

Received 12 September 2023, accepted 21 November 2023, date of publication 7 December 2023, date of current version 2 February 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3340442

RESEARCH ARTICLE

Test Scenario Fusion: How to Fuse Scenarios From Accident and Traffic Observation Data

MAXIMILIAN BÄUMLER¹ AND GÜNTHER PROKOP²

Chair of Automobile Engineering, Technische Universität Dresden, 01069 Dresden, Germany

Corresponding author: Maximilian Bäumlér (mb@mbaeumler.de)

This work was supported in part by Technische Universität Dresden (TU Dresden), in part by the Carl Gustav Carus Faculty of Medicine, in part by the SLUB Dresden, in part by the Open Access Publication Funding of the DFG, in part by the Motus Project under Grant 19FS2015A, and in part by the mFUND Grant from Germany's Federal Ministry for Digital and Transport.

ABSTRACT Scenario-based testing will help to validate automated driving systems (ADS) and establish safer road traffic. To date, most data-driven test scenario generation methods rely primarily on one data source such as police accident data (PD), naturalistic driving studies, or video-based traffic observations (VOs). However, none of these data sources perfectly satisfies all the layers of the six-layer model for the description of test scenarios. Moreover, not all available data sources cover the same location and period of time. Therefore, we fused information from 1,648 scenarios extracted from a German VO with information from 74 scenarios extracted from German PD into a comprehensive new PD* database. Finally, PD* consisted of 74 accident scenarios extended, for example, by variables containing the dynamic information of the VO scenarios. Thus, PD* contained more than 350 variables, whereas PD contained only 269 and VO only 122 variables. For fusion, we followed the Find-Unify-Synthesize-Evaluation (FUSE) for Representativity (FUSE4Rep) process model using statistical matching. Subsequently, we derived three logical scenarios from PD* to test an autonomous emergency braking system (AEB) in a stochastic traffic simulation incorporating driver-behavior models. The quality of the fusion itself was satisfactory, and we propose improving the VO data collection process and observation time to obtain even better results.

INDEX TERMS Advanced driver assistance systems, autonomous driving, system validation, vehicle safety, road traffic.

I. INTRODUCTION

The second wave of automated driving is starting right now [1], aiming to make road traffic safer by introducing automated driving systems (ADSs). Thus, ADSs should drive more safely than attentive human drivers to achieve safer road traffic [2]. One way of proving that ADSs drive safer than human drivers is to compare their safety performance with that of human drivers in selected test scenarios representing real-world traffic. In the early development stage of ADSs, these prospective assessments are performed virtually, for example, by comparing the number of accidents per test scenario for human drivers and ADSs in a large number of virtual simulation runs. To obtain the greatest possible variety of differently executed scenarios over all simulation runs, stochastic simulations can be used, which can vary

test scenario-specific parameters, such as ego-/agent speed, stochastically per simulation run [3]. Moreover, they can also incorporate driver behavior models to compare ADSs with human driving behavior [3], [4]. The idea of comparing the performance based on test scenarios emerges from the so-called scenario-based testing approach [5], which attempts to compress the interesting parts of the road traffic in the ADS operational design domain (ODD) into test scenarios. The ODD describes the domain and conditions for which the ADS is developed to operate securely, such as the road layout, speed ranges, and environmental conditions [6]. Generally, test scenarios are “a kind of flip book representing the temporal sequence of scenes with different actions (e.g. lane change) and events (e.g. collision)” [7, p. 226]. The description of the test scenarios can follow the six-layer model (6LM) [8], specifying, for example, road networks, traffic guidance objects, dynamic objects, and environmental conditions. Additionally, for ADS comparisons with human

The associate editor coordinating the review of this manuscript and approving it for publication was Xiangxue Li.

driver behavior, information on the road users involved, such as the driver's age and driving experience [4], is also helpful in the scenario description. Furthermore, Menzel et al. differentiate between logical and concrete scenarios [9]. Logic scenarios, suitable for stochastic traffic simulations [3], [7], provide parameter ranges of the information contained, whereas concrete scenarios, suitable for proving ground tests, provide specific parameters from the ranges.

One way to generate test scenarios is to extract them from the road traffic data [10]. Possible road traffic data sources include police accident data (PD), video-based traffic observations using drone/stationary cameras (VOs), and naturalistic driving studies (NDSs) [10], [11]. Ideally, one would continuously collect all of these different road traffic data in the corresponding ODD of the ADS to obtain a holistic scenario catalog covering the entire traffic event of the ODD, including accident and critical, complex, and normal driving scenarios [3], [12], [13]. In other words, the test scenario catalog should be representative of the ODD traffic event, meaning that the distributions of the scenarios in the test scenario catalog and in real traffic are similar at the time of the ADS testing. However, there are three main drawbacks that hinder the creation of holistic and representative test scenario catalogs:

- 1) **Limited availability:** To date, there has been no continuous, representative [13], or standardized traffic data collection covering real-world driving behavior, comparable to the nationwide collections of road traffic accidents by the police [14]. So far, various VO [11], [15], [16] and NDS [11], [17] data sets exist, which show, however, only a temporally and spatially limited section of road traffic, for example, two intersections over three months [16]. Owing to the large organizational and financial efforts involved, for example, continuous drone VOs of all ODD-relevant traffic sections are unrealistic in the future.
- 2) **Differing information:** Existing data sources vary in their information content available to describe all six layers of the 6LM and information on road users. For example, the German PD provide conflict situations leading to accidents ("AccidentType") and personal information about the road users involved (e.g., age, driving experience, car details) [18]; however, they do not contain any dynamic information (trajectory, speed course, etc.) of the road users to derive logical/concrete scenarios. By contrast, VOs provide dynamic information for all road users, such as detailed trajectories, traffic volume, and overall traffic behavior, but do not contain any personal information that is not visible to the sensor/camera. Finally, NDSs can provide personal and dynamic information of the road users involved, but are limited to the information perceivable by the sensors from the road users' vehicles participating in the study – the surrounding traffic is only perceived selectively here, depending on the "ego-vehicle-view".

- 3) **Reliable coverage:** Even if all data are available, it is almost impossible to reliably identify all relevant scenarios using supervised, unsupervised, or rule-based scenario identification approaches [10]. Owing to the large amount of necessary road traffic data, there are always unknown scenarios that cannot be identified by the algorithms used – be it due to missing rules, missing training data or incorrect hyperparameterization.

To overcome these drawbacks, we propose two approaches:

First, we propose fusing concrete scenarios [9] identified in different data sources to overcome the challenges of limited availability (1) and differing information (2), motivated by the Find-Unify-Synthesize-Evaluation (FUSE) for Representativity (FUSE4Rep) process model [7], [19]. The fusion of the scenarios identified from PD and VO can create a representative accident test scenario catalog containing the dynamic and personal information of the road users involved. Using this representative test scenario catalog, the following question can be answered to identify edge cases for testing: *Which is the most likely police-recorded traffic conflict leading to accidents, and how should it be parameterized?* To create the test scenario catalog, the FUSE4Rep process model uses statistical matching (SM) and provides a general procedure for fusing scenarios identified in road traffic data [7], [20]. SM is a technique to combine "statistically heterogeneous samples to construct a new sample that can be regarded as having come from an unobserved joint distribution of interest" [21, p. 6]. In contrast to other fusion techniques such as record linkage, the samples to be fused do not require identical observations linked by a unique identifier [21].

Second, we propose to derive logical scenarios from the identified (concrete) scenarios and vary them in a stochastic traffic simulation to overcome the coverage challenge (3), as shown in [3]. Thus, the accident scenarios are extended to normal driving and critical scenarios. Therefore, scenarios that were not included in the real-world data sample or were identified in real-world data are revealed using the stochastic component of the simulation.

Building on these two proposals, this study explores the following overall research question (RQ): *How can a representative test scenario catalog applicable for testing in stochastic traffic simulations be derived by fusing scenarios identified in PD and VO using the FUSE4Rep process model?* To answer this question, we applied the FUSE4Rep process model for the first time completely. The aim was to fuse the scenarios identified in the PD and VO data [16] from selected intersections in Dresden, Germany. To provide a practical example, the scenarios to be fused are relevant for a hypothetical autonomous emergency braking system (AEB), which supports drivers in car-car conflicts. The contributions of this study are as follows:

- Demonstration of all necessary steps for fusion, starting with collection of PD and VO data.
- Fusion of the scenarios identified in PD and VO using SM.

- Derivation of three exemplary logical scenarios for application in a stochastic traffic simulation to test a hypothetical AEB.

In the following section, we first introduce the existing data-driven scenario generation approaches, the basics of SM, and the FUSE4Rep process model. Next, we demonstrate how to apply the FUSE4Rep process model to the PD and VO data, and how to derive logical test scenarios. Finally, we conclude the paper with a discussion and directions for future research.

II. BACKGROUND/RELATED WORK

The following section provides an overview of the current data-driven scenario generation approaches and SM. Moreover, the section introduces the FUSE4Rep process model.

A. PROCESS OF DATA-DRIVEN TEST SCENARIO GENERATION

To provide an overview of the current data-driven test scenario generation approaches, we extended the data-driven test scenario generation process proposed in [10], [22], and [23], and added the proposed 5th step of scenario fusion to overcome the challenges of limited data availability and differing information (see Figure 1).

[Steps 1 & 2] After determining the ADS under test (SuT) and specifying the corresponding ODD in step one, step two requires the minimum types of data sources required to generate the test scenarios. To date, only a few approaches have required more than one primary data source to identify the scenarios. When using more than one primary data source, the approaches rely either on a combination of real driving (NDS/VO) and accident data [7], [19], [24], [25], [26], [27] or on a combination of real driving and simulated/synthetic data [28], for example, recorded in driving simulators.

[Step 3] Step three asks how the chosen primary data are collected. To date, most of these methods are based on existing accident or real driving data sets, as listed in [10], and [11]. Frequently used VO data sets are the highD [29] and inD [30] data sets. Currently, only [7], [19], [31] emphasize to perform real driving data collection following a fixed sampling plan to obtain a representative sample of the road traffic of the ODD.

[Step 4] Subsequently, step four requires the identification and extraction of scenarios, which is mostly done using rule-based approaches: rule-based approaches alone [7], [19], [22], [24], [25], [26], [27], [28], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55] or combined with unsupervised [56], [57], [58] or supervised [59], [60], [61], [62], [63], [64] approaches. An example of a rule-based approach applied to VO data is to first determine the trajectories of road users and then use surrogate safety measures (SSMs), such as time-to-collision (TTC), to detect critical scenarios [19].

[Step 5] Step five (optional) requires the fusion of scenarios identified from different road traffic data sources.

To date, [25] has matched the three-digit accident type (3AT) classification [65] used by the German police to driving situations from a NDS by comparing attributes such as the ego-maneuver using record linkage. However, [25] did not aim to create a statistically valid and representative database. In contrast, [26] fused VO and PD data for one intersection in Germany, Dresden, using SM. Hereby, the SM-process assigned traffic densities and average velocities, extracted from VO data, to past accidents based on time-related variables, such as day, hour, and minute [26]. Consequently, the matching in [26] did not create test scenarios, which is also due to the low level of common information, that is, common variables, shared between the two data sources to be fused. Therefore, [7] proposed the FUSE4Rep process model (see Section D) to maximize the common information to be fused by designing the data collection accordingly. Thus, [19] illustrates how to collect and prepare VO data following the FUSE4Rep process model. Consequently, [19] identified scenarios in VO data according to the 3AT classification scheme using a rule-based approach. However, a complete fusion of the scenarios identified in the PD and VO data using SM has not yet been achieved. We want to emphasize, that the fusion of e.g. weather-related information to already identified scenarios using timestamps, as proposed by [24], belongs more to the field of “data enrichment” relying on record linkage [21].

[Steps 6 & 7] Finally, steps seven and eight, which are not the focus of this study, aim to generate executable test scenarios by using e.g., parameter sampling or driver models [10], and to evaluate the created scenarios regarding, for example, their criticality [10].

B. APPLICATION OF GENERATED TEST SCENARIOS IN STOCHASTIC TRAFFIC SIMULATION

Scenario-based testing of ADSs in virtual traffic simulations will help compare ADSs with the current traffic event and thus help answer whether ADSs:

- reduce the occurrence of accidents/critical situations in the scenarios tested,
- mitigate the consequences of accidents/critical situations in the scenarios tested,
- and cause new critical situations/accidents [7], [66].

Representing a wide range of possible virtual traffic simulations, we focus on comparing ADS driving behavior with human driving behavior in a stochastic traffic simulation [3], [66]. Owing to the stochastic approach, only logical scenarios are required, which allows a new parameterization in each simulation run. Here, the required logical scenarios must contain the following information on the dynamic objects at a minimum:

- start positions of ego vehicles and agents (i.e., other road users),
- maneuvers of ego vehicles and agents (for example, going straight, turning left, or turning right),

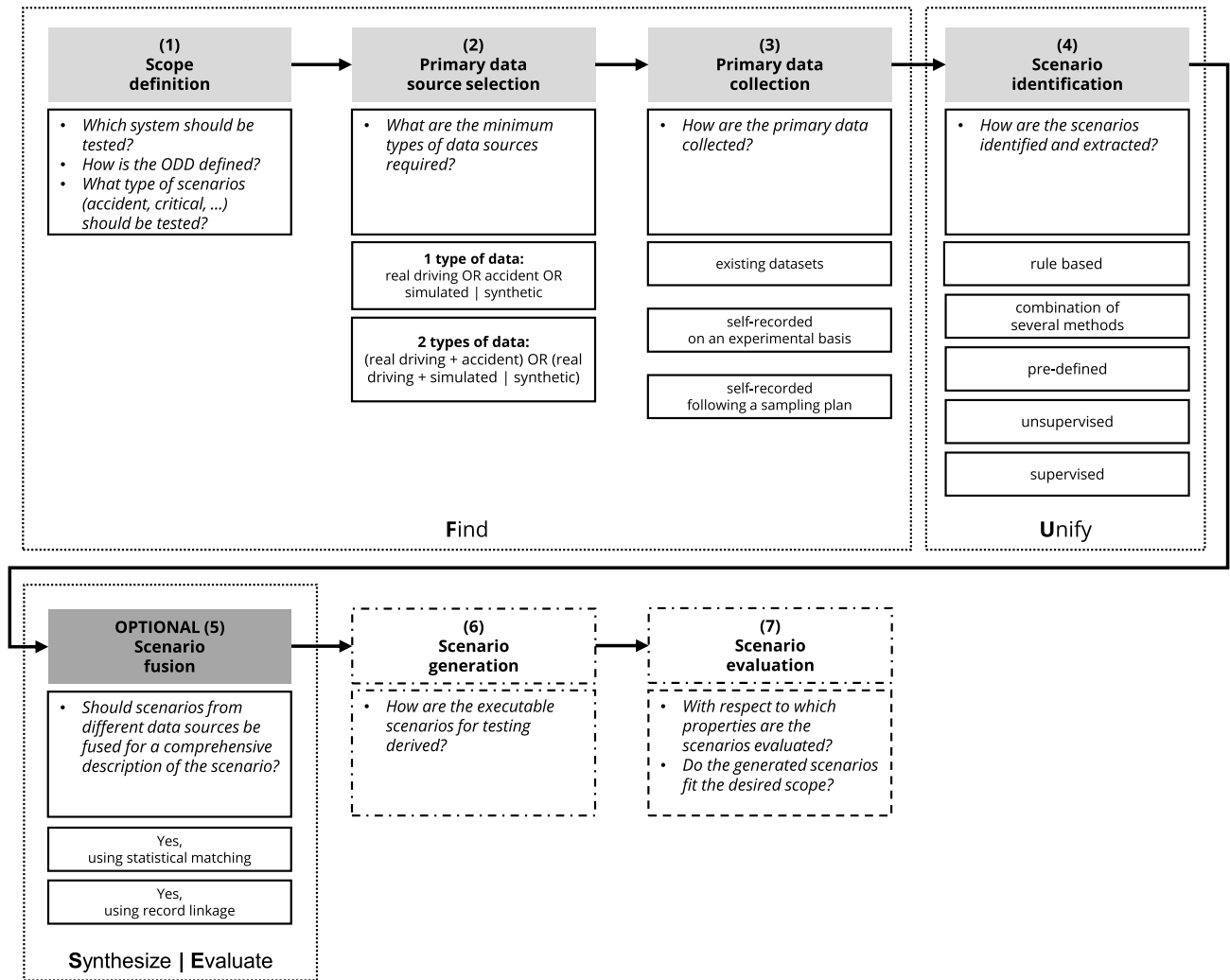


FIGURE 1. Data-driven scenario generation process based on [10], [22], and [23]. The proposed 5th step of scenario fusion to overcome the challenges of limited availability and differing information of data sources is highlighted. The overall process was mapped to the FUSE4Rep process model [7]. Steps six and seven are masked out because they are not in the main focus of this study and do not belong to the FUSE4Rep process model.

(c) and speed at start (e.g., min/mean/max) of ego vehicles and agents [3], [7].

Interestingly, the 3AT classification, originally developed for PD [64], [65] and applied by [19] to identify scenarios in VO data, helps determine both the start positions and maneuvers of the ego vehicle and agents. As stated, for ADS comparisons with human driver behavior, information on the road users involved, such as the driver’s age and driving experience, can also be helpful [4]. To date, the scenarios used in [3] for comparing an AEB system to human driving behavior have been selected based on accident statistics and expert opinion, and parameterized based on a literature review and a descriptive analysis of VO data.

C. STATISTICAL MATCHING (SM)

SM originates from social sciences [21] and aims to combine information from independently acquired samples into a new synthetic one [20]. The new synthetic sample can then

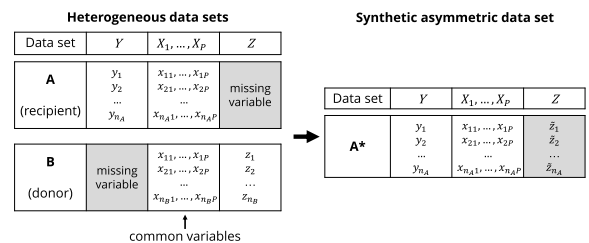


FIGURE 2. Schematic representation of an asymmetric data fusion using statistical matching [26].

be considered as a sample of a virtually joint distribution emerging from the fused samples [21]. Therefore, the two samples must already address the same population before fusion; for example, all traffic conflicts in a specific space and time [19].

Figure 2 [26] illustrates the SM for two independently acquired data sets A and B. Both data sets have some

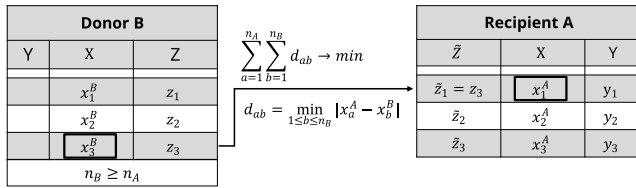


FIGURE 3. Example of the distance-hot-deck-method (DHD) [26].

variables X_I, \dots, X_P in common, as well as some missing variables: Data set **A** misses variable Z , and data set **B** misses variable Y . In the case of asymmetric SM, the smaller data set – here **A**, receives information about Z from **B**. Thus, **A** is also called “recipient” and **B** is called “donor”. After the fusion, **A*** represents a new, complete, and synthetic micro-data set containing all variables, X_I, \dots, X_P, Y , and Z . Depending on the fusion quality, **A*** can be used for further analysis.

For data sets containing categorical variables, such as PD, non-parametric fusion methods are suitable [67]. In the case of fusing PD with VO data, suitable non-parametric fusion methods, such as the distance-hot-deck (DHD) method, use the Gower distance metric [26], [68]. Furthermore, (ensembles of) machine learning classifiers are also suitable, such as random forests and their boosted derivatives, support vector machines, and neural nets [26], [69], [70]. However, because we want to fuse unique scenarios described by several single variables in this case, the aforementioned single-label machine learning classifiers are not applicable; thus, we focus on DHD, as illustrated in [26]. Furthermore, DHD outperforms other hot-deck methods, such as random-hot-deck, and also outperforms random forests in preserving marginal distributions (see the upcoming validity proof), according to [26]. We also emphasize that the Gower distance metric can handle categorical and metric variables simultaneously [67].

As presented in Figure 3 [26], DHD aims to replace the missing variable Z of the recipient by comparing the common variables **X** of **B** and **A**. Usually, DHD does not use all common variables **X**, but only these explaining the missing variable Z the best – this subset of the common variables X_{MV} used for comparing / fusing is also called matching variables (MV). In detail, DHD selects a suitable observation to be fused from **B** by minimizing the local and global distances of the MVs following a distance function, which here is the Gower distance, suitable for handling categorical and non categorical variables [67]. In constrained mode, the DHD selects every observation only once for fusing, whereas in unconstrained mode, the DHD can select every observation multiple times. When two observations have the same distance, DHD randomly chooses.

After fusing successfully, we can check the validity of the fusion using a plausibility check performed by experts [26] by comparing the results with real-world data and by comparing the (marginal) distributions [20]. Specifically, the (marginal) distributions of Z in **A***, that is f_Z , and **B**, that is f_Z , as well as

the common distributions of Z with single MVs in **A***, that is $f_{ZX_{MV}}$, and **B**, that is $f_{ZX_{MV}}$, must coincide [20]:

$$\tilde{f}_Z = f_Z \text{ and } \tilde{f}_{ZX_{MV}} = f_{ZX_{MV}} \quad (1)$$

To generate logical scenarios, the preservation of the (marginal) distributions is more important than assigning every observation correctly, which is measured by the “hit rate” [20]. This means that for logical scenario generation, which relies on parameter ranges/distributions, the fulfillment of the best possible hit rate, as a strong, never measurable indicator for valid data fusion [20], is not in focus.

D. THE FUSE4REP PROCESS MODEL

In the following, we present the Find-Unify-Synthesize-Evaluation (FUSE) for Representativity (FUSE4Rep) process model in detail, as proposed in [7], for fusing PD and VO data. In addition, we mapped the FUSE4Rep process model to the corresponding steps of the data-driven scenario generation process (see Figure 1). The FUSE4Rep process model incorporates the following four steps [7]:

- Find a possible common population and information shared between the PD and VO data. This information serves as the basis for shared variables. Accordingly, using smartphone apps in VO data collection to collect additional common information, such as the incorrect behavior of road users, is reasonable [19].
- Unify the possible common information in the form of shared variables; for example, identify scenarios according to the 3AT scheme in both data sets [19], [64]. This also includes an assessment of the crash risk of the scenarios identified in the VO data using generalized extreme value distributions (GEV) [19], [71].
- Synthesize both data sets by SM.
- Evaluate the fusion using statistical indicators, comparable real-world data, or expert opinions.

Regarding the data-driven scenario generation process, the step “Find” depends on the “scope definition” (1), the “primary data source selection” (2) and the “primary data collection” (3). In contrast, the step “Unify” belongs to the “scenario identification” (4) step because, for example, determining the 3AT is a type of scenario identification. Finally, the steps “Synthesize” and “Evaluate” belong to the optional fifth step “scenario fusion”.

E. RESEARCH GAP AND OBJECTIVE

The overall objective is to fuse PD and VO data for the first time using the FUSE4Rep process model to be able to derive a representative test scenario catalog, applicable for testing in stochastic traffic simulations. Therefore, we consider the following four RQs in detail to answer the overall RQ stated in the Introduction:

- 1) What are the suitable MVs for fusing the PD and VO data?
- 2) Does the constrained or unconstrained DHD perform better?

- 3) How good are the final fusion results obtained?
- 4) Is it possible to derive logical scenarios applicable for testing in stochastic traffic simulations?

III. METHODOLOGY

The methodology mainly follows the four steps of the FUSE4Rep process model [7] and is based on the exemplary data fusion presented in [26], as well as on VO data preparation according to the FUSE4Rep process model presented in [19]. We use the terms “accident” and “crash” in the following as synonyms, considering only accidents with two participants, excluding e.g. accidents with only one participant due to driving errors. We also use the terms “scenario” and “observation” synonymously: the scenario identified in the PD/VO data is equivalent to an observation stored in the data sets. Finally, we use the following convention regarding the different sets of variables considered:

- **Shared variables X_S** : All variables with the same name, content, and categories in PD and VO.
- **Common variables X_C** : All shared variables, which have a similar distribution in PD and VO and thus belong to the same common population.
- **Specific variables Z** : All variables that are specific to the donor data set, that is, VO.
- **Matching candidates X_{MC}** : All common variables investigated as candidates for matching variables.
- **Matching variables X_{MV}** : All matching candidates that are finally selected as matching variables for fusion.

A. FIND

The first step “Find” encompasses all steps to find a common population and common information between the PD and VO data [7] because a shared, unobserved superordinate population is essential for every data fusion. The identified population must be described spatially, factually and temporally [13], [72]. Following the data-driven scenario generation process (see Figure 1), the common population depends on the ODD of the ADS to be tested. Thus, the acquisition of PD should be based on the ODD defined beforehand. Subsequently, the VO data set should attempt to cover the same ODD as a random sample [13] of the common population. To define a common population, a link between PD and VO is required [19]. According to [73] and [74], every accident is preceded by a traffic conflict, but not every traffic conflict can lead to an accident. In Germany, PD includes the traffic conflict preceding the accident described by the variable “AccidentType” [65]; in some states, AccidentType is also described by the detailed 3AT scheme [65]. The 3AT describes the traffic conflict leading to an accident, whereby a conflict is defined as the simultaneous approach of road users to a road location where they may collide [65]. Consequently, VO can also reveal traffic conflicts [75], which is a possible link between PD and VO. Thus, the shared, unobserved superordinate population can consist of all traffic conflicts that occur in the ODD in a given period. To provide a

TABLE 1. “Find”: All steps.

Result: Common population and information between PD and VO.	
1	Define ODD of the system to be tested.
2	Acquire PD corresponding to the ODD.
3	Define superordinate, shared population between PD and VO.
4	Acquire / conduct VO following the defined population. Try to maximize the information shared between PD and VO.

specific example, a common superordinate population for an AEB system operating at intersections under good weather conditions in 2023 in Germany could be “*all traffic conflicts occurring in the absence of rain, fog, and snow [Factual] at German intersections [Spatial] in 2023 [Temporal]*”. With respect to the required sample sizes for the PD and VO data, the size of the PD must be sufficient to encompass the variations in the VO at the desired level of confidence and error tolerance and vice versa [26].

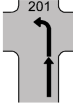

Regarding possible common information, the VO data set/data collection should be oriented towards PD collection, as PD collection is normally standardized and independently performed by the police. For example, in VO, a suitable approach would be to record the incorrect behavior of road users in observed traffic conflicts according to the PD categories with the help of drone pilots conducting VO [19], [76]. Moreover, the 6LM [8] can help find information associated with the six layers between both data sets, whereby the ListDB specification already provides a comprehensive codebook of possible variables [76]. Table 1 summarizes all “Find” steps necessary.

B. UNIFY

The second step “Unify” encompasses all steps to unify the possible common information in the form of shared variables X_S between the PD and VO data [7]. With respect to the required format, both PD and VO should be tabulated such that each row contains an accident or observed traffic conflict described by means of shared variables X_S , as well as variables Y and Z specific to each data set (see Figure 2). As an example of unification, we demonstrate in part how to determine the AccidentType variable according to the 3AT scheme as a shared variable as the link between PD and VO. The detailed instructions can be found in [19] and [64]. First, the 3AT should be, if not available, determined for all accidents contained in the PD by manually analyzing the accident descriptions (see Table 2) with expert knowledge or using machine learning approaches [64].

Second, after determining the set of different 3ATs occurring in the PD, the equivalent conflict situations according to the 3AT scheme must be identified in the VO. The prerequisites for identification are the trajectories of

TABLE 2. Exemplary accident descriptions of 3AT 201 and 3AT 302 recorded by the police (street names were replaced for easier reading). Pictograms provided by [65]. “W” in pictogram indicates a road user, which has to wait / give priority.

	Car 02 drives in front of car 01 on street X towards the city. Car 02 intends to turn left into street Y and thus stops for turning. Car 01 overlooks this and hits car 02, which is in front of it.
	Car driver 01 drives his car on street X in the direction of street Y. When car driver 01 turns left at the junction of streets X and Y onto street Y, he overlooks the car coming from the left (right-of-way regulation by sign 205).

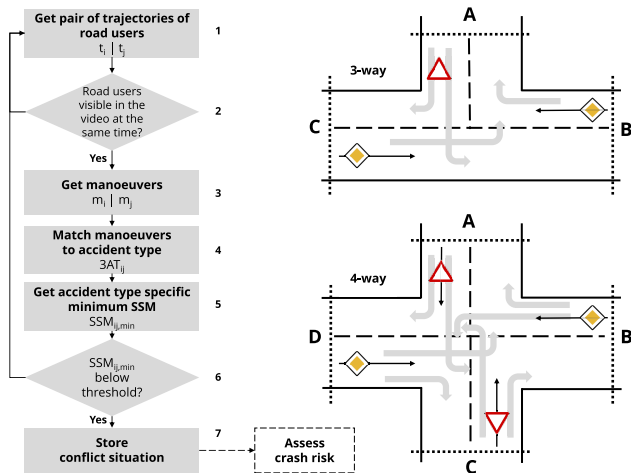


FIGURE 4. Process of 3AT determination in VO according to [19] (left). Possible maneuvers at a 3-/4-way intersection without turn-around (right). Virtual gates are indicated by the letters A-D.

road users already extracted from the video data using commercial [77], open-source [78], or self-developed [19], [79] video analysis frameworks. After determining the maneuvers (going straight, turning left, turning right) and travel directions (e.g., from virtual gate C to virtual gate A at a 3-way intersection) of all road users, the rule-based determination of conflict constellations following the 3AT scheme (for details, see [19]) is applicable (see Figure 4). For example, when two cars at a 3-way intersection are visible in the video simultaneously, they can have a potential conflict constellation according to 3AT 201 [65], when

- the first car, called “agent”, entering the intersection at virtual gate C wants to turn to virtual gate A and thus slows-down/stops to turn, and
- the second car, called “ego”, succeeds car A and goes meanwhile straight.

In the following, the road user causing the conflict is called “ego” and the opponent road user is called “agent”. After identifying a potential conflict constellation, the question is whether it can also be a traffic conflict, that is, whether it bears the potential risk of a crash. Therefore, every identified conflict constellation, called 3AT, is described by

the time course of an SSM to assess its maximum risk over time [80], [81]. Therefore, 3ATs involving a longitudinal approach between ego and agent, for example, 3AT 201, are described by the modified-time-to-collision (MTTC), [82] expanding the concept of time-to-collision (TTC) [83] by considering road users’ accelerations. The TTC describes “the time until a collision between the vehicles would occur if they continued on their present course at their present rates” [80, p. 155]. MTTC was calculated as follows:

$$MTTC = \frac{\Delta v \pm \sqrt{\Delta v^2 + 2\Delta a \Delta s}}{\Delta a} \quad (2)$$

where Δv is the relative speed, Δa is the relative (tangential) acceleration, and Δs is the relative distance between the ego and agent. The MTTC is equal to TTC when $\Delta a = 0$ and $\Delta v > 0$. When $\Delta a \neq 0$ and both results are positive, the MTTC corresponds to the smaller value [82]. When $\Delta a \neq 0$ and one result is negative and one is positive, the MTTC corresponds to the positive result [82]. In contrast, 3ATs, for example, 3AT 302 [65], consisting mostly of perpendicular approaches between ego and agent, are described using post-encroachment-time (PET) [84]. PET is defined as “the time between the moment that a road user (vehicle) leaves the area of potential collision and the other road user [vehicle] arrives [at the] collision area” [80, p. 155] and is calculated as follows:

$$PET = t_2 - t_1 \quad (3)$$

where t_1 is the moment at which the rear bumper of the first vehicle leaves the area of potential collision and t_2 is the moment the front bumper of the second vehicle enters the area of potential collision. After calculating the corresponding SSMs and their time-course $SSM(t)$, only the conflict constellations with a potential risk of crash are stored by removing those with a minimum SSM_{min} greater than five seconds. Hereby, five seconds corresponds to the length of the GIDAS pre-crash matrix [19], [85], which is a simulation format widely used in accident research. Subsequently, for each of the identified and stored 3ATs, the corresponding metadata, including the incorrect behavior of road users, of the VO can be matched. Finally, the population of traffic conflicts identified from VO should also be assessed for their overall risk of crashes using an approach based on real-time crash prediction modeling [19], [71]. Reference [19] proposed using only 3AT populations, with a positive risk of crashes $R_{C,3AT} > 0$. The $R_{C,3AT}$ is calculated as follows [19], [71], and [86]:

$$R_{C,3AT} = G_{3AT}(0) = \begin{cases} \exp\left\{-\left[1 + \xi_i \left(-\frac{\mu_i}{\sigma_i}\right)\right]^{-\frac{1}{\xi_i}}\right\}, & \xi \neq 0 \\ \exp\left[-\exp\left(\frac{\mu_i}{\sigma_i}\right)\right], & \xi = 0 \end{cases} \quad (4)$$

Here, G_{3AT} is the generalized extreme value (GEV) distribution obtained via maximum likelihood estimation based on the SSM_{min} distribution obtained for every type

TABLE 3. “Unify”: All steps.

Result: Unified PD/VO data sets suitable for fusion.	
1	Create a set of shared variables \mathbf{X}_S by aligning variables regarding name, content and categories.
2	Determine the 3AT for PD, when necessary.
3	Determine the 3AT for VO using suitable SSMs.
4	Remove all 3ATs, which do not bear a positive risk of crash.

of 3AT [19]. Specifically, G_{3AT} is described by scale parameter σ , location parameter μ , and shape parameter ξ . After sorting out the 3ATs that do not have a positive crash risk and transforming, if possible, all other variables to the shared variables \mathbf{X}_S , the actual fusion can then begin. Table 3 summarizes all the necessary “Unify” steps.

C. SYNTHESIZE

The third step, “Synthesize” encompasses all steps to fuse both data sets, the donor data set \mathbf{VO} , and the recipient data set \mathbf{PD} into the new data set \mathbf{PD}^* by SM [7]. First, the assumption that both data sets to be fused belong to the same population is verified by investigating the shared variables \mathbf{X}_S with the same name, content, and categories identified previously. The more shared variables exist and the more the distributions of the shared variables in \mathbf{PD} and \mathbf{VO} coincide (the more common variables \mathbf{X}_C exist), the more likely it is that a common population is given. Reference [87] recommended using the Hellinger Distance $H_D \in [0, 1]$ as a similarity measure to compare the distributions of categorical (nominal) shared variables. To compare a shared categorical (nominal) variable x between \mathbf{PD} and \mathbf{VO} , $H_D(p_{x,VO}, p_{x,PD})$ is calculated based on the relative frequencies p_x of the individual categories $j = 1, \dots, J$ of the shared variable x compared:

$$H_D(p_{x,VO}, p_{x,PD}) = \frac{1}{\sqrt{2}} \sqrt{\sum_{j=1}^J (\sqrt{p_{x,VO,j}} - \sqrt{p_{x,PD,j}})^2} \quad (5)$$

The more similar the distributions are, the closer $H_D(p_{x,VO}, p_{x,PD})$ is to zero. Reference [88] refer to the rule of thumb that distributions are similar when $H_D(p_{x,VO}, p_{x,PD}) \leq 0.05$, which must not be valid for every data fusion. Therefore, we determine the threshold for similarity empirically by splitting the VO data set n times randomly into two data sets following the same ratio as \mathbf{PD} to \mathbf{VO} and applying the maximum/median of all H_D measured between the synthetically created \mathbf{VO} data sets as the similarity threshold. Subsequently, all common variables \mathbf{X}_C belonging to the same population are investigated for their suitability as matching candidates for \mathbf{X}_{MC} . In this data fusion, the specific (missing) variables \mathbf{Z} to be fused, are all metric

and assumed to be normally distributed; for example, the variables EgoStartSpeed and AgentStartSpeed. By contrast, the shared variables \mathbf{X}_S are all categorical and nominal and can thus be displayed as dichotomous variables. Thus, we use the point-biserial correlation measure $r_{pb} \in [-1, 1]$ to find the categorical matching candidates x in $\mathbf{X}_{C,VO}$, which explain the (metric) specific variables \mathbf{Z} to be fused best [89]:

$$r_{pb,xz} = \frac{\bar{Z}_1 - \bar{Z}_0}{\frac{\sum Z^2 - (\sum Z)^2}{N-1}} \sqrt{\frac{N_1 N_0}{N(N-1)}} \quad (6)$$

where \bar{Z}_0 / \bar{Z}_1 are the means of the specific variable z when the common variable x is coded 0 / 1, respectively; N_0 / N_1 are the numbers of observations when x is 0 / 1; and N is the number of all observations, that is, the sum of N_0 and N_1 ($= \mathbf{VO}$ size). A perfect positive correlation is given at $r_{pb,xz} = +1$, no association is given at $r_{pb,xz} = 0$, and a perfect negative correlation is given at $r_{pb,xz} = -1$.

Because it is also difficult to define a selection threshold for the selection of matching candidates \mathbf{X}_{MC} using point-biserial correlation, we perform test fusions with the selected matching candidates and thereby identify their best combination/selection. The best combination/selection corresponds to the final matching variables \mathbf{X}_{MV} for the subsequent fusion. The best fusion is that which best preserves the distributions of the specific variables \mathbf{Z} in \mathbf{VO} and the fused data set \mathbf{PD}^* , among others (see Sections II-C and III-D). With regard to fusion using constrained/unconstrained DHD, we consider every scenario/observation stored in \mathbf{VO} as unique. Consequently, all the specific variables \mathbf{Z} that are fused always originate from the same scenario identified by one matching combination of MVs. This approach has the advantage that the parameter ranges/distributions required to describe logical scenarios can be derived flexibly, because values that have already been aggregated over multiple scenarios, such as the mean values of EgoStartSpeeds, are not matched during the fusion. Table 4 summarizes all “Synthesize” steps.

D. EVALUATE

The fourth and last step “Evaluate” encompasses all steps to evaluate the data fusion [7], whereby we do not compare the results with real-world data owing to missing access to, for example, the German-In-Depth-Accident-Study [90] providing comparable accident data with a five-second pre-collision simulation [85]. The metrics used to compare the (marginal) distributions f_Z of \mathbf{VO} and \mathbf{PD}^* and the common distributions $f_{Z_{X_{MV}}}$ of \mathbf{VO} and \mathbf{PD}^* depend on the type of variables compared (metric vs. categorical). In the case of distributions $f_Z / f_{Z_{X_{MV}}}$ formed by categorical variables, we use the Hellinger Distance $H_D(p_{VO}, p_{PD^*})$, as introduced in (5). When comparing distributions f_Z formed by metric \mathbf{Z} , we calculate the two-sample Smirnov test statistic D [91] for $\alpha = 0.05$, which is a common approach for comparing two empirical distributions [91]. The corresponding critical value

TABLE 4. “Synthesize”: All steps.

Result: Completed fusion, resulting in a new data set PD*	
1	Determine similarity thresholds empirically.
2	Identify all shared variables with common distributions, i.e. common variables \mathbf{X}_C .
3	Identify possible matching candidates \mathbf{X}_{MC} from the common variables \mathbf{X}_C , which have the highest explanatory power for the specific variables \mathbf{Z} .
4	Conduct test fusions to find the best matching variables \mathbf{X}_{MV} in the matching candidates \mathbf{X}_{MC} and if fusion should be constrained/unconstrained.
5	Fuse PD and VO using the matching variables \mathbf{X}_{MV} and fusion settings.

TABLE 5. “Evaluate”: All steps.

Result: Quality of fusion is evaluated. PD* can be used.	
1	Determine similarity thresholds empirically.
2	Compare (marginal) distributions f_Z in VO and PD* for similarity.
3	Compare common distributions $f_{Z\mathbf{X}_{MV}}$ in VO and PD* for similarity.
4	[Optional] Judge fusion results using expert knowledge.
5	[Optional] Compare fusion results to corresponding real world data.

of the test statistic $D_{crit,\alpha=0.05}$, forming the upper threshold for judging similarity and used for large samples, such as n_{VO} and n_{PD^*} , is calculated as follows [92]:

$$D_{crit,\alpha=0.05} = 1.36 \sqrt{\frac{n_{VO} + n_{PD^*}}{n_{VO}n_{PD^*}}} \quad (7)$$

However, a comparison of the common distributions $f_{Z\mathbf{X}_{MV}}$ formed by metric \mathbf{Z} and categorical MVs \mathbf{X}_{MV} is not possible. Hence, we compare how the correlation between \mathbf{Z} and \mathbf{X}_{MV} is preserved between **VO** and **PD***, and thus compare the corresponding difference $\Delta r_{pb,\mathbf{X}_{MV}\mathbf{Z},VOPD^*}$ using (6). As stated in Section III-C, applying pre-defined thresholds to judge similarity is difficult. Additionally, relying only on statistical tests may be too strict. Hence, we empirically determine thresholds for all comparisons, as introduced, based on the mean of all metrics computed for n random splits of the donor data set **VO**. Please note that some of the determined thresholds are already necessary in the step “Synthesize” to judge, for example, the common population of the common variables. Table 5 summarizes all “Evaluate” steps.



FIGURE 5. Aerial images [93], [94], [95] of all 20 intersections included in the **PD**. The intersections were mostly connected via primary and secondary roads to ensure a comparable volume of traffic. The two intersections also included in **VO** are largely displayed (Tharandter Straße/Frankenbergsstraße; Kohlenstraße/Dorfhainer Straße). All 20 intersections are located in Dresden, Germany: (Tharandter Straße AND (Altfränkener Straße OR Anton-Weck-Straße OR Chauseehausstraße OR Clara-Viebig-Straße OR Frankenbergsstraße OR Reiserwitzer Straße OR Schillingstraße)) AND (Reiserwitzer Straße AND (Schillingstraße OR Anton-Weck-Straße OR Bonhoefferplatz OR Frankenbergsstraße OR Mohorner Straße)) AND (Chemnitzer Straße AND (Bayreuther Straße OR Bamberger Straße OR Hegerstraße OR Bienertstraße)) AND (Kohlenstraße AND (Dorfhainer Straße OR Muldaer Straße OR Döbraer Straße OR Höckendorfer Weg)).

IV. RESULTS

The results follow the structure of the FUSE4Rep process model.

A. FIND

The ODD of the hypothetical AEB for which the test scenarios were identified is defined as follows: “The AEB shall operate in 2022 at 20 different, but in their geometry, traffic volume and traffic signage comparable, inner-city intersections (see Figure 5) in Dresden, Germany, at good weather conditions (dry road, no rain, no fog, no snow, no strong wind) during daylight and shall support the drivers in car-car turning / crossing / longitudinal conflicts.”

To address this ODD, we used for demonstration purposes two, according to the FUSE4Rep process model collected [19], data sets of the representative and publicly available ListDB data set collection [16] as **VO** data: ListDBRepOne [93] and ListDBRepTwo [94] data sets. Both contain each representative three-month **VO** (June to August 2022) at one intersection in Dresden, Germany. According to [19], June to August are accident-prone months in Dresden and may thus lead to more observed traffic conflicts in the data sets used. The ListDBRepOne data set focuses on a three-way intersection, called “Tharandter Straße/Frankenbergsstraße” and encompasses 790 min of video and 10,324 car trajectories. The ListDBRepTwo data set focuses on a four-way intersection,

called “Kohlenstraße/Dorfhainer Straße” and encompasses 855 min of video and 7,570 car trajectories. Both intersections were recorded during daylight at four different time slots throughout the day, four times a month, and under good weather conditions (no rain, fog, snow or stronger¹ wind) [19]. In addition to the original video material and trajectories of all road users analyzed by DataFromSky [77], both data sets contain extensive metadata (122 variables) according to the ListDB specification [76], which is already geared towards the closest possible conformity with PD [19]. Thus, the metadata include variables of environmental information (e.g., temperature and road surface temperature), location information, and information on incorrect behavior (e.g., failure to observe the right of way) of road users [19], [76]. Both intersections are located in the city and have a speed limit of 50 km/h on priority roads.

We then obtained all PD available to us from 01/01/2005 to 12/31/2021² for all 20 intersections (see Figure 5) specified in the ODD (2x VO+18x additional intersections) and filtered them according to the following criteria:

- Car-car accidents not involving trailers.
- No influence of alcohol or drugs was observed, which is not observable in VO.
- In the case of the seven intersections connected with “Tharandter Straße”: Accidents since 2009 due to construction work in 2007-2008, which removed embedded train tracks on the road.
- Plausible accident information coded by the police.
- Available accident description that allows the determination of 3AT.
- 3AT is either of type “turning”, “crossing” or “longitudinal” [65].

Finally, we obtained a PD data set of 74 car-car accidents described by 269 variables and an accident description. Overall, the shared superordinate population between the acquired PD and VO data sets is estimated as:

- **[Factual]** All turning, crossing or longitudinal traffic conflicts that potentially lead to accidents between two cars. Conflicts occur during daylight under good weather conditions (no rain, fog, or snow).
- **[Spatial]** At 20 specified intersections in Dresden, Germany.
- **[Temporal]** From 2005 to 2022.

B. UNIFY

Overall, ten different types of 3ATs describing the conflict situations of the 74 car-car accidents are represented in the PD (see Figure 6): three turning conflicts (3ATs 201, 211, 231), four crossing conflicts (3ATs 302, 303, 321, 322), and three longitudinal conflicts (3ATs 681, 601, 621). According to the ten 3ATs represented in PD, we identified the same conflict constellations in the VO by analyzing the provided

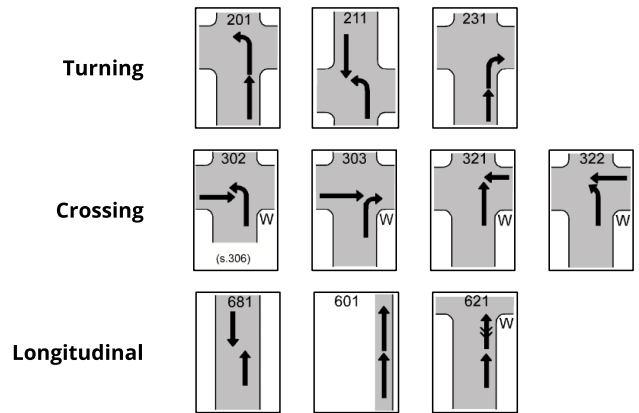


FIGURE 6. Ten different 3ATs represent the conflict situations of 74 car-car accidents in the PD. Pictograms provided by [65]. “W” in pictogram indicates a road user, which has to wait / give priority.

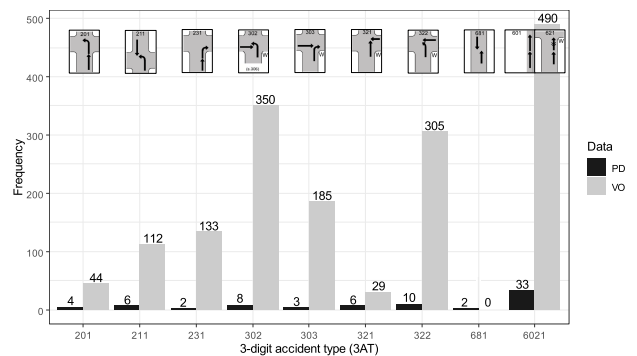


FIGURE 7. Bar chart showing the 3AT distribution in police accident (PD, n=74) and video-based traffic observation data (VO, n=1,648).

trajectory data. However, we collapsed 3ATs 601 and 621 into one category, called “6021”, resulting in the identification of nine different 3ATs. The reason was, that the 3ATs 601 and 621 are difficult to distinguish at the observed intersections. In fact, they differ only in that 3AT 621 specifies a possible reason for the leading car to slow down: it must wait at the intersection entry.

Figure 7 illustrates the distribution of the nine different 3ATs identified in the 74 accidents recorded in PD and in the 1,648 conflict constellations, that is, the scenarios identified in VO. Accordingly, the three most frequent 3ATs are in both data sets: 3ATs 6021, 302, and 322, where 3AT 6021 leads by a large margin to the second most common 3AT in both data sets. Interestingly, 3AT 681 was recorded twice in PD but was never observed in VO.

Table 6 presents 3AT-dependent crash risk $R_{C,3AT}$ estimates based on the calculated minimum SSMs (SSM_{min}). Accordingly, every 3AT GEV distribution has a positive crash risk, meaning that the identified conflict constellations bear the inherent risk of a crash. But, the risk of crashes for 3AT 231, $R_{C,231} = 1e^{-8}$, is relatively small; therefore, inclusion seems questionable. However, when inspecting the empirical and modeled GEV distributions for 3AT 231 (see Figure 8), it is evident that the modeled GEV underestimates

¹The drones were able to fly at wind speeds of up to 10.5 m/s.

²Unfortunately, the PD for 2022 was not available at the time of processing, and only the VO data for 2022 were available.

TABLE 6. Results of modeled extreme value distributions. Minimum and maximum standard error estimates are bold in the columns. Scale σ , location μ , shape ξ .

3AT	SSM metric	n	σ	μ	ξ	$R_{C,3AT}$
			(standard error estimates)			
201	MTTC	44	1.0267 (0.1189)	1.9840 (0.1712)	-0.2305 (0.0973)	0.0071
211	PET	112	1.4213 (0.1432)	2.6236 (0.1584)	-0.5739 (0.1175)	0.0296
231	MTTC	133	0.5086 (0.0451)	0.9125 (0.0513)	0.3664 (0.0874)	$1e^{-8}$
302	PET	350	1.0151 (0.0583)	3.0078 (0.0676)	1.0151 (0.0743)	0.0010
303	PET	185	1.0703 (0.0906)	2.7001 (0.1056)	1.0703 (0.1208)	0.0014
321	PET	29	1.3434 (0.2470)	2.9355 (0.2798)	1.3434 (0.1823)	0.0184
322	PET	305	1.0267 (0.0551)	3.0931 (0.0678)	1.0267 (0.0588)	0.0013
6021	MTTC	490	1.0697 (0.0466)	3.5531 (0.0518)	1.0697 (0.0346)	0.0046

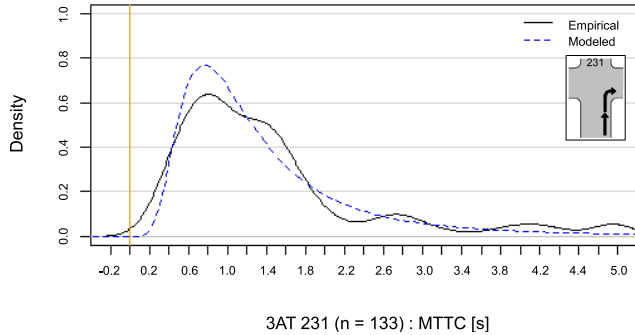


FIGURE 8. Comparison of empirical and modeled extreme value distributions of 3AT 231 with MTTC [s] as SSM. The vertical yellow line marks zero.

the risk of a crash $R_{C,231}$ compared to the real observation (see the left area under the corresponding curve, next to the yellow vertical line). Therefore, we included the 3AT 231 population for subsequent data fusion. Finally, when comparing the standard error estimates (see Table 6), it is remarkable that the modeling quality correlates with the available number of extreme values, and the highest error estimates are observed for the GEV of 3AT 321, which is based on only 29 extreme values.

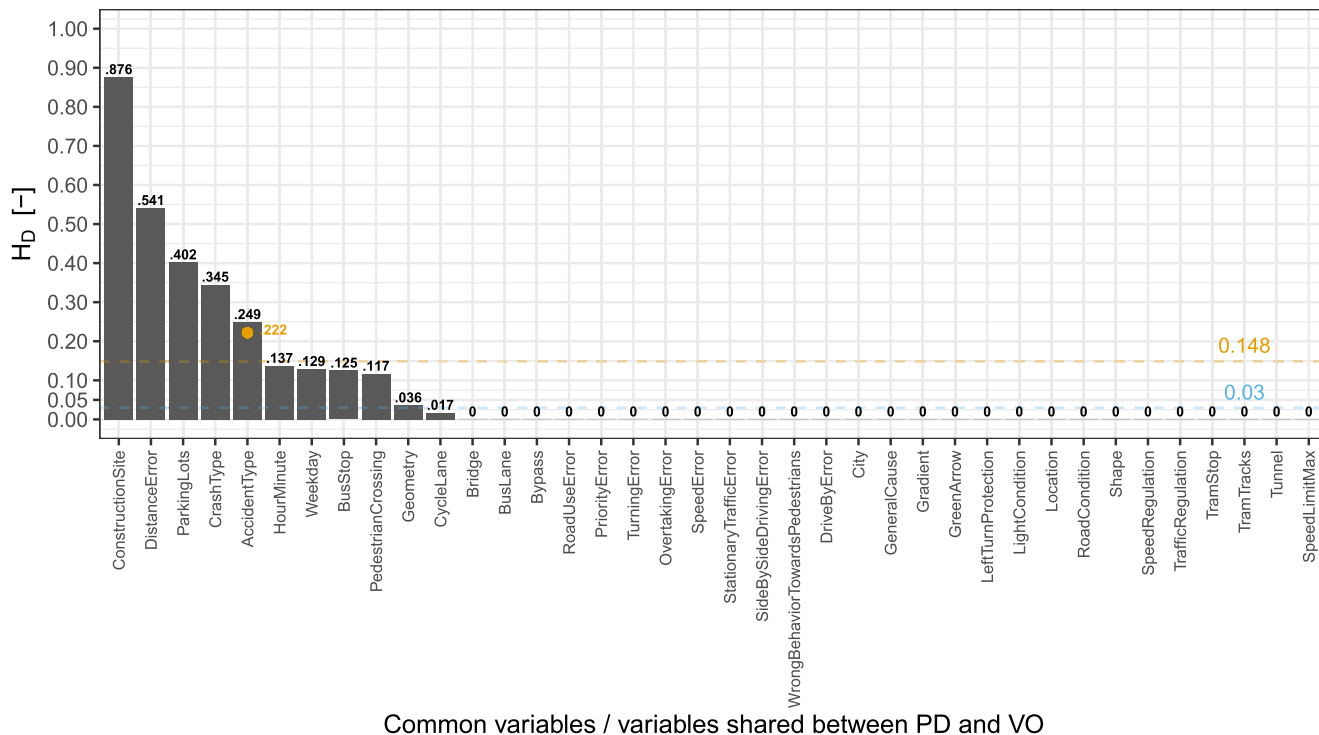
In conclusion, the recipient data set PD encompasses 74 scenarios/observations and the donor data set VO encompasses 1,648 scenarios/observations. Moreover, PD and VO share 38 categorical variables X_S (Table 7), which are further described in the ListDB codebook [76].

TABLE 7. Shared X_S and specific variables Z (alphabetically sorted).

Shared variables X_S		Specific variables Z (in VO)	
	Categorical	Categorical	Metric
AccidentType	Bridge	AgentDirection	Agent Acceleration LatStart [m/s ²]
BusLane	BusStop	AgentEntryGate	Agent Acceleration Start [m/s ²]
Bypass	City	EgoDirection	Agent Acceleration TanStart [m/s ²]
ConstructionSite	CrashType	EgoEntryGate	Agent AverageSpeed [m/s]
CycleLane	DistanceError	Fog	Agent SpeedStart [m/s]
DriveByError	GeneralCause	PublicHoliday	Ego Acceleration LatStart [m/s ²]
Geometry	Gradient	Rain	Ego Acceleration Start [m/s ²]
GreenArrow	HourMinute	RoadUserMost	Ego Acceleration TanStart [m/s ²]
LeftTurn Protection	LightCondition	RoadUser SecondMost	Ego AverageSpeed [m/s]
Location	OvertakingError	Snow	Ego SpeedStart [m/s]
ParkingLots	Pedestrian Crossing	Sunshine	RoadSurface Temperature [°C]
PriorityError	RoadCondition	TrafficJam	SSM _{min} [s]
RoadUseError	Shape	TrafficVolume	Temperature [°C]
SideBySide DrivingError	SpeedError	Wind	WindSpeed [m/s]
SpeedLimitMax	SpeedRegulation		
StationaryTraffic Error	TrafficRegulation		
TramStop	TramTracks		
Tunnel	TurningError		
Weekday	WrongBehavior TowardsPedestrians		

C. SYNTHESIZE

Of the 38 shared categorical variables X_S between PD and VO, 27 did not show any variance between PD and VO, that is $H_D = 0$ (Figure 9). Subsequently, we can assume a common population between PD and VO regarding these 27 variables, whereby 16 of the 27 variables specify the



Common variables / variables shared between PD and VO

FIGURE 9. Hellinger distance for all 38 variables shared between PD and VO. The variables are explained in [76]. The yellow dot indicates the Hellinger distance for the variable “AccidentType” when dropping the 3AT 681 in PD. The dashed lines indicate empirically determined thresholds for similarity (yellow: maximum/blue: median) by measuring the Hellinger distance for 100 random splits of VO according to the ratio of PD/VO.

locations of the VO/PD (e.g., Bridge, BusLane, and Bypass). Of the remaining 11 variables showing a $H_D > 0$, six were below the empirically determined similarity threshold of $H_D = 0.148$ (maximum H_D observed in 100 random splits of VO). However, AccidentType, the variable specifying 3AT, did not seem to share a common population with $H_D = 0.249$, exceeding the (maximum) similarity threshold by 68%. When calculating H_D only for the 3ATs occurring in PD and VO, that is, when neglecting 3AT 681 only occurring in PD, H_D for AccidentType decreased to 0.222, but still exceeded the (maximum) threshold by 50%. Overall, VO and PD appeared to share a common population with 33 common variables: X_C .

To select the matching candidates X_{MC} , we investigated the explanatory power of the common variables X_C for selected specific variables Z (see Table 7), describing mainly the dynamic behavior of the ego at the start of the scenario: EgoSpeedStart, EgoAccelerationStart, EgoAccelerationTanStart, and SSM_{min} . All variables (except SSM_{min}) were calculated using the median of the first ten data points of the corresponding trajectory. Regarding the common variables X_C , 27 common variables showing zero variance between PD and VO are negligible; they cannot be used for matching.

Instead, we also included the shared variables AccidentType and CrashType in the selection process because these variables are assumed to have a high correlation, and thus explanatory power, with Z variables. Figure 10 supports this assumption: The variable AccidentType had

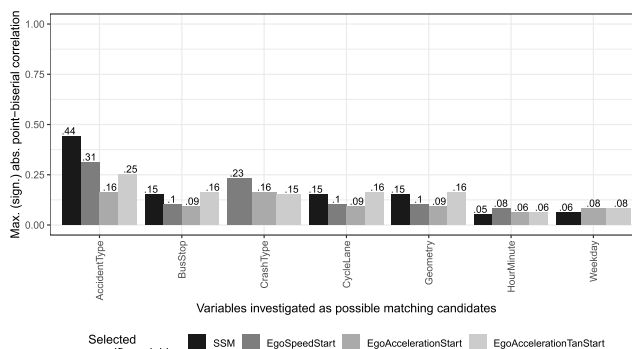


FIGURE 10. Maximum significant ($\alpha = 0.05$) absolute point-biserial correlation between selected specific variables Z and possible matching candidates (measured in VO).

the highest, significant positive correlation with all selected specific variables: SSM_{min} (0.44), EgoSpeedStart (0.31), EgoAccelerationStart (0.16) and EgoAccelerationTanStart (0.25). In contrast, the variables HourMinute and Weekday had the weakest correlations, resulting in a maximum correlation of 0.08 for e.g., HourMinute \sim EgoSpeedStart. Moderate correlations were observed for BusStop (max: 0.16), CrashType (max: 0.23), CycleLane (max: 0.16) and Geometry (max: 0.16). Based on these results, we selected the following five variables as matching candidates X_{MC} for the test fusions, despite not all five variables belonging to the common population between PD and VO: AccidentType, CrashType, Geometry, CycleLane, and BusStop. The variable CrashType specifies the (estimated) type of collision and

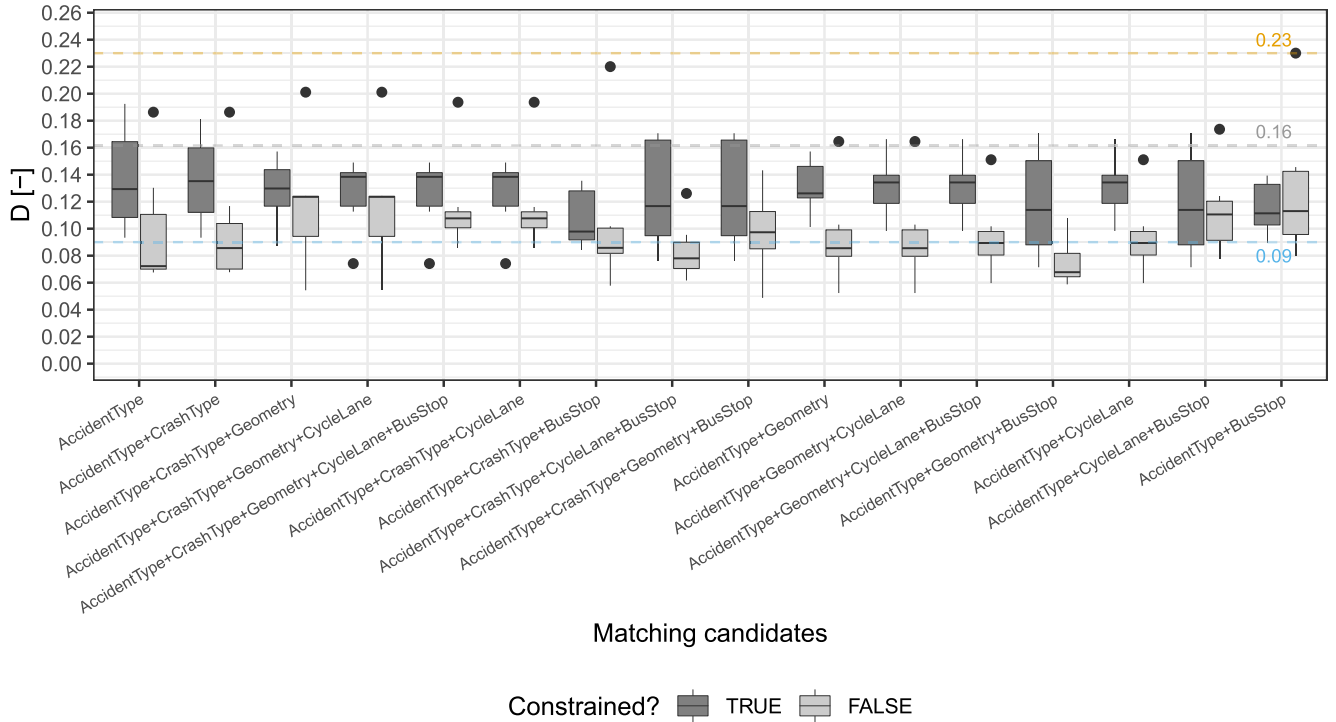


FIGURE 11. Boxplots illustrating the calculated Smirnov test statistics D of the selected specific variables Z between VO and PD*. The results depend on the combination of matching candidates used for the test fusions and the type of DHD (constrained/unconstrained). The blue and yellow dashed lines indicate empirically determined thresholds for similarity (yellow: maximum/blue: median) by measuring the Hellinger distance for 100 random splits of VO according to the ratio of PD/VO. The gray dashed line indicates the critical value of the test statistic ($\alpha = 0.05$; 0.162) under which all distributions can be considered similar.

distinguishes in the present data set between collisions with leading cars, oncoming cars, and turning/crossing cars. The variable CrashType was estimated for VO based on the 3AT during the step “Unify”. The variable Geometry differentiates between three- and four-way intersections. The variables CycleLane and BusStop indicate whether a cycle lane or bus station are located at an intersection, respectively.

In the 32 test fusions, we tried all 16 combinations of the five selected matching candidates X_{MC} , dependent on constrained/unconstrained DHD. However, the variable AccidentType always had to be included because it is the strongest link between PD and VO and had the highest correlation with the specific variables Z . To identify the best test fusion, we compared the distributions $f_{Z_{sel}}$ of the seven most important specific variables, Z_{sel} , between VO and PD*, to save time and computational resources. The most important specific variables Z_{sel} are those that specify the dynamic behavior of ego and agent at the start of the scenario (except SSM_{min}) and are all of metric nature: Ego/AgentSpeedStart, Ego/AgentAccelerationStart, Ego/AgentAccelerationTanStart, SSM_{min} .

Figure 11 illustrates the 32 test fusion results, where all test fusions were below the upper threshold $D_{Z,upper} = 0.23$ (yellow line). 26 of the 32 test fusions produced similar distributions, according to the Smirnov test ($D_{crit,\alpha=0.05} < 0.162$). Seven test fusions were also below the lower threshold

$D_{Z,lower} = 0.09$, whereby the best results were delivered by an unconstrained DHD, relying on the matching candidates AccidentType, Geometry, and BusStop (median: 0.068). Interestingly, in 31 of the 32 test fusions, the unconstrained DHD outperformed the constrained DHD, except for the fusion that used the matching candidates AccidentType and BusStop. Furthermore, more matching candidates did not necessarily lead to better fusion results, as demonstrated by the fusion using all matching candidates; this phenomenon coincides with the experience in [88]. Table 9 (see Appendix) provides an overview of the 74 scenarios with an excerpt of 13 selected variables.

D. EVALUATE

In the following, we evaluate the best fusion PD* (unconstrained; MVs = AccidentType, Geometry, BusStop) by comparing the (marginal) distributions f_Z of all specific variables Z (see Table 7) and the common distributions $f_{Z_{X_{MV}}}$ of specific variables Z with the MVs in VO and PD*.

Figure 12 shows that the median (0.07) of all compared distributions of metric and specific variables Z_{metric} is below all the determined thresholds and thus can be considered similar. Furthermore, all the compared metric distributions of Z_{metric} were below the critical value of the Smirnov statistic ($D_{crit,\alpha} = 0.162$), with the variable AgentAverageSpeed (0.118) closest to the threshold.

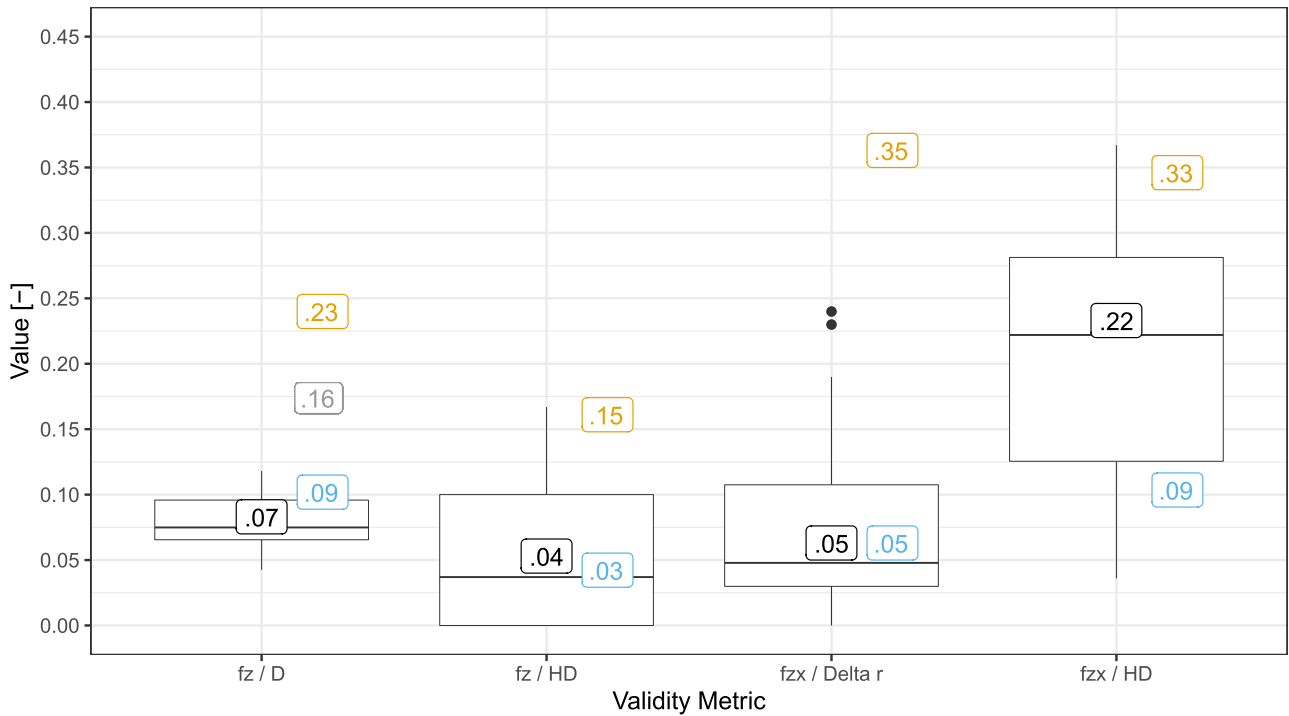


FIGURE 12. Boxplots illustrating the 4th validity level. The blue and yellow labels indicate empirically determined thresholds for similarity (yellow: maximum/blue: median). The gray label indicates the critical value of the test statistic ($\alpha = 0.05$; 0.162) under which all distributions can be considered similar.

Six of the seven most important specific variables, Z_{sel} , were below or equal to the lower threshold $D_{Z,lower} = 0.09$: EgoSpeedStart (0.07), Ego/AgentAccelerationStart (0.059/0.065), Ego/AgentAccelerationTanStart (0.093/0.068), and SSM_{min} (0.063). Only AgentSpeedStart was slightly above the lower threshold (0.108).

In contrast, the median of the compared categorical distributions (0.04) for Z_{cat} was slightly higher than the lower empirical threshold ($H_{D,Z,lower} = 0.03$). However, all categorical specific variables, Z_{cat} , were below the upper empirical threshold ($H_{D,Z,upper} = 0.15$), except for EgoDirection (0.167). Regarding the preservation of the correlation between the metric specific variables, Z_{metric} , and the MVs, X_{MV} , the median of the compared differences in the correlations in VO and PD* exactly coincided with the identified lower threshold ($\Delta_{rpb,MVZ,lower} = 0.05$), and all were below the upper threshold $\Delta_{rpb,MVZ,upper} = 0.35$.

Finally, the common distributions of the categorical specific variables, Z_{cat} , and categorical MVs exceeded the lower threshold, $H_{D,MVZ,lower} = 0.09$, by 144%, resulting in a median value of 0.22. Only 21 out of 42 compared common distributions were below the median of 0.22, and thus, can be considered similar. In addition, the common distributions of AccidentType + Wind (0.356) and BusStop + RoadUserSecondMost (0.367) exceeded the upper threshold $H_{D,MVZ,upper} = 0.33$.

Figure 13 shows the cumulative density function plots for Ego/AgentSpeedStart and Ego/AgentAccelerationTanStart for the VO and PD*. As revealed by the Smirnov test statistic,

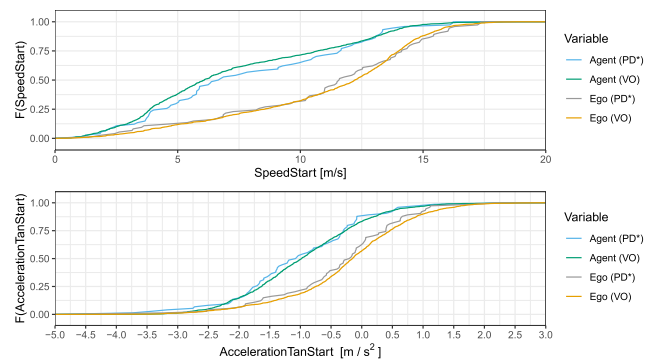


FIGURE 13. CDF plot comparing specific variables “SpeedStart” and “AccelerationTanStart” between VO and PD*.

the AgentSpeedStart curves did not coincide as well as the others, and the fused start speeds in PD* (median: 6.7 m/s) were, on average, 10% higher than the original donor speeds in VO (median: 6.1 m/s). In conclusion, we consider data fusion to be valid for the subsequent derivation of logical scenarios because of the similar distributions of VO and PD*. Moreover, the unconstrained DHD used 73 different VO scenarios to fuse them with the 74 scenarios contained in PD*. The identical scenario fused twice by the unconstrained DHD was a scenario based on the 3AT “322”.

E. DERIVATION OF LOGICAL SCENARIOS

The most likely traffic conflict leading to an accident in the considered ODD, which the hypothetical AEB

could face, was 3AT 6021 (44.6%, see Table 8). Thereby the observed EgoStartSpeeds were between 2.43 m/s and 17.32 m/s, with a mean of 10.71 m/s. Interestingly, in the VO data set for scenario 6021, the minimum observed EgoStartSpeed was 0.91 m/s, and thus 62.5% lower than in PD*. The variable EgoAccelerationStart demonstrates a comparable occurrence, as its minimum initial acceleration rises from 0.05 m/s² in VO to 0.16 m/s² in PD*. The observed AgentStartSpeeds were between 3.23 m/s and 16.23 m/s, with a mean of 10.64 m/s. The same phenomenon as for the ego occurs for the agent, showing an observed minimum AgentStartSpeed, which increases from 0.05 m/s in VO to 0.16 m/s in PD*, and an observed minimum AgentAccelerationStart, which increases from 0.01 m/s² to 0.11 m/s² in PD*. The accident severity was between no one injured and the agent being slightly injured, whereby the maximum property damage observed was 8.500€. The driver age of the egos was between 19 and 71 years, whereas that of the agent was between 19 and 79 years.

Table 8 also introduces exemplary logical scenarios for 3ATs 231 and 321, which display the minimum, mean, and maximum values of the considered PD* variable. As expected, the EgoSpeeds at start differ between the 3ATs: while the 3AT 6021 has a minimum speed at the start of 2.43 m/s (mean: 10.71 m/s), 3AT 231 has 9.72 m/s (mean: 10.01 m/s) and 3AT 321 12.24 m/s (mean: 14.72 m/s) due to probably wrongly assuming the right of way. For testing in a stochastic simulation, it is necessary to transfer real intersection geometries into an OpenDRIVE file [96]. Table 8 provides the location IDs (LocIDs) for OpenDRIVE file creation, that can be used to create the corresponding files from aerial images.

V. DISCUSSION

In the following section, we answer the research questions, relate the results to the literature, and discuss the limitations of the study. Finally, we provide an outlook for researchers and practitioners.

A. RESEARCH QUESTIONS

The overall RQ asked for a method to create a representative test scenario catalog suitable for testing in stochastic traffic simulations. We answered the overall RQ by fusing 74 accidents obtained from PD and 1,648 scenarios identified in the VO data following the FUSE4Rep process model. The generated representative test scenario database PD*, containing 74 scenarios described by over 350 variables, can be considered representative of the defined population derived from AEB's ODD. We want to emphasize that it would not have been possible to derive concrete scenarios from PD by analytically reconstructing accidents because of missing information. Even if the PD contains extensive information for reconstruction, the analytical reconstruction of any accident is very complex and expensive.

RQ1 asked for suitable MVs to fuse PD and VO data. Among the 32 test fusions conducted, AccidentType, Geometry, and BusStop were identified as the best combinations of MVs. Unfortunately, AccidentType did not belong to the common population of PD and VO, and attention should be paid to this in future fusions. An improvement could possibly already be achieved by a VO that overlaps more closely in time with the PD and/or a VO that is carried out over a longer period of time instead of just three months.

RQ2 asked which type of DHD performed better: constrained or unconstrained. In this study, unconstrained DHD performed the best and outperformed constrained DHD in 31 of the 32 test fusions.

RQ3 asked how good the final fusion results were overall: Almost all conducted comparisons of distribution similarity between VO and PD* showed that these are similar, which is the most important quality criterion for a valid data fusion. In particular, the distributions of the most important metric variables – Ego/AgentSpeedStart, Ego/AgentAccelerationStart, Ego/AgentAccelerationTanStart, and SSM_{min} – show a high degree of similarity.

Finally, RQ4 asked whether it was possible to derive logical scenarios applicable for testing in a stochastic traffic simulation. In fact, the logical scenarios derived from PD* do contain all necessary variables [3], [4], [7]: Ego/AgentEntryStart specifying the start position as the distance to the intersection entry, AccidentType and Ego/AgentManeuver describing the maneuvers of ego and agent, Ego/AgentSpeedStart describing the initial speed at the start, Ego/AgentAge as the corresponding driver's age, and Ego/AgentDrivingLicenseYear as the year in which the driver's license was acquired. The latter can be an indicator of the driving experience. It has also been shown that logical scenarios derived from PD* instead of VO have a narrower range of values, e.g. for the initial start speeds Ego/AgentStartSpeed, which allows the speed ranges to be narrowed down to higher start speeds in the stochastic simulation.

In addition, the fused data set PD* allows the derivation of the parameters in the way they are needed to generate logical scenarios for stochastic traffic simulations [3], [7]: be it min/max specifications for metric variables or, for example, normal distributions. The quality of the data fusion is suitable for the derivation of logical scenarios, because of the good distribution reproduction of the specific variables Z in PD*. Moreover, it is also possible to derive, for example, the Ego/AgentSpeed describing variables dependent on their current intersection phase (approaching, deceleration, crossing, and exit) [97]. Next, the three identified logical scenarios can be used for AEB evaluation in a stochastic simulation, as in [3]. The logical scenarios and the stochastic component of the simulation can then be used to test the AEB in normal driving, critical, and accident scenarios [3]. However, we emphasize that the concrete scenarios stored

in PD* do not necessarily have to be plausible, because the DHD minimizes both, the single and overall distances of the matching variables [26].

B. COMPARISON TO EXISTING RESEARCH

When comparing the results to the PD-VO fusion conducted in [26], it is remarkable that the MVs used are no longer time-related (e.g., Weekday, Hour, Minute), but conflict- (AccidentType) and location-related (Geometry, BusStop). Moreover, the quality of the data fusion itself is significantly better than that of [26], for example, with regard to the Z distributions in PD* and VO (a direct comparison of figures is not reasonable, because in [26] all specific variables Z were categorical). The reasons for this may be found in the use of the unconstrained DHD instead of the constrained one, in the more meaningful MVs, and in the longer VO (eight days vs. three months). Regarding the FUSE4Rep process model proposed in [7], no major drawbacks were identified in the application. Only matching the manually recorded incorrect behavior of road users (provided in VO [93], [94]) to the identified 3ATs (see [19]) did not work properly. In the future, location-based information (e.g., the zone of incorrect behavior) should be included in the matching process. Furthermore, while convenient, the 3AT scheme for describing a conflict situation, finds only pre-defined conflict patterns. This situation is exacerbated by the fact that the SSMs used (MTTC and PET) are unsuitable for describing any form of traffic conflict [80].

C. LIMITATIONS

Limitations arise owing to the conditional independence assumption (CIA) necessary for valid data fusion, according to [88], [98], and [99]. Thereby, CIA cannot be measured and must be assumed by an expert performing SM [88]. CIA assumes that the variables specific to each data set Y and Z are independent, given the common variables X . In other words, the joint distribution of Y and Z is unknown, whereas it is also difficult to estimate. However, the more common variables X exist, the more likely CIA can be assumed [100]. We assume that the CIA is given by 38 shared and 33 common variables.

We also emphasize that only the fourth validity level, the similarity of f_Z and $f_{Z_{X_{MV}}}$ between PD* and VO, according to [20], can be verified in real-world fusions. Therefore, subsequent statistical analyses of PD* must be performed carefully because of potential limitations. In addition, the overlap between the periods of PD (16 years) and VO (3 months) in the example shown is small; therefore, care should be taken to ensure a better match for future fusions.

Another limitation arises from the empirically determined thresholds, which were obtained by randomly splitting VO 100 times into two data sets. Further studies should investigate the optimal number of random splits to properly determine thresholds.

Finally, we want to point out that the trajectories contained in the VO data sets [93], [94] describe only the estimated center point of every road user, combined with an estimated standard geometry (length/width). Therefore, the calculated SSMs can overestimate or underestimate the corresponding conflicts owing to varying real-world geometries.

D. FOR RESEARCHERS

We recommend further research efforts to determine the incorrect behavior of road users in VO to gain additional MVs. Moreover, the 3AT determination, especially the SSM calculation in VO, can be improved using video analysis techniques, that allow the determination of exact road user geometries. In addition, scenarios/traffic conflicts can be identified in PD and VO by relying on rule-based techniques as well as supervised/unsupervised techniques [10].

Regarding SM, we recommend investigating imprecise imputation techniques, which are independent of CIA [101], and testing further validation techniques, such as relying on bias and variance estimates [102]. Finally, the construction of a fused data set using a symmetric SM [101] can lead to a larger database of test scenarios.




E. FOR PRACTITIONERS

We recommend that practitioners compare the demonstrated fusion results with real-world data, such as with comparable reconstructed accidents stored in the GIDAS database [90]. In addition, we recommend extending the shared variables between VO and PD by improving the VO data collection process. Also, the collection of metric variables instead of categorical variables can be expanded (e.g., precipitation amount instead of Rain: yes or no). Moreover, we recommend performing longer VO observation periods and conducting these continuously alongside PD collections so that the superordinate shared population between the VO and PD matches better. Furthermore, we recommend the additional use of weather-independent sensors. The drones used in [93], and [94] can fly only under good weather conditions, excluding bad weather-related traffic conflicts.

VI. CONCLUSION

Fusing concrete scenarios identified from different data sources helps close the gaps in data availability and differing information when creating representative test scenario databases and catalogs. As part of the FUSE4Rep process model, statistical matching can generate a new representative accident test scenario database, PD*, by asymmetrically fusing scenarios identified in police accident data (PD) with those identified in video-based traffic observation (VO) data. The necessary matching variables are AccidentType (describing the underlying traffic conflicts), BusStop (indicating the presence of bus stops), and Geometry (describing, for example, the number of nodes of intersections). The best matching algorithm is the unconstrained distance-hot-deck method. Consequently, the fused scenario database

TABLE 8. Logic scenarios derived from the fused database containing $n = 74$ scenarios following the 6LM when applicable (extract).

Source	6LM-Layer	Variable	3AT 6021 ($n = 33 44\%$)	3AT 231 ($n = 2 2.7\%$)	3AT 321 ($n = 6 8.1\%$)
					
			min/mean*/max		(*median for years)
X_S	1	LocID	ThaRei; ThaAlt; ThaCla; ReiMoh; CheBay; CheBie; ThaAnt; KohHoc; ThaCha; ReiAnt	ReiMoh; CheBie	ReiSch; ReiFra; ReiAnt
X_S	1	Geometry	3W; 4W	4W	4W
X_S	1	BusStop	No; Yes	No; Yes	No
X_S	1	TrafficRegulation	Yield-controlled	Yield-controlled	Yield-controlled
X_S	3	ConstructionSite	No	No	No
X_S	4	GeneralCause	None	None	None
X_S	4	CauseI	None; Distance	None; Distance	None
PD	4	TotalPropertyDamage [€]	30/2,172/8,500	300/900/1,500	500/3,083/4,000
VO	4	EgoEntryStart [m]	0/n.a./6	0	0
VO	4	EgoSpeedStart [m/s]	2.43/10.71/17.32	9.72/10.01/10.30	12.24/14.72/17.22
VO	4	EgoAccelerationStart [m/s ²]	0.16/1.09/3.21	0.53/0.83/1.13	0.33/0.64/0.97
VO	4	EgoAccelerationTanStart [m/s ²]	-3.21/-0.60/2.16	0.45/0.80/1.13	-0.72/0.03/0.96
X_S	4	EgoManeuver	Go Straight	Go Straight	Go Straight
PD	4	EgoAge [Years]	19/42/71	36/49/61	22/42/86
PD	4	EgoDrivingLicenseYear [Year]	1966/1994/2010	1993/n.a./2009	1963/2002/2012
PD	4	EgoInjury	None	None	None
PD	4	EgoPropertyDamage [€]	30/1,219/5,000	150/325/500	0/1,417/2,000
VO	4	AgentEntryStart [m]	0/n.a./6	0	0
VO	4	AgentSpeedStart [m/s]	3.23/10.64/16.23	5.89/6.32/6.75	1.84/2.54/3.80
VO	4	AgentAccelerationStart [m/s ²]	0.11/1.07/3.73	1.54/1.67/1.79	1.54/1.88/2.51
VO	4	AgentAccelerationTanStart [m/s ²]	-3.74/-0.63/2.54	-0.84/-0.75/-0.65	-2.51/-1.87/-1.54
X_S	4	AgentManeuver	Go Straight	Turn Right	Go Straight
PD	4	AgentAge [Years]	19/42/79	58/60/61	27/47/79
PD	4	AgentDrivingLicenseYear [Year]	1964/1998/2018	1991/n.a./1994	1964/1996/2008
PD	4	AgentInjury	None; slightly injured	None	None
PD	4	AgentPropertyDamage [€]	0/953/3,500	150/575/1,000	500/1,667/2,000
VO	4	SSM [s]	(M)TTC: 0/3.50/4.99	(M)TTC: 0/1.01/1.39	PET: 0/3.31/4.64
VO	4	TrafficVolume	Middle; High	High	Middle; High
X_S	5	RoadCondition	Dry	Dry	Dry
VO	5	Temperature [°C]	12.5/21.6/29.0	19.0/20.5/22.0	18.0/20.2/24.0
VO	5	RoadSurfaceTemperature [°C]	15.8/29.9/55.7	25.0/34.9/44.8	33.1/37.7/44.8
VO	5	Sunshine	No; Light; Strong	Light; Strong	No; Strong
VO	5	WindSpeed [m/s]	1.39/2.96/4.44	3.06/3.75/4.44	3.06/4.17/5.0
VO	5	Rain	No	No	No

PD* includes information on the road users involved in an accident (age, year of driving license acquisition), the conflict situation, and the corresponding starting speeds and accelerations. Thus, the most frequent conflicts leading to accidents and how they should be parameterized for ADS testing in stochastic simulations can be determined and transferred to a test scenario catalog.

The FUSE4Rep process model was proposed and applied in part to determine scenarios from VO data. As our results demonstrate for the first time, it is also possible to use the FUSE4Rep process model to fuse PD and VO scenario data sets into PD* and subsequently derive a test scenario catalog containing logical scenarios. These logical scenarios can then be used to assess, for example, autonomous emergency braking systems (AEBs), in stochastic traffic simulations. Future research should focus on improving the scenario identification process in the PD and VO

data sets and increasing the number of shared variables between the scenario data sets to be fused. Moreover, the process of statistical matching can be improved by investigating imprecise imputation techniques, performing symmetric statistical matching, and expanding validation techniques.

Overall, the collection of VO data (samples) should be continuous with the PD collection and a fixed VO collection system should be established - if this happens, we expect that our method will help to develop test scenario databases containing comprehensively described test scenarios that are representative of clearly defined operational design domains. In this way, we contribute to safe ADSs and, thus, to safer road traffic in the future.

APPENDIX

See Tables 8 and 9.

TABLE 9. Newly created data set PD*, described by 13 of 350+ selected variables. "9999" is coded, if unknown.

Nr	LocID	Geometry	BusStop	Accident Type	Ego Property Damage [€]	Ego Age [Years]	Ego Speed Start [m/s]	Ego Acceleration TanStart [m/s ²]	Agent Property Damage [€]	Agent Age [Years]	Agent Speed Start [m/s]	Agent Acceleration TanStart [m/s ²]	SSM _{min} [s]
1	ThaCha	3W	No	302	3,000	58	13.93	1.05	5,000	55	5.56	-0.80	3.57
2	ThaRei	3W	No	6021	1,500	32	15.42	-1.47	1,500	50	16.08	-1.66	4.62
3	ThaAnt	3W	No	322	1,500	52	16.07	0.05	1,500	33	5.77	-1.03	4.60
4	ThaAlt	3W	Yes	6021	1,000	71	10.98	-1.61	2,000	30	11.66	-1.71	1.29
5	ThaCla	3W	Yes	6021	750	55	6.66	1.11	750	43	3.67	2.54	3.34
6	ThaCla	3W	Yes	302	1,000	28	11.14	0.27	3,000	80	5.88	-1.99	2.70
7	ThaCla	3W	Yes	302	1,000	28	16.01	-0.01	3,000	80	3.90	-2.10	2.30
8	ReiSch	4W	No	303	500	54	12.94	0.54	1,500	35	3.74	-1.89	1.63
9	KohDor	4W	Yes	211	6,000	47	5.39	-0.52	6,000	46	12.63	0.55	0.83
10	ThaRei	3W	No	6021	1,000	51	10.23	-0.85	500	76	9.65	-0.08	2.41
11	ReiMoh	4W	Yes	6021	1,000	60	14.46	0.63	500	68	14.24	0.40	4.49
12	CheBay	3W	No	6021	100	25	9.18	0.66	100	47	9.28	0.90	4.68
13	ThaAlt	3W	Yes	322	1,500	59	13.21	-0.05	1,500	45	4.72	-1.08	3.20
14	CheBie	4W	No	6021	100	32	9.59	-0.58	50	31	9.96	-0.30	1.03
15	ThaCha	3W	No	302	4,000	26	8.57	-0.10	5,000	46	4.02	-0.51	4.57
16	ReiFra	4W	No	681	4,000	66	12.34	0.40	2,000	64	4.56	-0.30	3.94
17	ReiSch	4W	No	322	500	87	13.89	0.06	500	67	1.09	-1.22	2.20
18	ThaAlt	3W	Yes	6021	100	69	17.32	-0.15	300	34	16.23	-0.39	4.07
19	CheBay	3W	No	303	1,000	72	12.09	0.28	500	48	5.06	-1.84	2.77
20	ReiMoh	4W	Yes	322	500	31	14.13	-0.16	1,500	40	1.87	-1.35	2.87
21	ThaAlt	3W	Yes	322	2,000	42	16.07	0.05	6,000	28	5.77	-1.03	4.60
22	ThaAnt	3W	No	6021	500	39	10.20	-0.61	500	79	12.09	-0.97	2.61
23	CheBay	3W	No	6021	4,000	47	10.95	-1.21	1,000	59	12.19	-0.88	4.38
24	ThaCla	3W	Yes	6021	2,000	39	2.43	-1.00	500	24	4.66	-1.92	3.25
25	ReiMoh	4W	Yes	231	150	61	9.72	0.45	150	61	5.89	-0.84	0.63
26	ThaAnt	3W	No	6021	5,000	64	13.09	0.64	3,500	48	12.90	-0.12	4.24
27	CheBie	4W	No	231	500	36	10.30	1.13	1,000	58	6.75	-0.65	1.39
28	ThaAnt	3W	No	201	1,500	42	2.61	-1.90	1,500	46	7.36	-3.17	3.80
29	ThaRei	3W	No	6021	300	57	6.86	-1.89	600	23	6.65	-0.63	4.98
30	ThaAnt	3W	No	6021	1,500	35	11.60	-0.29	1,000	41	11.67	-0.08	4.22
31	CheBie	4W	No	6021	1,500	21	3.14	2.16	1,000	45	5.61	-3.74	3.78
32	KohHoc	3W	Yes	6021	2,000	30	6.95	-3.21	2,000	20	9.09	-3.44	4.82
33	ThaRei	3W	No	322	0	23	16.29	-0.90	0	26	6.24	-1.47	2.07
34	CheBay	3W	No	322	2,000	79	13.66	-0.57	500	34	5.76	-1.68	4.80
35	CheBie	4W	No	322	3,000	28	11.39	0.40	3,500	44	3.70	-1.29	4.27
36	ReiMoh	4W	Yes	6021	500	29	12.54	-0.30	500	49	13.07	-0.31	4.32
37	ThaCha	3W	No	6021	500	56	11.48	-1.95	1,500	69	11.49	-1.37	3.97
38	KohDor	4W	Yes	211	500	35	6.86	-0.78	2,500	22	11.57	0.53	0.20
39	ReiMoh	4W	Yes	6021	50	46	11.50	-0.36	250	48	5.86	-2.11	0.54
40	ReiSch	4W	No	321	1,500	86	12.24	-0.30	2,000	29	2.51	-1.78	3.60
41	ReiMoh	4W	Yes	211	500	20	5.64	-0.63	1,500	25	13.22	-0.23	3.24
42	ThaRei	3W	No	201	1,500	79	3.53	-0.54	1,500	24	6.37	-1.40	2.04
43	ReiFra	4W	No	211	5,000	50	14.35	0.02	5,000	49	8.35	-0.47	3.00
44	ThaAlt	3W	Yes	6021	900	20	8.32	-0.37	100	36	9.09	-0.13	3.76
45	ReiAnt	4W	No	303	2,000	26	11.57	0.41	3,000	27	5.42	-1.50	2.47
46	ThaRei	3W	No	6021	500	31	10.38	-0.09	500	42	10.51	-0.37	3.02
47	ReiAnt	4W	No	302	1,000	22	1.50	1.82	1,000	53	13.75	-0.10	3.94
48	CheBie	4W	No	6021	2,000	56	10.82	-0.09	2,000	44	11.02	-0.27	3.86
49	CheBie	4W	No	6021	2,500	19	10.99	0.97	1,000	19	10.65	0.59	3.40
50	KohHoc	3W	Yes	6021	200	19	12.53	-0.65	500	52	13.31	-0.24	3.49
51	ReiMoh	4W	Yes	322	400	31	17.38	-0.21	700	46	3.86	-2.18	2.27
52	ThaAnt	3W	No	6021	2,000	68	12.21	-2.52	2,000	31	12.47	-2.14	3.63
53	ThaAlt	3W	Yes	302	2,500	74	11.80	1.04	2,500	43	3.37	-1.40	4.74
54	ThaAnt	3W	No	6021	500	29	13.38	-0.89	1,000	36	13.35	-0.57	3.66
55	ThaRei	3W	No	6021	30	25	13.86	0.39	0	37	13.35	1.18	1.94
56	ReiSch	4W	No	302	2,000	37	13.68	0.06	2,000	61	1.24	-1.19	3.54
57	ReiFra	4W	No	321	2,000	80	14.60	-0.72	2,000	79	3.80	-1.54	2.27
58	ReiAnt	4W	No	321	1,500	22	17.22	0.39	1,500	24	2.51	-1.78	4.37
59	CheBie	4W	No	322	500	72	14.85	-0.40	500	20	3.83	-1.20	4.67
60	ThaAnt	3W	No	6021	500	28	14.39	-1.62	1,000	21	13.06	-1.60	2.69
61	ReiMoh	4W	Yes	6021	500	51	15.67	-0.11	2,500	25	16.19	-0.35	3.04
62	ThaRei	3W	No	6021	3,000	66	7.27	-1.27	1,000	40	6.87	-0.36	4.99
63	ReiFra	4W	No	321	0	9999	14.89	0.96	500	64	2.07	-1.87	4.64
64	ThaFra	3W	No	201	2,000	71	3.63	-1.65	2,000	31	4.93	-2.73	2.95
65	ThaRei	3W	No	302	1,000	20	6.49	-2.52	1,000	25	10.43	0.48	4.70
66	ReiFra	4W	No	321	2,000	35	14.12	-0.35	2,000	27	1.84	-2.51	2.74

TABLE 9. (Continued.) Newly created data set PD*, described by 13 of 350+ selected variables. "9999" is coded, if unknown.

67	ThaAnt	3W	No	681	7,000	41	8.19	0.75	5,000	67	5.03	-1.38	2.74
68	ThaRei	3W	No	6021	1,000	42	14.96	0.63	500	41	11.89	-0.08	2.38
69	ReiFra	4W	No	321	1,500	30	15.25	0.18	2,000	60	2.51	-1.78	2.24
70	ThaRei	3W	No	201	1,500	32	1.89	-1.09	1,500	23	6.32	-2.75	2.69
71	ReiAnt	4W	No	6021	3,000	20	3.19	-1.75	500	23	3.23	0.11	3.74
72	KohDor	4W	Yes	211	5,000	58	4.53	-0.82	6,000	40	7.83	-1.66	1.74
73	ReiMoh	4W	Yes	211	2,000	46	15.44	-0.17	2,000	21	7.60	-1.60	3.74
74	ThaRei	3W	No	6021	200	40	11.00	-2.14	800	56	10.17	-0.79	4.64

ACKNOWLEDGMENT

The authors thank Qilin Tang and Kevin Gowinkowski for their preliminary work, support, and fruitful discussions. They also thank Lena Otto for her extensive feedback.

REFERENCES

- [1] A. Hohm, "The second wave of automated driving," *ATZ Worldwide*, vol. 124, no. 5, p. 64, May 2022.
- [2] L. T. Bergmann, "Ethical issues in automated driving—Opportunities, dangers, and obligations," in *User Experience Design in the Era of Automated Driving*. Berlin, Germany: Springer, 2022, pp. 99–121.
- [3] C. Siebke, M. Bäuml, K. Blenz, M. Lehmann, M. Ringhand, M. Mai, and G. Prokop, "Predicting the impact on road safety of an intersection AEB at urban intersections. Using a novel virtual test field for the assessment of conflict prevention between cyclists/pedelecs and cars," *Transp. Res. Interdiscipl. Perspect.*, vol. 17, Jan. 2023, Art. no. 100728.
- [4] M. Witt, L. Wang, F. Fahrenkrog, K. Kompaß, and G. Prokop, "Cognitive driver behavior modeling: Influence of personality and driver characteristics on driver behavior," in *Proc. AHFE Int. Conf. Hum. Factors Transp.*, vol. 786, Orlando, FL, USA, Jul. 2018, pp. 751–763, doi: 10.1007/978-3-319-93885-1_69.
- [5] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick, and F. Diermeyer, "Survey on scenario-based safety assessment of automated vehicles," *IEEE Access*, vol. 8, pp. 87456–87477, 2020.
- [6] K. Czarniecki, "Operational design domain for automated driving systems," Waterloo Intell. Syst. Eng. (WISE) Lab., Univ. Waterloo, ONT, Canada, Tech. Rep., 2018, doi: 10.13140/RG.2.2.18037.88803.
- [7] M. Bäuml and G. Prokop, "Validating automated driving systems by using scenario-based testing: The Fuse4Rep process model for scenario generation as part of the Dresden method," *Zeitschrift Für Verkehrssicherheit*, vol. 68, no. 3, pp. 226–230, 2022.
- [8] M. Scholtes, L. Westhofen, L. R. Turner, K. Lotto, M. Schuldes, H. Weber, N. Wagener, C. Neurohr, M. H. Bollmann, F. Körte, J. Hiller, M. Hoss, J. Bock, and L. Eckstein, "6-layer model for a structured description and categorization of urban traffic and environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021.
- [9] T. Menzel, G. Bagschik, and M. Maurer, "Scenarios for development, test and validation of automated vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Cangshu, China, Jun. 2018, pp. 1821–1827.
- [10] J. Cai, W. Deng, H. Guang, Y. Wang, J. Li, and J. Ding, "A survey on data-driven scenario generation for automated vehicle testing," *Machines*, vol. 10, no. 11, p. 1101, Nov. 2022.
- [11] H. Ren, H. Gao, H. Chen, and G. Liu, "A survey of autonomous driving scenarios and scenario databases," in *Proc. 9th Int. Conf. Dependable Syst. Their Appl. (DSA)*, Wulumuqi, China, Aug. 2022, pp. 754–762.
- [12] D. Nalic, T. Mihalj, A. Eichberger, M. Bäuml, and M. Lehmann, "Scenario based testing of automated driving systems: A literature survey," in *Proc. FISITA World Congr. Tech. Programme*, Sep. 2021, pp. 1–10.
- [13] M. Lehmann, M. Bäuml, G. Prokop, and D. Hamelow, "Use of a criticality metric for assessment of critical traffic situations as part of SePIA," in *Proc. Int. Stuttgarter Symp.*, Stuttgart, Germany, 2019, pp. 1154–1167.
- [14] ITF. (2022). *Road Safety Annual Report 2022*. OECD, Paris, France. [Online]. Available: <https://www.itf-oecd.org/sites/default/files/docs/irtad-road-safety-annual-report-2022.pdf>
- [15] fka GmbH. *levelXData—Real-World Scenario Data*. fka GmbH, Aachen, Germany. Accessed: Nov. 1, 2023. [Online]. Available: <https://levelxdata.com/>
- [16] M. Bäuml and M. Lehmann. *ListDB: Representative Datasets of Video-Based Traffic Observations*. Dresden, Germany. Accessed: Jul. 17, 2023. [Online]. Available: <https://opara.zih.tu-dresden.de/xmlui/handle/123456789/5882>
- [17] J. M. Hankey, M. A. Perez, and J. A. McClafferty. (2016). *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets*. Virginia Tech Transportation Institute (VTTI), Blacksburg, Virginia. [Online]. Available: <http://hdl.handle.net/10919/70850>
- [18] Statistisches Bundesamt Deutschland. *GENESIS-Online: Die Datenbank des Statistischen Bundesamtes: Verkehrsunfälle*. Wiesbaden, Germany. Accessed: Aug. 9, 2023. [Online]. Available: <https://www-genesis.destatis.de/genesis/online>
- [19] M. Bäuml, M. Lehmann, and G. Prokop, "Generating representative test scenarios: The fuse for representativity (Fuse4Rep) process model for collecting and analysing traffic observation data," in *Proc. 27th ESV Conf.*, Yokohama, Japan, 2023, pp. 1–26.
- [20] S. Rässler, *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York, NY, USA: Springer, 2002.
- [21] A. J. Braverman, "Data fusion," in *Encyclopedia of Quantitative Risk Analysis and Assessment*, E. L. Melnick and B. S. Everitt, Eds. Hoboken, NJ, USA: Wiley, 2008.
- [22] S. Feng, Y. Feng, H. Sun, S. Bao, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles, Part II: Case studies," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 9, pp. 5635–5647, Sep. 2021.
- [23] X. Zhang, J. Tao, K. Tan, M. Törngren, J. M. G. Sánchez, M. R. Ramli, X. Tao, M. Gyllenhammar, F. Wotawa, N. Mohan, M. Nica, and H. Felbinger, "Finding critical scenarios for automated driving systems: A systematic mapping study," *IEEE Trans. Softw. Eng.*, vol. 49, no. 3, pp. 991–1026, Mar. 2023.
- [24] W. Ko, S. Park, S. Park, H. Jeong, and I. Yun, "Development of freeway-based test scenarios for applying new car assessment program to automated vehicles," *PLoS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0271532.
- [25] S. Krause and F. Busch, "Identification of hazards in urban driving situations using accident categorizations," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Las Vegas, NV, USA, Oct. 2020, pp. 156–161.
- [26] M. Bäuml, L. Dziuba-Kaiser, Z. Yin, M. Lehmann, and G. Prokop, "Use information you have never observed together: Data fusion as a major step towards realistic test scenarios," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Rhodes, Greece, Sep. 2020, pp. 1–8, doi: 10.1109/ITSC45102.2020.9294649.
- [27] S. Krause, "Development of a database for the description and analysis of use cases for automated driving," in *Proc. 98th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2019. [Online]. Available: <https://mediatum.ub.tum.de/1476889>
- [28] A. Keler, P. Malcolm, G. Grigoropoulos, S. A. Hosseini, H. Kath, F. Busch, and K. Bogenberger, "Data-driven scenario specification for AV–VRU interactions at urban roundabouts," *Sustainability*, vol. 13, no. 15, p. 8281, Jul. 2021.
- [29] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Maui, HI, USA, Nov. 2018, pp. 2118–2125.

- [30] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The inD dataset: A drone dataset of naturalistic road user trajectories at German intersections," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Las Vegas, NV, USA, Oct. 2020, pp. 1929–1934.
- [31] T. Winkle, C. Erbsmehl, and K. Bengler, "Area-wide real-world test scenarios of poor visibility for safe development of automated vehicles," *Eur. Transp. Res. Rev.*, vol. 10, no. 2, Jun. 2018, Art. no. 32.
- [32] Y. Akagi, R. Kato, S. Kitajima, J. Antona-Makoshi, and N. Uchida, "A risk-index based sampling method to generate scenarios for the evaluation of automated driving vehicle safety," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, New Zealand, Oct. 2019, pp. 667–672.
- [33] T. Braun, L. Ries, F. Körtke, L. Turner, S. Otten, and E. Sax, "Collection of requirements and model-based approach for scenario description," in *Proc. 7th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2021, pp. 634–645.
- [34] C. Chang, D. Cao, L. Chen, K. Su, K. Su, Y. Su, F.-Y. Wang, J. Wang, P. Wang, J. Wei, G. Wu, X. Wu, H. Xu, N. Zheng, and L. Li, "MetaScenario: A framework for driving scenario data description, storage and indexing," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 2, pp. 1156–1175, Feb. 2023.
- [35] P. Elspas, J. Lindner, M. Brosowsky, J. Bach, and E. Sax, "Towards a scenario database from recorded driving data with regular expressions for scenario detection," in *Proc. 8th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2022, pp. 400–409.
- [36] S. Feng, Y. Feng, C. Yu, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles, Part I: Methodology," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1573–1582, Mar. 2021.
- [37] A. Gambi, V. Nguyen, J. Ahmed, and G. Fraser, "Generating critical driving scenarios from accident sketches," in *Proc. IEEE Int. Conf. Artif. Intell. Test. (AITest)*, Newark, CA, USA, Aug. 2022, pp. 95–102.
- [38] Q. Goss, Y. AlRashidi, and M. I. Akbas, "Generation of modular and measurable validation scenarios for autonomous vehicles using accident data," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Nagoya, Japan, Jul. 2021, pp. 251–257.
- [39] I. Han, "Scenario establishment and characteristic analysis of intersection collision accidents for advanced driver assistance systems," *Traffic Injury Prevention*, vol. 21, no. 6, pp. 354–358, Aug. 2020.
- [40] D. Kaufmann, L. Klampfl, F. Klück, M. Zimmermann, and J. Tao, "Critical and challenging scenario generation based on automatic action behavior sequence optimization: 2021 IEEE autonomous driving AI test challenge group 108," in *Proc. IEEE Int. Conf. Artif. Intell. Test. (AITest)*, Oxford, U.K., Aug. 2021, pp. 118–127.
- [41] C. Knies and F. Diermeyer, "Data-driven test scenario generation for cooperative maneuver planning on highways," *Appl. Sci.*, vol. 10, no. 22, p. 8154, Nov. 2020.
- [42] W. Ko, S. Park, J. Yun, S. Park, and I. Yun, "Development of a framework for generating driving safety assessment scenarios for automated vehicles," *Sensors*, vol. 22, no. 16, p. 6031, Aug. 2022.
- [43] F. Montanari, C. Stadler, J. Sichermann, R. German, and A. Djanatliev, "Maneuver-based resimulation of driving scenarios based on real driving data," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Nagoya, Japan, Jul. 2021, pp. 1124–1131.
- [44] H. Nakamura, H. Muslim, R. Kato, S. Préfontaine-Watanabe, H. Nakamura, H. Kaneko, H. Imanaga, J. Antona-Makoshi, S. Kitajima, N. Uchida, E. Kitahara, K. Ozawa, and S. Taniguchi, "Defining reasonably foreseeable parameter ranges using real-world traffic data for scenario-based safety assessment of automated vehicles," *IEEE Access*, vol. 10, pp. 37743–37760, 2022.
- [45] D. Nalic, A. Eichberger, G. Hanzl, M. Fellendorf, and B. Rogic, "Development of a co-simulation framework for systematic generation of scenarios for testing and validation of automated driving systems," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, New Zealand, Oct. 2019, pp. 1895–1901.
- [46] S.-W. Park, K. Patil, W. Wilson, M. Corless, G. Choi, and P. Adam, "Creating driving scenarios from recorded vehicle data for validating lane centering system in highway traffic," *Soc. Automot. Eng. (SAE)*, Warrendale, PA, USA, Tech. Rep., 2020, Art. no. 2020-01-0718. [Online]. Available: <https://www.sae.org/publications/technical-papers>
- [47] R. Pfeffer, J. He, and S. Eric, "Development and implementation of a concept for the meta description of highway driving scenarios with focus on interactions of road users," in *Proc. 6th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2020, pp. 440–447.
- [48] C. Reichenbacher, M. Rasch, Z. Kayatas, F. Wirthmüller, J. Hipp, T. Dang, and O. Bringmann, "Identifying scenarios in field data to enable validation of highly automated driving systems," in *Proc. 8th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2022, pp. 134–142.
- [49] D. Shen, Q. Yi, L. Li, R. Tian, S. Chien, Y. Chen, and R. Sherony, "Test scenarios development and data collection methods for the evaluation of vehicle road departure prevention systems," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 3, pp. 337–352, Sep. 2019.
- [50] A. Tenbrock, A. König, T. Keutgens, and H. Weber, "The ConScenD dataset: Concrete scenarios from the highD dataset according to ALKS regulation UNECE R157 in OpenX," in *Proc. IEEE Intell. Vehicles Symp. Workshops (IV Workshops)*, Nagoya, Japan, Jul. 2021, pp. 174–181.
- [51] P. Themann, A. Pütz, A. Zlocki, and L. Eckstein, "Database of relevant traffic scenarios as a tool for the development and validation of automated driving," in *Proc. 13th Int. Symp. Adv. Vehicle Control*, Munich, Germany, 2016, pp. 489–496.
- [52] L. Tötzel, M. Zipfl, D. Bogdoll, M. R. Zofka, and J. M. Zöllner, "Reliving the dataset: Combining the visualization of road users' interactions with scenario reconstruction in virtual reality," in *Proc. 6th Int. Conf. Intell. Transp. Eng.*, 2022, pp. 436–454.
- [53] L. Xu, C. Zhang, Y. Liu, L. Wang, and L. Li, "Worst perception scenario search for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Las Vegas, NV, USA, Oct. 2020, pp. 1702–1707.
- [54] D. Zhao, Y. Guo, and Y. J. Jia, "TrafficNet: An open naturalistic driving scenario library," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 1–8, doi: [10.1109/ITSC.2017.8317860](https://doi.org/10.1109/ITSC.2017.8317860).
- [55] A. Zlocki, A. König, J. Bock, H. Weber, H. Muslim, H. Nakamura, S. Watanabe, J. Antona-Makoshi, and S. Taniguchi, "Logical scenarios parameterization for automated vehicle safety assessment: Comparison of deceleration and cut-in scenarios from Japanese and German highways," *IEEE Access*, vol. 10, pp. 26817–26829, 2022.
- [56] C. Cheng, "Design and implementation of test scenario library data system for autonomous vehicles," Tech. Rep., 2020, Art. no. 2022-01-5219.
- [57] N. Kolb, F. Hauer, M. Golagha, and A. Pretschner, "Data-driven assessment of parameterized scenarios for autonomous vehicles," in *Computer Safety, Reliability, and Security (Lecture Notes in Computer Science)*, vol. 13414, M. Trapp, F. Saglietti, M. Spislander, and F. Bitsch, Eds. Cham, Switzerland: Springer, 2022, pp. 350–364.
- [58] J. Langner, H. Grolig, S. Otten, M. Holzapfel, and E. Sax, "Logical scenario derivation by clustering dynamic-length-segments extracted from real-world-driving-data," in *Proc. 5th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, Heraklion, Crete, 2019, pp. 458–467.
- [59] N. E. Chelbi, D. Gingras, and C. Sauvageau, "Worst-case scenarios identification approach for the evaluation of advanced driver assistance systems in intelligent/autonomous vehicles under multiple conditions," *J. Intell. Transp. Syst.*, vol. 26, no. 3, pp. 284–310, May 2022.
- [60] P. Elspas, Y. Klose, S. Isele, J. Bach, and E. Sax, "Time series segmentation for driving scenario detection with fully convolutional networks," in *Proc. 7th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2021, pp. 56–64.
- [61] P. Elspas, J. Langner, M. Aydinbas, J. Bach, and E. Sax, "Leveraging regular expressions for flexible scenario detection in recorded driving data," in *Proc. IEEE Int. Symp. Syst. Eng. (ISSE)*, Oct. 2020, pp. 1–8, doi: [10.1109/ISSE49799.2020.9272025](https://doi.org/10.1109/ISSE49799.2020.9272025).
- [62] A. Harder, J. Ranjit, and M. Behl, "Scenario2Vector: Scenario description language based embeddings for traffic situations," in *Proc. ACM/IEEE 12th Int. Conf. Cyber-Phys. Syst.*, M. Maggio, J. Weimer, M. Al Faruque, and M. Oishi, Eds. New York, NY, USA, Jun. 2021, pp. 167–176.
- [63] F. Montanari, H. Ren, and A. Djanatliev, "Scenario detection in unlabeled real driving data with a rule-based state machine supported by a recurrent neural network," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–5, doi: [10.1109/VTC2021-Spring51267.2021.9449032](https://doi.org/10.1109/VTC2021-Spring51267.2021.9449032).
- [64] M. Bäumler and G. Prokop, "Generating ADS test scenarios from police accident data: How to predict the type of road traffic accident accurately," Technische Universität Dresden, Chair Automobile Eng., Dresden, Germany, Tech. Rep., 2022, doi: [10.2139/ssrn.4295798](https://doi.org/10.2139/ssrn.4295798).
- [65] Gesamtverband der Deutschen Versicherungswirtschaft. (2016). *Unfalltypen-Katalog: Leitfaden zur Bestimmung des Unfalltyps*. Unfallforschung der Versicherer, Berlin. [Online]. Available: <https://www.udv.de/resource/blob/80022/89b4d80028aac8cab649d3a3c6157a0/unfalltypenkatalog-data.pdf>

- [66] C. Siebke, M. Mai, and G. Prokop, "What do traffic simulations have to provide for virtual road safety assessment? Human error modeling in traffic simulations," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 2, pp. 1419–1436, Nov. 2023.
- [67] M. D'Orazio, *Statistical Matching: Theory and Practice*. Chichester, U.K.: Wiley, 2006.
- [68] J. C. Gower, "A general coefficient of similarity and some of its properties," *Biometrics*, vol. 27, no. 4, pp. 857–871, Dec. 1971.
- [69] G. Siedel. (2021). *Evaluation von Machine-Learning-Modellen und Konzeptionierung eines Modell-Ensembles für die Vorhersage von Unfalldaten*. Technische Univ. Dresden, Dresden. [Online]. Available: <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-739723>
- [70] H. Chen. (2021). *Verbesserung und Evaluation eines Modell-Ensembles für die Vorhersage von Unfalldaten anhand synthetischer Daten*. Technische Univ. Dresden, Dresden. [Online]. Available: <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-764956>
- [71] L. Zheng and T. Sayed, "A novel approach for real time crash prediction at signalized intersections," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102683.
- [72] S. Häder, "GESIS survey guidelines: Stichproben in der Theorie v1.1.," GESIS-Leibniz-Inst. für Sozialwissenschaften, Mannheim, Germany, Tech. Rep., 2015, doi: [10.15465/gesis-sg_014](https://doi.org/10.15465/gesis-sg_014).
- [73] A. P. Tarko, *Measuring Road Safety With Surrogate Events*. Amsterdam, The Netherlands: Elsevier, 2019.
- [74] F. Orsini, G. Gecchele, R. Rossi, and M. Gastaldi, "A conflict-based approach for real-time road safety analysis: Comparative evaluation with crash-based models," *Accident Anal. Prevention*, vol. 161, Oct. 2021, Art. no. 106382.
- [75] E. Polders and T. Brijs, "How to analyse accident causation? A handbook with focus on vulnerable road users (deliverable 6.3)," Horizon 2020 EC Project, Transp. Res. Inst. (IMOB), InDeV, Hasselt Univ., Hasselt, Belgium, Tech. Rep., 2018. [Online]. Available: <https://documentserver.uhasselt.be/bitstream/1942/27208/2/Handbook-InDeV.pdf>
- [76] M. Bäuml, M. Lehmann, and S. Arndt, *ListDBOnto: ListDB—Leverage Information on Street Traffic*. Accessed: Aug. 17, 2023. [Online]. Available: <https://w3id.org/listdb>
- [77] RCE Systems SRO. *DataFromSky TrafficSurvey*. Accessed: Aug. 18, 2023. [Online]. Available: <https://ai.datafromsky.com/picker>
- [78] Platomo GmbH. *OpenTrafficCam: Fully Integrated Open Source Workflow for Video-Based Recording and Automated Analysis of Road Traffic*. Accessed: Aug. 18, 2023. [Online]. Available: <https://opentrafficcam.org/>
- [79] M. A. Khan, W. Ectors, T. Bellemans, D. Janssens, and G. Wets, "Unmanned aerial vehicle-based traffic analysis: Methodological framework for automated multivehicle trajectory extraction," *Transp. Res. Record: J. Transp. Res. Board*, vol. 2626, no. 1, pp. 25–33, Jan. 2017.
- [80] 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 Res.*, vol. 41, no. 4, pp. 153–163, Dec. 2017.
- [81] W. Ding, C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao, "A survey on safety-critical driving scenario generation—A methodological perspective," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 6971–6988, Mar. 2023.
- [82] K. Ozbay, H. Yang, B. Bartin, and S. Mudigonda, "Derivation and validation of new simulation-based surrogate safety measure," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2083, no. 1, pp. 105–113, Jan. 2008.
- [83] J. C. Hayward. (1972). *Near Miss Determination Through Use of a Scale of Danger*. [Online]. Available: <https://onlinepubs.trb.org/onlinepubs/hrr/1972/384/384-004.pdf>
- [84] B. L. Allen, B. T. Shin, and P. J. Cooper. (1978). *Analysis of Traffic Conflicts and Collisions*. [Online]. Available: <https://trid.trb.org/view/85806>
- [85] A. Schubert, C. Erbsmehl, and L. Hannawald, "Standardized pre-crash scenarios in digital format on the basis of the VUFO simulation," in *Proc. 5th Int. Conf. ESAR*, Hannover, Germany, 2013, pp. 1–7. [Online]. Available: <https://bast.opus.hbz-nrw.de/opus45-bast/frontdoor/index/index/docId/630>
- [86] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Berlin, Germany: Springer, 2001.
- [87] M. D. Orazio. (2013). *Statistical Matching: Methodological Issues and Practice With R-StatMatch: Introduction to Statistical Matching*. Italian National Institute of Statistics—Istat, Vitoria-Gasteiz. [Online]. Available: https://en.eustat.eus/productosServicios/datos/Seminario_55_Slides.pdf
- [88] A. Leulescu and M. Agafitei, "Statistical matching: A model based approach for data integration," Publications Office Eur. Union, Luxembourg, U.K., Tech. Rep., 2013, doi: [10.2785/44822](https://doi.org/10.2785/44822).
- [89] D. Kornbrot, "Point biserial correlation," in *Wiley StatsRef: Statistics Reference Online*. Hoboken, NJ, USA: Wiley, 2014, doi: [10.1002/9781118445112.stat06227](https://doi.org/10.1002/9781118445112.stat06227).
- [90] Verkehrsunfallforschung an der TU Dresden GmbH. *GIDAS: German In-Depth Accident Study*. Dresden, Germany. Accessed: Aug. 22, 2023. [Online]. Available: <https://www.gidas.org/start-en.html>
- [91] V. W. Berger and Y. Zhou, "Kolmogorov–Smirnov test: Overview," in *Wiley StatsRef: Statistics Reference Online*. Hoboken, NJ, USA: Wiley, 2014, doi: [10.1002/9781118445112.stat06558](https://doi.org/10.1002/9781118445112.stat06558).
- [92] P. Wessel. (2013). *Critical Values for the Two-Sample Kolmogorov–Smirnov Test (2-Sided)*. University Hawaii at Manoa (SOEST). Accessed: Aug. 22, 2023. [Online]. Available: <https://sparky.rice.edu/astr360/kstest.pdf>
- [93] M. Bäuml and M. Lehmann, "ListDB RepOne: 3 months (Jun'22–Aug'22) of drone videos and trajectories at a 3-way intersection," ListDB: Representative datasets of video-based traffic observations," Chair Automobile Eng., Technische Univ. Dresden, Dresden, Germany, Tech. Rep., 2023, doi: [10.25532/OPARA-229](https://doi.org/10.25532/OPARA-229).
- [94] M. Bäuml and M. Lehmann, "ListDB RepTwo: 3 months (Jun'22–Aug'22) of drone videos and trajectories at a 4-way intersection," ListDB: Representative datasets of video-based traffic observations," Chair Automobile Eng., Technische Univ. Dresden, Dresden, Germany, Tech. Rep., 2023, doi: [10.25532/OPARA-230](https://doi.org/10.25532/OPARA-230).
- [95] GeoSN. *Geoportal Sachsenatlas*. Accessed: Aug. 16, 2023. [Online]. Available: www.geoportal.sachsen.de
- [96] ASAM. *ASAM OpenDrive*. Accessed: Jan. 26, 2023. [Online]. Available: <https://www.asam.net/standards/detail/opendrive/>
- [97] M. Plavsic. (2010). *Analysis and Modeling of Driver Behavior for Assistance Systems at Road Intersections*. Fakultät für Maschinenwesen, Technische Universität, München, München, Germany. [Online]. Available: <https://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20101004-973138-1-2>
- [98] C. A. Sims, "Rejoinder," *Ann. Econ. Social Meas.*, vol. 1, no. 3, pp. 355–357, 1972.
- [99] T. de Waal, "Statistical matching: Experimental results and future research questions," The Hague, Statist. Netherlands, Tech. Rep. 19, 2015.
- [100] I. Tsamardinos, S. Traintafillou, and V. Lagani, "Towards integrative causal analysis of herogeneous data sets and studies," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 1097–1157, 2012.
- [101] E.-M. Endres. (2019). *Statistical Matching Meets Probabilistic Graphical Models*. Fakultät für Mathematik, Informatik und Statistik, Ludwig-Maximilians-Universität München, München, Germany. [Online]. Available: <https://edoc.ub.uni-muenchen.de/24400/>
- [102] F. Goudie, A. van Delden, and T. de Waal, "Quality measure to evaluate statistical matching methods," in *Proc. Conf. New Techn. Technol. Statist.*, Brussels, Belgium, 2023, pp. 1–5.



MAXIMILIAN BÄUMLER received the Diploma Engineering degree from TU Dresden, Germany, in 2017, and the M.B.A. degree from Collège des Ingénieurs, in 2018. Since 2018 he is researching at the Chair of Automobile Engineering, TU Dresden, in the field of scenario generation for assessing automated driving systems. His current research interests include the fusion of police accidents and video-based traffic observation data to generate representative test scenarios.



GÜNTHER PROKOP received the Dr.-Ing. degree from the Technical University of Munich, Germany, in 1998. After more than ten years of work in German automotive industry, he became a Professor with the Chair of Automobile Engineering, TU Dresden, in 2010. From 2018 to 2022, he was the Dean of the "Friedrich List" Faculty of Transport and Traffic Sciences, TU Dresden.

...