \quad (8)$$

where the indicator function $\mathbf{1}_{s_i \geq s}$ is defined by

$$\mathbf{1}_{s_i \geq s} = \begin{cases} 1, & \text{if } s_i \geq s, \\ 0, & \text{if } s_i < s. \end{cases} \quad (9)$$

Due to the fact that GIDAS does not contain accidents without personal injury, we have that $P(\text{Severity} \geq 1 | \text{Accident}, CP, \text{OD}) = 1$ for all CP, so the interesting cases for estimating that probability are $s \in \{2, 3\}$.

(4) *Estimation of CP-Associated Accident Risk.* Plugging the estimates of equations (6)–(8) into the Bayesian risk

decomposition of equation (5), we approximate the risk associated with CP_j as

$$\begin{aligned} & \text{Risk}(CP_j, \text{Accident}, \text{Severity} \geq s | \text{OD}) \\ & \approx \mathcal{E}_{\text{Accident,OD}} \cdot \text{rel_freq}(CP_j) \cdot \frac{\sum_{i=1}^n 1_{s_i \geq s} w_i a_{i,j} \text{acc. of severity} \geq s}{\text{abs_freq}(CP_j) \text{accidents}} \\ & = 827 \cdot \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n 1_{s_i \geq s} w_i a_{i,j} \frac{\text{acc. of severity} \geq s}{10^9 \text{ km}}, \end{aligned} \quad (10)$$

where $n = 15417$. These estimates can now be used in a safety case, for example, as reference values for relative acceptance criteria, or within a criticality analysis for estimating the criticality association of phenomena, cf. Section 4.3.2.

(5) *Evaluation for the Running Example.* Using equation (10), we can now estimate the risk for the CP of the running example by utilizing the constant $\mathcal{E}_{\text{Accident,OD}}$, the case-phenomenon relation matrix of Section 4.2, and the three-point ordinal accident severity provided by GIDAS. Table 4 shows the calculated probability of severity and the associated risk of an accident with at least serious injuries, respectively, fatal injuries for CP_{17} , CP_{31} , and CP_{131} . The calculation of these values for the entire CP-catalog is provided as supplementary material, including the case of $s = 1$, which is left out here for brevity [41].

4.3.2. *Relevance Estimation according to Associated Risk.* Having obtained an estimate for the accident risk associated with CP, we are now able to instantiate the process step “estimation of criticality association,” cf. Figure 1, by ranking the CP according to the quantity $\text{Risk}(CP, \text{Accident}, \text{Severity} \geq s | \text{OD})$ of equation (5). More precisely, in the criticality analysis we are interested in all critical situations, which translates to $s = 1$. Therefore, with equation (10), we use

$$\text{Risk}(CP_j, \text{Accident}, \text{Severity} \geq 1 | \text{OD}) \approx \mathcal{E}_{\text{Accident,OD}} \cdot \text{rel_freq}(CP_j), \quad (11)$$

for estimating the criticality association of CP. Figure 4 shows, as a bar chart, the top twenty CP ranked according to equation (11). Note that, the factor $\mathcal{E}_{\text{Accident,OD}}$ does not depend on CP_j , so this is essentially a ranking according to $P(CP_j | \text{Accident}, \text{OD})$, the first factor of equation (10), which is approximated using the occurrence of the CP in the master dataset, i.e., $\text{rel_freq}(CP_j)$.

Among the top twenty CP, several relate to the dynamics of the involved TPs, such as “strong braking maneuver of TP,” “intersecting planned trajectories of TPs,” “speed,” “reduced friction on road,” “high relative speed,” and “overtaking.” These dynamical factors concerning bad planning, unadapted speeds, and harsh braking maneuvers

are significantly associated with accident risk, likely due to their temporal proximity to collisions. Let us mention that the high-risk phenomenon “strong braking maneuver of TP” is translated to the database scheme by referring to a deceleration rate of at least 4.7ms^{-2} , which approximates the limit force that an average driver is still able to handle safely [45]. Other high-risk CP do refer to the infrastructure, for example, “intersection,” “presence of VRUs/URUs with road access,” “degraded road quality,” “bad road surface,” or to the environment, for example, “road weather.” Moreover, the top twenty also include the CP “occlusion” and “dark clothing of VRU” which work on the level of perception due to blockage of electromagnetic waves by an opaque object and lack of reflection thereof, respectively. We note that these CP can also be relevant for ADS that rely on machine-perception based on cameras, lidar, and radar sensors, although for these technologies “occlusion” or “dark clothing of VRU” might arise on a different range of the electromagnetic spectrum. Finally, “nonego-TP violating right of way” refers to the misbehavior of traffic participants, so the relevance for ADSs is given as well.

Even though the data at hand are highly biased towards situations of maximal criticality, i.e., accidents with damage to persons, the ranking of CP according to $\text{Risk}(CP, \text{Accident}, \text{Severity} \geq 1 | \text{OD})$ provides us with evidence for the process step “estimation of criticality association” in the method branch of the criticality analysis, cf. Figure 1. For example, the accident risk associated with the CP in Figure 4 makes them potential candidates for subsequent causal analysis and provides well-founded information about typical scenario classes in urban accidents with passenger car involvement.

4.4. *Risk Assessment for Combinations of CP.* In Section 1 and equation (5), in particular, we have only considered the accident risk associated with a single CP. However, as can be seen from Figure 5, which displays the distribution of CP numbers in the master dataset, accident cases usually feature several CP. In fact, in approximately 55% of all cases in the master dataset between four and seven CP were present.

Let us remark that the distribution of Figure 5 can easily be shifted by adding trivial abstractions or concretizations of highly prevalent CP. In order to mitigate the effect of double-counting CP due to existing abstraction/concretization

TABLE 4: The estimated probability of severity and the associated risk for the three criticality phenomena from the running example.

ID	Criticality phenomenon	Severity (equation (8))		Risk (equation (10))	
		With $s = 2$	With $s = 3$	With $s = 2$	With $s = 3$
#17	Intersecting planned trajectories of TPs	15.43%	0.30%	70.3	1.4
#31	Nonego-TP violating right of way	17.94%	0.52%	30.2	0.9
#131	Occlusion	17.20%	0.21%	32.6	0.4

The values are calculated using equations (8) and (10), respectively, for $s = 2$, i.e., serious injuries, and $s = 3$, i.e., fatal injuries. The units of severity and risk, cf. (8) and (10), respectively, are omitted for brevity. Here, the probability of severity is rounded to four digits and risk is rounded to one digit.

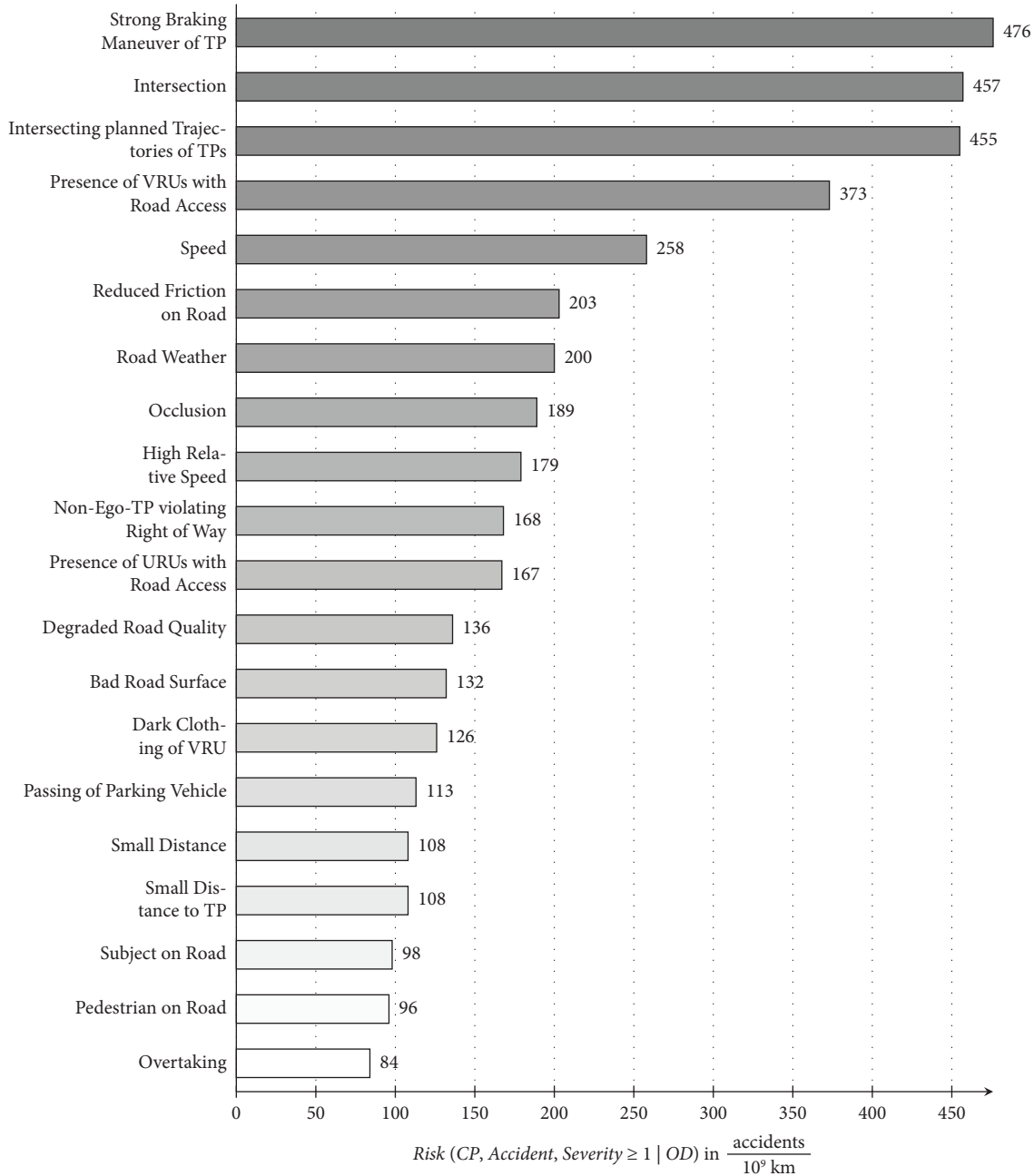


FIGURE 4: Top twenty criticality phenomena ordered according to the quantity Risk(CP, Accident, Severity ≥ 1|OD), estimated using equation (11), with unit accidents with passenger car involvement and damage to persons per billion kilometers.

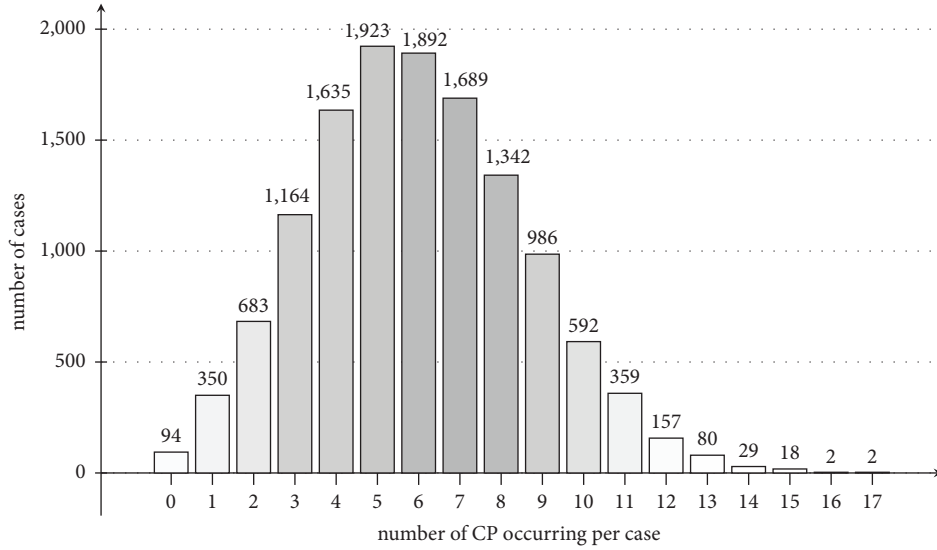


FIGURE 5: Plot of the distribution of accident cases in the master dataset, cf. Section 3.2, along the number of criticality phenomena occurring per accident case. This statistic has been adjusted to exclude common abstraction-concretization relations in order to avoid double-counting of CP.

relations, we have excluded the following abstractions for the distribution of Figure 5: “small distance,” “small distance to TP,” “speed,” “occlusion,” “glare,” “subject on road,” “pedestrian on road,” “overtaking,” and “road weather.”

In any case, the CP-number distribution cannot be used to draw definite conclusions, but should be seen as a motivation to investigate the risk associated with combinations of CP more rigorously. Furthermore, combinatorial aspects of CP are important to various verification and validation activities, for example, when deriving realistic test scenarios for ADS behavior [38, 46]. So, in order to prevent the exponential blow-up of considering all combinations of CP in a safety case, we are interested in (a) the risk associated with combinations of CP and (b) how to find potentially relevant combinations of CP.

First, we generalize the risk assessment from Section 4.3 to address (a) and then, in Section 4.4.2, we evaluate

associations among CP to generate hypotheses about interesting CP-combinations, addressing (b).

4.4.1. Generalized Risk Estimation for CP-Combinations.

In the following, with a slight abuse of notation, let $CP_1, \dots, CP_l, CP_{l+1}, \dots, CP_k$ $1 \leq l < k$ denote different CP and set $CP^+ := CP_1 \wedge \dots \wedge CP_l$ and, likewise, $CP^- := (CP_{l+1} \vee \dots \vee CP_k)$. Based on the case-phenomenon relation matrix from Section 4.2 and the weights from Section 2.2.2, we can directly calculate the absolute frequency of accident cases in the master dataset for which the CP-conjunction $CP^+ \wedge CP^-$ occurred. Hence, we generalize the absolute and relative frequencies from equation (4) to multiple CP as

$$\begin{aligned} \text{abs_freq}(CP^+ \wedge CP^-) &= \sum_{i=1}^n w_i \prod_{j=1}^l a_{i,j} \prod_{j=l+1}^k (1 - a_{i,j}), \\ \text{rel_freq}(CP^+ \wedge CP^-) &= \frac{\text{abs_freq}(CP^+ \wedge CP^-)}{\sum_{i=1}^n w_i}, \\ \text{proj_freq}(CP^+ \wedge CP^-) &= \sum_{i=1}^n e_i \prod_{j=1}^l a_{i,j} \prod_{j=l+1}^k (1 - a_{i,j}). \end{aligned} \quad (12)$$

Figure 6 shows the more general absolute frequencies for different combinations of the three CP “intersecting planned trajectories of TPs,” “nonego TP violating right of

way,” and “occlusion” as a proportional Venn diagram. As the overlap is quite substantial, their conjunctions, such as $CP_{17} \wedge CP_{31}$, might be an interesting CP-combination for risk assessment.

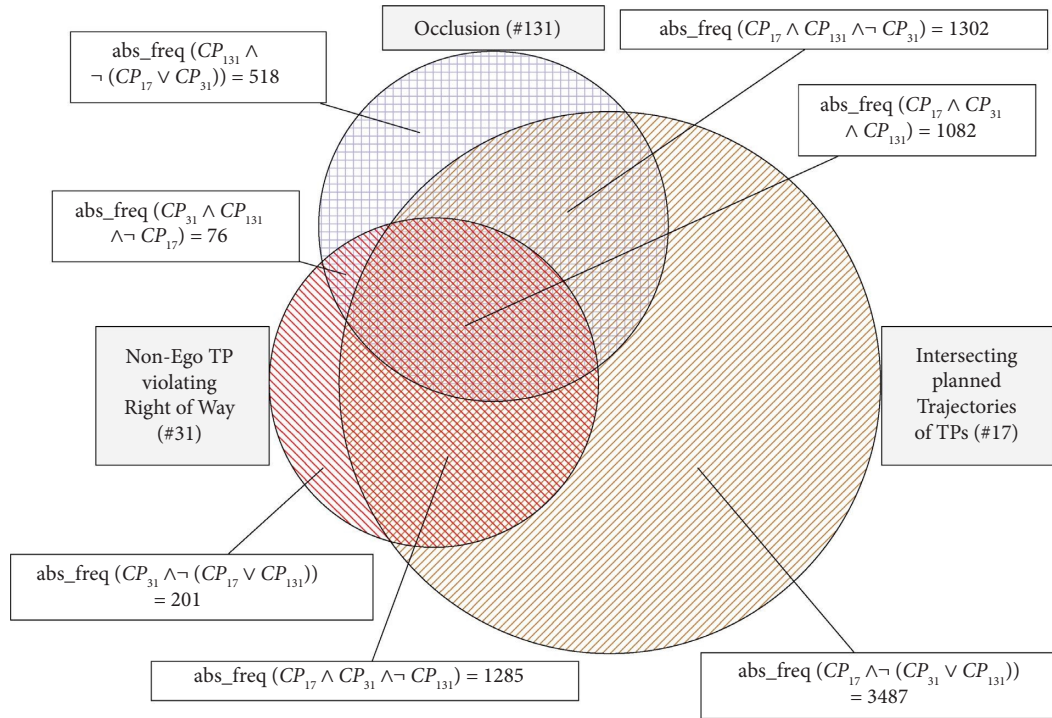


FIGURE 6: Proportional Venn diagram showing the absolute frequencies in the master dataset for different combinations of the three running example criticality phenomena.

With the probability of severity being generalized to CP-combinations as

$$P(\text{Severity} \geq s | \text{Accident}, CP^+ \wedge CP^-, OD) \approx \frac{\sum_{i=1}^n 1_{s_i \geq s} w_i \prod_{j=1}^l a_{i,j} \prod_{j=l+1}^k (1 - a_{i,j})}{\text{abs_freq}(CP^+ \wedge CP^-)} \frac{\text{acc. of severity} \geq s}{\text{accidents}}. \quad (13)$$

$$\text{Risk}(CP^+ \wedge CP^-, \text{Accident}, \text{Severity} \geq s | OD)$$

$$= P(CP^+ \wedge CP^- | \text{Accident}, OD) P(\text{Accident} | OD)$$

$$\approx \mathcal{C}_{\text{Accident}, OD} \cdot \text{rel_freq}(CP^+ \wedge CP^-) \frac{\sum_{i=1}^n 1_{s_i \geq s} w_i \prod_{j=1}^l a_{i,j} \prod_{j=l+1}^k (1 - a_{i,j})}{\text{abs_freq}(CP^+ \wedge CP^-)} \frac{\text{acc. of severity} \geq s}{\text{accidents}} \quad (14)$$

$$= 827 \cdot \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n 1_{s_i \geq s} w_i \prod_{j=1}^l a_{i,j} \prod_{j=l+1}^k (1 - a_{i,j}) \frac{\text{acc. of severity} \geq s}{10^9 \text{ km}}$$

Figure 5, which is also contained in the supplementary material [41], shows the estimated probability of severity and

Here, we obtain almost analogously to equation (10), the following estimate for the accident risk associated with such conjunctions of CP:

risk for four different combinations of the CP from the running example. A comparison with the respective values for single CP, cf. Table 4, indicates that while conjunctions of

TABLE 5: The estimated probability of severity and the associated risk for combinations of the three CP from the running example.

Criticality phenomena combination	Severity (equation (13))		Risk (equation (14))	
	With $s = 2$	With $s = 3$	With $s = 2$	With $s = 3$
$CP_{17} \wedge CP_{31}$	18.74%	0.51%	28.2	0.8
$CP_{31} \wedge CP_{131}$	20.89%	0.48%	15.4	0.4
$CP_{17} \wedge CP_{131}$	18.17%	0.23%	27.6	0.3
$CP_{17} \wedge CP_{31} \wedge CP_{131}$	21.42%	0.44%	14.7	0.3

The values are calculated using equations (13) and (14), respectively, for $s = 2$, i.e., serious injuries, and $s = 3$, i.e., fatal injuries. The units of severity and risk, cf. (13) and (14), respectively, are omitted for brevity. Here, the probability of severity is rounded to four digits and risk is rounded to one digit.

CP are associated with more serious injuries, they occur disproportionately less frequently such that the associated accident risk is significantly reduced compared to the risk associated with individual CP.

4.4.2. Associations among Criticality Phenomena. In order to avoid the combinatorial blow-up of assessing the risk of all k -wise combinations of CP, we are interested in associations among them. Taking a certain level of association as a prerequisite for interesting CP-combinations, can lead to a significant reduction of effort for the safety engineer. Besides, by Reichenbach’s common cause principle, stochastically dependent random variables are likely to have a common cause, so knowledge about associations among CP will be helpful when modeling and analyzing their causal relations [13].

We recall that in this work CP corresponds to binary-valued random variables restricted to the accident with passenger car involvement and damage to persons in urban areas in Germany. As a measure of association, we relied on the Φ -coefficient [47] to calculate the pairwise correlation between the two CP. By referencing the entries of Table 6, the Φ -coefficient is given by the formula

$$\Phi(CP_1, CP_2) = \frac{h_{11}h_{22} - h_{12}h_{21}}{\sqrt{(h_{11} + h_{12})(h_{21} + h_{22})(h_{11} + h_{21})(h_{12} + h_{22})}} \quad (15)$$

$$\in [-1, 1].$$

The values of the Φ -coefficient range from -1 to $+1$, where $\Phi = 0$ means that there is no correlation in the data, at $\Phi = -1$ the CP are mutually exclusive, and at $\Phi = +1$ there is a perfect correlation, i.e., the CP are always present in combination. Note that, for $\Phi > -1$, however, the CP might still be mutually exclusive when $h_{22} = 0$, but not when $h_{11} > 0$.

For each of the $\binom{116}{2} = 6\,670$ pairwise combinations of different CP, the entries of Table 6, therefore Φ , can be computed directly from the case-phenomenon relation matrix $(a_{i,j})$ from Section 4.2. Note that, CP with zero absolute frequency were excluded from this subanalysis. A table containing the Φ -coefficient for the remaining 6 105

TABLE 6: Generic contingency table for the definition of the Φ -coefficient.

CP	CP ₁		
	Presence?	0	1
CP ₂	0	h_{11}	h_{12}
	1	h_{21}	h_{22}

pairwise CP-combinations, ranked according to its absolute value, is contained in the supplementary material [41].

Table 7 contains the Φ -coefficient for all pairwise combinations of CP_{17} , CP_{31} , and CP_{131} as well as for the respective entries of h_{11} , h_{12} , h_{21} , h_{22} for each combination. Between them, we observe significant correlations of 0.35, 0.27, and 0.25, respectively. While there is a significant positive correlation among these CP, they clearly do not exhibit any abstraction/concretization relation. Besides indicating that the respective conjunctions are appropriate candidates for risk assessment, as previously conducted in Section 4.4, a downstream causal analysis of either phenomenon is likely to reveal some common confounding variable being present, for example, the CP “intersection.”

Generally, values of Φ close to $+1$ indicate an abstraction/concretization relationship between the two CP’s. Numerically, this is due to either h_{12} or h_{21} from Table 6 being close to zero. Also, such correlations can be distorted through bias introduced from the dataset or from the database scheme. We consider the example of “pedestrian on road” and “subject on road,” while “subjects” include wild animals and pet animals, in the master dataset, subjects are overwhelmingly pedestrians. As we have restricted the master dataset to urban accidents, there were only 23 cases where a “subject on road” is referred to as a nonhuman, resulting in φ (“subject on road” and “pedestrian on road”) = 0.99.

As an example for bias introduced by the database scheme, we consider the CP-pair “small distance” and “small distance to TP” where the latter is a concretization of the former, already indicating Φ being close to $+1$. In this case, as distances to nonTP-objects are not identifiable within the GIDAS, biased was introduced through the impossibility of an exact translation of the natural language specification to SQL queries and therefore, through the database scheme, i.e., the phenomenon “small distance” was underapproximated in the translation as the scheme did not encompass the relevant fields for an exact identification. Hence, their

TABLE 7: The Φ -coefficient for the three pairwise combinations of CP from the running example together with the entries from the respective contingency tables.

CP ₁	CP ₂	Φ	0-0	0-1	1-0	1-1
Intersecting planned trajectories of TPs (#17)	Nonego-TP violating right of way (#31)	0.35	5 564	277	4 789	2 367
Intersecting planned trajectories of TPs (#17)	Occlusion (#131)	0.27	5 249	593	4 771	2 384
Nonego-TP violating right of way (#31)	Occlusion (#131)	0.25	8 534	1 819	1 486	1 158

translated SQL queries are identical and we have Φ (“small distance” and “small distance to TP”) = +1.

On the negative side of Φ -values we register CP that tend to exclude each other, which can again be due to reasons of semantics or due to the source of data that Φ was evaluated on. In the case of the master dataset, the lowest $\Phi = -0.41$ was realized between the “small distance” and “intersecting planned trajectories of TPs.” Usually in GIDAS, “small distance” indicates rear-end collisions while “intersecting planned trajectories of TPs” hints at crossing or turning accidents. The reason why Φ is not closer to -1 is the high number of accidents where neither of these two CP occur, namely, $h_{11} = 4 184$. Another example of CP with negative Φ is given by Φ (“curvature of road” and “intersection”) = -0.3 . While these CP exclude each other semantically, Φ is not closer to -1 because of many accidents with none of these CP, namely, $h_{11} = 4 940$ cases on straight roads that were neither curves nor intersections.

4.5. Analysis of Edge Cases. Besides the most frequent CP-number combinations, as provided by Figure 5, the tails of the CP-number distribution are valuable. From these tails, for example, accident cases with 0 CP (94 cases) or ≥ 14 CP (41 cases), it is possible to extract indications on the quality of the CP-catalog such as its completeness or the abstraction levels.

The cases with very few or even without any CP are interesting in manifold ways. On the one hand, an individual inspection of these cases might uncover gaps in the CP-catalog and can be used to find CP candidates from a detailed accident case analysis. This process iteratively facilitates completeness of the CP-catalog. On the other hand, the CP might be available in the CP-catalog, but currently not identifiable within the GIDAS, or at least not automatically. Let us mention that during the analysis the phenomenon “occluded traffic light” from the initial CP-catalog has been added to the GIDAS codebook, and examples of which may only be identified in the GIDAS through natural language analysis of individual case descriptions which include CP’s such as “push-to-front motorcyclist” and “construction site remains.” Performing such analyses of case descriptions, however, is out of the scope of this work. Moreover, accidents with only one participant, for example, caused by a heart attack or some other health issue of the driver, show up in the 0 CP-class, as the CP-catalog has been elicited with respect to ADSs at SAE Levels 4 and 5.

On the other end of the CP-number distribution, we have accident cases of high complexity featuring up to 17 CP. Again, these extreme accident cases are worth further consideration. First, we consider the example provided by Figure 7, sketching an accident case from the master dataset where a left-turning passenger car collided with a child not

observing the red light on a pedestrian crossing at an intersection. This accident features 16 CP, several of which relate to the infrastructure, namely, “intersection,” “pedestrian crossing,” and “intersecting tram rails” as well as to its degradation, as “degraded road quality” and “degraded lane markings” were also present. Regarding environmental conditions, this case exhibits a “limited global light source” due to nighttime as well as “rain,” and therefore “reduced friction on road.” Moreover, we have “intersecting planned trajectories of TPs,” “presence of VRUs with road access,” “pedestrian crossing road directly,” “nonego-TP running a red traffic light” and, therefore also “nonego-TP violating right of way.” Additionally, the clothing of the child had low visibility, and due to the age, the child counts as an unpredictable road user (URU). Finally, the driver of the car tried to avoid the collision by emergency braking. Hence, the “Presence of VRUs with Road Access,” “Dark Clothing of VRU,” and “Strong Braking Maneuver of TP” apply as well.

Note that, although all these factors were present, the actual reasons for the accident may involve only a subset of them. For example, while “intersecting tram rails” or “degraded lane markings” may have been irrelevant to the emergence of the accident, other factors such as “reduced friction on road” may not have caused the accident, but may have worsened its severity. As to distinguish these possibilities, a causal analysis of the CP is necessary.

Accident cases with this large number of CP provide interesting scenarios for the criticality analysis and scenario-based verification and validation in general. From the perspective of the criticality analysis, such examples can be used to guide data collection efforts in the real-world or in a simulation, to validate criticality metrics, and for the evaluation of safety principles. On the side of scenario-based verification and validation, such complex accident scenarios can be used to define requirements on the ADS’s behavior and as high-yielding test cases covering multiple CP as shown by Scanlon et al. [48].

Finally, let us remark that our approach to risk estimation for conjunctions of CP, as presented in Section 4.4, fails for cases featuring such a high number of CP. For example, the accident of Figure 7 is the only case in the master dataset featuring this exact combination of CP. While the relative frequency is close to zero, and the empirical estimation of the accident severity using one sample is completely invalid.

5. Discussion

This section discusses the preparation and the results of the presented analysis, as described in Sections 3 and 4, regarding the safety aspects of the automated driving, their

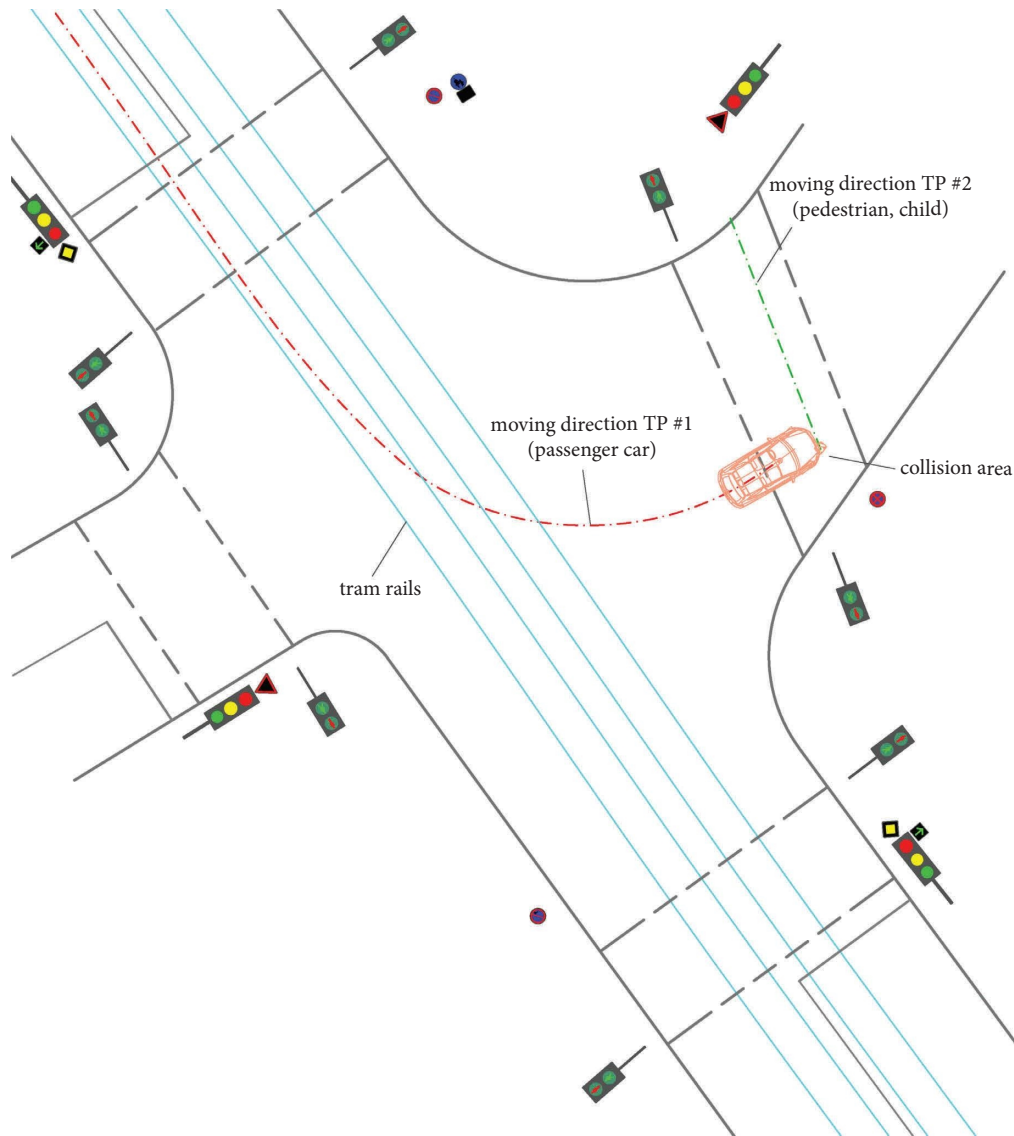


FIGURE 7: An accident scenario featuring 16 CP. Such highly complex accident scenarios can be used as test scenarios for the AV performance.

impact on the criticality analysis, and on the GIDAS database.

5.1. Lessons Learned

5.1.1. The Necessity of Formalizing Criticality Phenomena.

While an initial description in natural language is unavoidable when identifying CP, the work at hand has clearly exposed the need for a formal representation thereof. As described in detail in Section 3.3, we skipped the step of formalizing the CP-catalog and instead translated the CP directly to a representation in the GIDAS database query language. Since the CP had no underlying formal description and were defined in (sometimes vague) natural language, this action required immense coordination between the expert knowledge of the criticality analysts who provided the CP-catalog and the GIDAS database experts.

As elaborated on in Section 5.2.1, safety cases may rely on system-dependent quantities (for example, probabilities conditioned on the ODD). Such quantities have to be estimated using the system-dependent data from simulations, proving grounds, or fleets so as to make them comparable to the quantity of equation (5), and the semantics of the CP within the system-dependent data has to be unambiguously mappable to those used for the OD data.

In this regard, it seems highly impractical to perform a manual translation of the CP once for every data scheme in order to ensure the consistency between the relevant estimates. Therefore, the need for a formalization of CP that is independent of the specific source of data is given and is examined by Westhofen et al. [43].

5.1.2. Extensions of the GIDAS Database Scheme. The work at hand made it clear to the authors that a catalog of CP, as

elicited by a criticality analysis, almost directly functions as a source of requirements for data collection efforts regarding safety aspects within the transportation system, in general, and for ADS, in particular.

Since, at the time of analysis, the GIDAS database did not contain accident cases with ADS-operated vehicles, we conducted the analysis with a strong focus on CP that are also relevant to human traffic, cf. Section 2.1. Although the GIDAS codebook, as described in Section 2.2.1, codes for a vast number of variables, translating the CP-catalog to the database scheme furthered its completeness by uncovering white spots. For example, we identified the impossibility to observe the phenomenon “occluded traffic light” within GIDAS, which has led to the addition of a respective variable in the codebook.

In particular, when the collection of CP is extended to include phenomena specifically related to the machine-perception technology of ADS via sensors, then this extended CP-catalog can be used to drive the extension of the GIDAS database scheme towards preparedness for accident cases involving ADS-operated vehicles.

On an even more general level, a comprehensive CP-catalog is relevant, by definition of the term criticality phenomenon, for all data related to the safety of ADSs. Therefore, a sufficiently complete and unambiguously formalized CP-catalog can become a driver for requirements on data collection for ADS safety. Recently developed normative initiatives, such as the UN regulation UNECE R157 [49], already hint at such data collection requirements. For example, the UNECE R157 requires that “each vehicle equipped with ALKS (the system) shall be fitted with a data storage system,” which is, among others, required to record various variables in case the system is “involved in a detected collision”.

5.2. Utilization & Interpretation of Results. After these remarks on valuable insights regarding the practical execution, we discuss how the analysis results, presented in Section 4, can be leveraged for the safety of ADSs, in general, and for a criticality analysis, in particular.

5.2.1. Leverage of Results for ADS Safety Cases. In this work, we demonstrated how quantities of the form $P(CP_j | \text{Accident}, OD)$ and $P(CP^+ \wedge CP^- | \text{Accident}, OD)$, cf. Sections 4.3 and 4.4, can be estimated from accident data in order to assess the risk associated with CP. Here, conditioning on the OD reflects the assumption that our analysis is performed system-independently, before its design. During the design phase of an ADS, the ODD is specified, which are the “operating conditions under which a given ADS or feature thereof is specifically designed to function [1].” In SOTIF-related scenario-based analyses, this ODD is typically broken down into scenario clusters with special relevance to safety, for example, scenarios with occluded pedestrians at urban intersections. In turn, this allows estimating the risk for single clusters instead of analyzing the entire ODD at once for which it is known that brute-force

approaches will fail due to the size of data required to observe enough extremely rare events [4].

To obtain such a system-dependent risk estimate, a valid strategy is to condition the quantities of equation (5) on the ODD instead of the OD. However, the system’s behavior can drastically influence the exposure to CP within the ODD, therefore system-dependent data are required. Both estimates can then be used for acceptance criteria, for example, comparing the current level of risk in the OD and the system-induced risk in the ODD for estimating a positive risk balance of the analyzed system [14, 50]. However, as stated in Section 5.1.1, this requires a consistent semantics of CP between both system-dependent and system-independent datasets.

5.2.2. Leverage of Results within a Criticality Analysis. The criticality analysis, as conducted within VVM project, profited from the work at hand in many ways. Most importantly we showed the feasibility of the process steps “identification and formalization of criticality phenomenon” and “estimation of criticality association,” cf. Figure 1, through the elicitation of a CP-catalog, and for its translation into the GIDAS database scheme and subsequent risk assessment, we produced initial pieces of evidence for the relevance of a subset of this CP-catalog. In particular, Section 4.3.2 explains how estimating the risk of equation (10) can be used for relevance estimation of CP, leading to the prioritization of CP for causal analysis, for example, “occlusion” [11] or “reduced friction on road” [13].

Furthermore, analyzing the edge cases featuring either few or many CP contributes significant value to a criticality analysis. Subjecting accident cases featuring few (or no) CP to detailed individual case analysis facilitates either completeness of the CP-catalog by uncovering potential gaps or completeness of the GIDAS codebook by adding CP when examining these cases. On the other end of the distribution, we have accident cases with many (for example, 14 or more) CP that lend themselves as examples for studying the emergence of criticality in a concrete example, that is, by studying them in computer simulations [51].

After identifying the most relevant CP in accident scenarios within a criticality analysis, one has to select a set of representative instances for those abstract scenarios. Reconstructions of accidents represent a possibility to obtain valid instances of maximal criticality for the respective scenario classes. These instances can then be used for safety assurance in manifold ways. For example, they can be

- (i) Leveraged to derive realistic but system-independent test scenarios for the analyzed CP as to verify the performance of different systems under test
- (ii) Used as an input to an expert-based case analysis to ensure the validity of causal models within the criticality analysis, or
- (iii) Replayed to simulation tools and subjected to a sensitivity analysis of its parameters

Hence, the performed analysis of the GIDAS database, albeit being a specific source of data, provides a blueprint for how traffic data in general can be used within a criticality analysis.

6. Conclusion

In this work, we decomposed the accident risk associated with CP into exposure, controllability, and severity such that, by using Bayes' theorem, the respective quantities could be estimated using accident data. For this, we analyzed a representative dataset of traffic accidents in urban areas of Germany, provided by the GIDAS database, regarding the presence of a catalog of CP. Of the 166 CP elicited by the criticality analysis with respect to ADS at the SAE Level 4 and 5 in urban areas, 116 could be identified within the GIDAS database and were translated to their respective SQL queries. Based on the resulting 15417×116 case-phenomenon relation matrix, we managed to estimate the risk of an accident with passenger car involvement and damage to the persons in the urban areas of Germany associated with a single CP as well as combinations thereof. On the one hand, the resulting estimates can be used as reference values in a safety case that relies on demonstrating a positive risk balance within the scenario classes. On the other hand, this risk assessment allows for estimating the criticality association of CP within a criticality analysis before subjecting them to causal analysis.

Concerning future work, the most obvious extension is to perform similar analyses based on the CP-catalog for a more diverse set of traffic data where a subset of CP is identifiable. Moreover, it will be interesting to do so with a more complete CP-catalog, especially when extended to contain CP specific to machine-perception technology. Such efforts will necessarily be complemented by a formalization of CP that is less dependent on the data scheme, as already presented by Westhofen et al. [43].

Appendix

A. Examples of GIDAS Database Queries

In this section, we present three examples of SQL queries that were used in the presented analysis to assess the GIDAS database regarding the presence of CP.

A.1. Database Query: Intersecting Planned Trajectories of TPs.

```

1 SELECT [01_Umwelt].FALL, [01_Umwelt].UTYP, [01_Umwelt].STFUHO, [01_Umwelt].
   VKREG, [01_Umwelt].HURSU
2
3 FROM [01_Umwelt]
4
5 WHERE ([01_Umwelt].UTYP LIKE '2%' OR [01_Umwelt].UTYP LIKE '3%' OR [01
   _Umwelt].UTYP LIKE '4%' OR [01_Umwelt].UTYP IN (721, 722, 723)) AND [01
   _Umwelt].UTYP NOT IN (201, 215, 231, 251, 252, 259, 261, 262, 286, 306,
   323, 326)
6
7 GROUP BY [01_Umwelt].FALL, [01_Umwelt].UTYP, [01_Umwelt].STFUHO, [01_Umwelt].
   VKREG, [01_Umwelt].HURSU

```

A.2. Database Query: Nonego-TP Violating Right of Way.

```

1 SELECT [01_Umwelt].FALL, [01_Umwelt].UTYP, [01_Umwelt].STFUHO, [01_Umwelt].
   VKREG, [01_Umwelt].HURSU
2
3 FROM [01_Umwelt]
4 INNER JOIN [02_FZG] ON [01_Umwelt].FALL = [02_FZG].FALL
5 INNER JOIN [20_REKO] ON [02_FZG].FALL = [20_REKO].FALL AND [02_FZG].BETNR =
   [20_REKO].BETNR
6 INNER JOIN [26_BETEIL] ON [26_BETEIL].FALL = [20_REKO].FALL
7
8 WHERE [02_FZG].BETNR != [26_BETEIL].BETNR AND [02_FZG].FART = 3 AND ([26
   _BETEIL].URSWIS1 BETWEEN 27 And 32 OR [26_BETEIL].URSWIS1 BETWEEN 60 And
   66 OR [26_BETEIL].URSWIS1 IN (24, 25, 37) OR [26_BETEIL].URSWIS2
   BETWEEN 27 And 32 OR [26_BETEIL].URSWIS2 BETWEEN 60 And 66 OR [26_BETEIL
   ].URSWIS2 IN (24, 25, 37) OR [26_BETEIL].URSWIS3 BETWEEN 27 And 32 OR
   [26_BETEIL].URSWIS3 BETWEEN 60 And 66 OR [26_BETEIL].URSWIS3 IN (24, 25,
   37) OR [26_BETEIL].URSWIS4 BETWEEN 27 And 32 OR [26_BETEIL].URSWIS4
   BETWEEN 60 And 66 OR [26_BETEIL].URSWIS4 IN (24, 25, 37))
9
10 GROUP BY [01_Umwelt].FALL, [01_Umwelt].UTYP, [01_Umwelt].STFUHO, [01_Umwelt].
   VKREG, [01_Umwelt].HURSU

```

A.3. Database Query: Occlusion.

```

1 SELECT [01_Umwelt].FALL, [01_Umwelt].UTYP, [01_Umwelt].STFUHO, [01_Umwelt].
   VKREG, [01_Umwelt].HURSU
2
3 FROM [01_Umwelt]
4 INNER JOIN [02_FZG] ON [02_FZG].FALL = [01_Umwelt].FALL
5 INNER JOIN [27_STRASSE] ON [27_STRASSE].FALL = [02_FZG].FALL AND [27_STRASSE
   ].BETNR = [02_FZG].BETNR
6 INNER JOIN [26_BETEIL] ON [26_BETEIL].FALL = [27_STRASSE].FALL
7 LEFT JOIN [18_FGDAT] ON [26_BETEIL].FALL = [18_FGDAT].FALL AND [26_BETEIL].
   BETNR = [18_FGDAT].BETNR
8
9 WHERE [27_STRASSE].VSSICHT IN (1, 3, 4, 5, 6, 8) OR [27_STRASSE].VZSICHT IN
   (1, 3, 4, 5, 6, 8) OR [26_BETEIL].URSWIS1=63 OR [26_BETEIL].URSWIS2=63
   OR [26_BETEIL].URSWIS3=63 OR [26_BETEIL].URSWIS4=63 OR [18_FGDAT].UQFB2
   BETWEEN 2 And 8 OR [01_Umwelt].UTYP IN (404, 405, 411, 412, 413, 414,
   419, 422, 423, 424, 434, 435, 441, 442, 443, 444, 449, 452, 453, 454,
   463, 464, 465, 472, 473, 494) OR [27_STRASSE].SICHTBV IN (1, 3, 4, 8)
10
11 GROUP BY [01_Umwelt].FALL, [01_Umwelt].UTYP, [01_Umwelt].STFUHO, [01_Umwelt].
   VKREG, [01_Umwelt].HURSU

```

Data Availability

While access to the GIDAS database and the associated GIDAS codebook is limited to paying members, the authors made available all other ingredients of the conducted analysis as supplementary material [41]. This repository contains a version of the CP-catalog, as described in Section 3.1, the case-phenomenon relation matrix from Section 4.2 together with the weighting factors, the extrapolation factors, the associated accident severity, and the respective calculations for all frequentist and risk-related quantities. Moreover, we included lists of the identifiable CP ordered according to the associated accident risk on three different levels of severity, as well as a table containing the correlation coefficients for all pairs of CP, sorted by absolute value. Based on these data, all the results described in Section 4 can be reproduced and validated.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research study leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project "Verification and Validation Methods for Automated Vehicles in Urban Environments." The authors would like to thank the consortium for the successful cooperation. Open Access funding enabled and organized by Projekt DEAL.

Supplementary Materials

The provided supplementary material consists of four files [41]. The first file called “criticality-phenomena-catalog” contains the catalog of criticality phenomena (CP-catalog), as described in Section 3.1, on which the presented analysis is based. The second file “criticality-phenomena-risk-calculation” contains a datasheet which consists of anonymized case numbers pointing to the entries in the GIDAS database in the 1st column, the weighting factors w_i in the 2nd column, the extrapolation factors e_i in the 3rd column, and the associated accident severity s_i in the 4th column followed by the $15\,417 \times 116$ case-phenomenon relation matrix $(a_{i,j})$, as defined in Section 4.2. Below these data, the second file contains the calculation of the frequentist quantities of Section 4.2 and the risk-related quantities of Sections 4.3 and 4.4. The third file “criticality-phenomena-sorted-by-risk” contains three ordered lists, each containing the IDs of the 116 identifiable CP ordered according to the risk estimation of (7) for $s \in \{1, 2, 3\}$, respectively. For $s = 1$, the ordering corresponds to the results of Section 4.3.2 and, in particular, Figure 4. The fourth file “criticality-phenomena-phi-coefficient” contains a table ranking all pairs of CP according to the absolute value of the Φ -coefficient, as shown in Section 4.4.2. (*Supplementary Materials*)

References

- [1] SAE Taxonomy, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE International Standard, Warrendale, PA, USA, 2021.
- [2] A. Poddey, T. Brade, J. E. Stellet, and W. Branz, “On the validation of complex systems operating in open contexts,” 2019, <http://arxiv.org/abs/1902.10517>.
- [3] W. Wachenfeld and H. Winner, “The release of autonomous vehicles,” *Autonomous Driving*, pp. 425–449, Springer, Berlin, Germany, 2016.
- [4] N. Kalra and S. M. Paddock, “Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability?” *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.
- [5] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, S. Fabian, and M. Maurer, “Defining and substantiating the terms scene, situation, and scenario for automated driving,” in *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 982–988, Canary Islands, Spain, September 2015.
- [6] C. Neurohr, L. Westhofen, T. Henning, T. de Graaff, E. Möhlmann, and E. Böde, “Fundamental considerations around scenario-based testing for automated driving,” *IEEE Intelligent Vehicles Symposium*, vol. 4, pp. 121–127, 2020.
- [7] International Organization for Standardization, *ISO 21448: Road vehicles – Safety of the intended functionality*, International Organization for Standardization, Geneva, Switzerland, 2022.
- [8] Z. Zhu, R. Philipp, C. Hungar, and F. Howar, “Systematization and identification of triggering conditions: a preliminary step for efficient testing of autonomous vehicles,” in *Proceedings of the 2022 IEEE Intelligent Vehicles Symposium*, vol. 4, pp. 798–805, Aachen, Germany, January 2022.
- [9] A. Adee, R. Gansch, P. Liggesmeyer, C. Glaeser, and F. Drews, “Discovery of perception performance limiting triggering conditions in automated driving,” in *Proceedings of the 2021 5th International Conference on System Reliability and Safety (ICRSRS)*, pp. 248–257, Palermo, Italy, November 2021.
- [10] X. Xing, J. Tong, C. Junyi, X. Lu, and Y. Zhuoping, “An ontology-based method to identify triggering conditions for perception insufficiency of autonomous vehicles,” 2022, <https://arxiv.org/abs/2210.08724>.
- [11] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas, “Criticality analysis for the verification and validation of automated vehicles,” *IEEE Access*, vol. 9, pp. 18016–18041, 2021.
- [12] D. Otte, J. Michael, and H. Carl, “Scientific approach and methodology of a new in-depth investigation study in Germany called GIDAS,” in *Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles*, Eindhoven, Netherlands, November 2003.
- [13] T. Koopmann, C. Neurohr, L. Putze, L. Westhofen, R. Gansch, and A. Adee, “Grasping causality for the explanation of criticality for automated driving,” 2022, <https://arxiv.org/abs/2210.15375>.
- [14] N. Kauffmann, F. Fahrenkrog, L. Drees, and F. Raisch, “Positive risk balance: a comprehensive framework to ensure vehicle safety,” *Ethics and Information Technology*, vol. 24, no. 1, pp. 15–16, 2022.
- [15] International Organization for Standardization, *ISO 26262: Road vehicles – Functional safety*, International Organization for Standardization, Geneva, Switzerland, 2018.
- [16] L. Westhofen, C. Neurohr, T. Koopmann et al., “Criticality metrics for automated driving: a review and suitability analysis of the state of the art,” *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 1–35, 2023.
- [17] Statisches Bundesamt, *Verkehrsunfälle - Fachserie 8*, DESTATIS, Wiesbaden, Germany, 2020.
- [18] S. M. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, “Application of proximal surrogate indicators for safety evaluation: a review of recent developments and research needs,” *IATSS Research*, vol. 41, no. 4, pp. 153–163, 2017.
- [19] W. Damm and R. Galbas, “Exploiting learning and scenario-based specification languages for the verification and validation of highly automated driving,” in *Proceedings of the 2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS)*, pp. 39–46, IEEE, Gothenburg, Sweden, May 2018.
- [20] J. Antin, S. Lee, J. Hankey, and T. Dingus, *Design of the In-Vehicle Driving Behavior and Crash Risk Study: In Support of the SHRP 2 Naturalistic Driving Study*, Transportation Research Board, Washington, DC, USA, 2011.
- [21] R. Eenink, Y. Barnard, M. Baumann, X. Augros, and F. Utesch, “UDRIVE: the European naturalistic driving study,” in *Proceedings of the Transport Research Arena. IFSTTAR*, Paris, France, April 2014.
- [22] J. M. Hankey, M. A. Perez, and J. A. McClafferty, *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets*, Virginia Tech Transportation Institute, Blacksburg, VA, USA, 2016.
- [23] E. Brühning and S. Berns, “IRTAD- international road traffic and accident database,” *Zeitschrift für Verkehrssicherheit*, vol. 39, no. 3, pp. 121–124, 1993.
- [24] E. Brühning and S. Berns, “IRTAD- international road traffic and accident database,” *Zeitschrift für Verkehrssicherheit*, vol. 43, no. 1, 1997.
- [25] J. Ziegler, H. Liers, A. Chanove, and L. Drees, “Systematic approach to the objective evaluation of traffic accident databases,” *ATZ worldwide*, vol. 123, no. 12, pp. 76–79, 2021.

- [26] E. Esenturk, S. Khastgir, A. Wallace, and P. Jennings, "Analyzing real-world accidents for test scenario generation for automated vehicles," in *Proceedings of the 2021 IEEE Intelligent Vehicles Symposium*, pp. 288–295, Nagoya, Japan, December 2021.
- [27] E. Esenturk, A. G. Wallace, S. Khastgir, and P. Jennings, "Identification of traffic accident patterns via cluster analysis and test scenario development for autonomous vehicles," *IEEE Access*, vol. 10, pp. 6660–6675, 2022.
- [28] J. R. Treat, S. T. Nicholas, S. T. McDonald et al., "Tri-level study of the causes of traffic accidents: final report," Institute for Research in Public Safety, Bloomington, Indiana, 1979.
- [29] K. Goniewicz, M. Goniewicz, W. Pawłowski, and P. Fiedor, "Road accident rates: strategies and programmes for improving road traffic safety," *European Journal of Trauma and Emergency Surgery*, vol. 42, no. 4, pp. 433–438, 2016.
- [30] D. Otte, M. Jänsch, and C. Haasper, "Injury protection and accident causation parameters for vulnerable road users based on German In-Depth Accident Study GIDAS," *Accident Analysis & Prevention*, vol. 44, no. 1, pp. 149–153, 2012.
- [31] V. Vinayak, C. Sai, and N. J. Divya, "Autonomous vehicles: disengagements, accidents and reaction times," *PLoS One*, vol. 11, no. 12, Article ID 168054, 2016.
- [32] F. Favaro, S. Eurich, and N. Nader, "Autonomous vehicles' disengagements: trends, triggers, and regulatory limitations," *Accident Analysis & Prevention*, vol. 110, pp. 136–148, 2018.
- [33] A. Sinha, V. Vu, S. Chand, K. Wijayarathna, and V. Dixit, "A crash injury model involving autonomous vehicle: investigating of crash and disengagement reports," *Sustainability*, vol. 13, no. 14, pp. 7938–1050, 2021.
- [34] S. Wang and Z. Li, "Exploring causes and effects of automated vehicle disengagement using statistical modeling and classification tree based on field test data," *Accident Analysis and Prevention*, vol. 129, pp. 44–54, 2019.
- [35] M. Alexandra, R. Boggs, and J. Khattak, "Exploring the who, what, when, where, and why of automated vehicle disengagements," *Accident Analysis and Prevention*, vol. 136, Article ID 105406, 2019.
- [36] B. Mohamed, P. Andreas, Z. Adrian, and E. Lutz, "Eurofot: field operational test and impact assessment of advanced driver assistance systems: final results," *Proceedings of the FISITA 2012 World Automotive Congress*, pp. 537–547, Springer, Berlin, Germany, 2013.
- [37] N. J. Goodall, "Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data," *Accident Analysis and Prevention*, vol. 154, Article ID 106056, 2021.
- [38] G. Bagschik, T. Menzel, and M. Maurer, "Ontology based scene creation for the development of automated vehicles," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium*, vol. 4, pp. 1813–1820, Changshu, China, June 2018.
- [39] B. Kramer, C. Neurohr, M. Büker, E. Böde, M. Fränzle, and W. Damm, "Identification and quantification of hazardous scenarios for automated driving," in *Model-Based Safety and Assessment*, Z. Marc and H. Kai, Eds., pp. 163–178, Springer International Publishing, Cham, Switzerland, 2020.
- [40] E. Böde, M. Büker, W. Damm et al., "Identifikation und Quantifizierung von Automationsrisiken für hochautomatisierte Fahrfunktionen," Technical Report, OFFIS, Oldenburg, Germany, 2019.
- [41] S. Babisch, C. Neurohr, L. Westhofen, S. Schoenawa, and H. Liers, *Supplementary Material For 'Leveraging The Gidas Database For The Criticality Analysis Of Automated Driving Systems*, Springer, 2023, <https://doi.org/10.5281/zenodo.7704542>, Berlin, GermanySpringer.
- [42] M. Scholtes, L. Westhofen, L. R. Turner et al., "6-Layer model for a structured description and categorization of urban traffic and environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021.
- [43] L. Westhofen, C. Neurohr, M. Butz, M. Scholtes, and M. Schuldes, "Using ontologies for the formalization and recognition of criticality for automated driving," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 5, 2022.
- [44] M. Bäumer, H. Hautzinger, M. Pfeiffer et al., "Fahrleistungserhebung 2014 – Inlandsfahrleistung und Unfallrisiko. Bundesanstalt für Straßenwesen (BASt)," 2017, https://bast.opus.hbz-nrw.de/opus45-bast/frontdoor/deliver/index/docId/1775/file/BASt_V_291_barrierefreies_Internet_PDF.pdf.
- [45] B. B. Park, Y. Chen, and J. Hourdos, "Opportunities for preventing rear-end crashes: findings from the analysis of actual freeway crash data," *Journal of Transportation Safety & Security*, vol. 3, no. 2, pp. 95–107, 2011.
- [46] J. Tao, Y. Li, F. Wotawa, H. Felbinger, and M. Nica, "On the industrial application of combinatorial testing for autonomous driving functions," in *Proceedings of the 2019 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, pp. 234–240, IEEE, Xi'an, China, April 2019.
- [47] G. U. Yule, "On the methods of measuring association between two attributes," *Journal of the Royal Statistical Society*, vol. 75, no. 6, pp. 579–652, 1912.
- [48] J. M. Scanlon, K. D. Kusano, T. Daniel, C. Alderson, A. Ogle, and T. Victor, "Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain," *Accident Analysis & Prevention*, vol. 163, Article ID 106454, 2021.
- [49] United Nations Economic Commission for Europe (UNECE), *UN Regulation No. 157: Uniform Provisions Concerning the Approval of Vehicles with Regard to Automated Lane Keeping Systems*, United Nations Economic Commission for Europe (UNECE), Geneva, Switzerland, 2022.
- [50] T. Brade, B. Kramer, and C. Neurohr, "Paradigms in scenario-based testing for automated driving," in *Proceedings of the 2021 International Symposium on Electrical, Electronics and Information Engineering*, pp. 108–114, Coimbatore, India, February 2021.
- [51] J. S. Becker, T. Koopmann, B. Neurohr et al., "Simulation of abstract scenarios: towards automated tooling in criticality analysis," *Autonomes Fahren. Ein Treiber Zukunftsiger Mobilität*, Zenodo, Paris, France, pp. 42–51, 2022.