It is a capital mistake to theorize before one has data.
– Sir Arthur Conan Doyle

Following the reasoning of Sir Arthur Conan Doyle, instead of assuming that automated driving systems (ADSs) drive more safely than human drivers, their safety performance should be assessed using real-world data. For that reason, this thesis encompasses eight publications, P1–P8 [1–8], on how to develop a method for deriving test scenario catalogs from real-world data in order to prospectively assess ADSs.

1.1. Motivation and background

The second wave of automated driving is taking off [9]. Mercedes-Benz and BMW already offer traffic jam pilots classified as Level 3 by SAE International[1] [10, 11], which, as such, autonomously control vehicles in good weather conditions up to 60 km/h in highway traffic jams [11]. Level 3, also called "conditional driving automation," requires the driver to take over in an appropriate time frame when the automated driving system (ADS) reaches its limits [10]. To be approved, ADSs need to prove that they drive more safely than attentive human drivers and to show a "positive risk balance" [12]. One efficient way to prove a positive risk balance is to compare the estimated safety performance of ADSs with that of human drivers in a set of test scenarios [12, 13], also called a "test scenario catalog" [1].

The concept of using test scenarios instead of conventional test drives for assessing the safety performance of ADSs relies on the "scenario-based testing approach" [1,15]. In scenario-based testing, interesting parts of road traffic are transferred into test scenarios. Consequently, uninteresting parts of road traffic, including monotonous driving at a constant speed on a motorway with no change in scenery, are excluded from the test. Overall, a test scenario is "*a kind of flip book representing the temporal sequence of scenes with different actions (e.g. [,] lane change) and events (e.g.[,] collision)*" [4, p.226]. Figure 1.1 illustrates an example of a car–bicycle scenario observed in a video-based traffic observation (VO).

For that reason, a (prospective) assessment of the safety performance of an ADS can be valid only if all relevant scenarios are included in the assessment [4]. Against that background, the method developed in this thesis addresses the problem of creating test scenario catalogs that are "representative" of all scenarios occurring in the operational design domain (ODD) of the ADS to be tested in order to ensure a comprehensive assessment of the safety performance of the ADS. To date, no road traffic data source exists that can be used solely to derive a "representative" test scenario catalog that reliably covers all scenarios occurring in an ODD [7, 14].

[1] SAE International was formerly called "Society of Automotive Engineers."

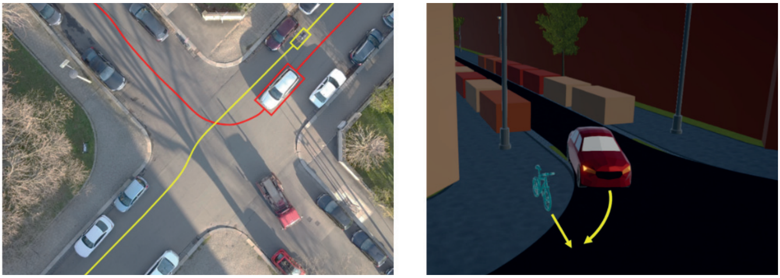

Figure 1.1.: Example of a car–bicycle scenario, observed in VO (left, see P4) and applied in a stochastic traffic simulation (right, see P8). Figure taken in parts from Bäumler and Prokop [4].

The reasons for that circumstance can be found, among others, in the limited availability of real-world data [2, 7] and divergent information among road traffic data sources [2, 7]. Accordingly, the motivation of this thesis is to solve the challenges of such limited data availability and divergent information using the fusion of information (i.e., scenarios) identified from different road traffic data sources, thereby enabling the generation of representative test scenario catalogs.

To better understand those challenges, it is important to understand how scenarios can be described, what road traffic data sources are possible, what the assessment process may resemble, and how a representative test scenario catalog is defined.

The content of a test scenario can be described using the six-layer model (6LM), in which the six layers describe, for example, the road network, traffic guidance and dynamic objects, as well as environmental conditions [16], as shown in Table 1.1. Witt et al. have also recommended considering information on the road users involved, including the drivers' age and driving experience, when using human driver behavior models to compare the safety performance of ADSs with that of human drivers [17]. In addition, Menzel et al. have hierarchically differentiated test scenarios at the level of detail in which the six layers are described as functional, logical, and concrete scenarios [18]. While functional scenarios are described semantically, logical scenarios specify parameter ranges, and concrete scenarios provide specific parameters obtained from the parameter ranges.

In the conceptualization or pre-development phase of ADSs, the comparison of the safety performance of ADSs with that of human drivers has to be performed prospectively. For that reason, the comparison is usually performed virtually by examining, e.g., the number of crashes per test scenario for ADSs and human drivers in a large number of stochastic traffic simulation runs [8, 13].

Table 1.1.: **Six-layer model (6LM), as proposed and illustrated by Scholtes et al. [16].**

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
Road network and traffic guidance objects	Roadside structures	Temporary modifications of Layers 1 and 2	Dynamic objects	Environmental conditions	Digital information
					
Roads, sidewalks, parking lots, road markings, traffic signs and lights	Buildings, guardrails, street lamps	Roadwork, temporary signs, covered markings	Road users (moving and non-moving), animals, miscellaneous objects	Illumination, precipitation, road weather, wind	Vehicle2X messages

Stochastic traffic simulations also vary the parameters of the given test scenarios [8,13,19] in order to cover the scenario space (i.e., scenario population), which consists of accident, critical, and normal driving scenarios, as thoroughly as possible [8,13]. In Siebke et al. [20], normal driving scenarios are scenarios, which are free of road traffic accidents (RTAs) [20], with a RTA occurring "*when at least one road user is injured or property damage is caused*" [20, p. 1420]. In critical scenarios, the probability of RTAs increases compared with normal driving scenarios [20]. The test scenario parameters to be varied within a simulation run are typically parameters, including initial ego/agent speeds or initial positions on roads [8]. Furthermore, when human driver behavior models are incorporated, human driver-related parameters such as reaction time and gaze duration, can be varied [8]. Depending on the sampled gaze distribution, for example, a scenario that previously led to an accident due to the driver's averting their gaze can now become a non-critical (i.e., normal driving) scenario, and vice versa. Consequently, scenarios not observed in the real world are revealed by using the stochastic component of the simulation. Figure 1.1 also illustrates the application of a car–bicycle scenario in the stochastic traffic simulation used in the literature [8].

Owing to the variety of existing driver behavior models and simulation approaches [20], this thesis focuses on providing scenarios for assessing ADSs using the stochastic traffic simulation as part of the *Dresdner Methode* ('Dresden method' or 'DD method') [13].

The term *stochastic traffic simulation* in this thesis therefore refers to the simulation presented by Siebke et al. [8,20] incorporating a human driver behavior model. The DD method is a data-driven ADS assessment framework covering the complete toolchain, starting from road traffic data collection, through test scenario catalog generation, and on to virtual assessment [13]. Consequently, the stochastic traffic simulation requires logical scenarios according to Menzel et al.'s classification [18] that can be stochastically varied within the given parameter ranges [8, 21].

It is important to note that there are parameters, and thus parameter ranges, that are part of the driver model, including gaze fixation and duration, and that include initial start speeds, which are provided by the logical scenarios fed into the simulation [8]. Thus, logical scenarios do not have to contain all the parameters to be varied within the stochastic traffic simulation.

At a minimum, the following information about dynamic objects (i.e. Layer 4, see Table 1.1), particularly of the ego vehicle (i.e., the vehicle with the ADS being tested) and agents (i.e., other road users), have to be included in the logical scenarios [4, 8]: start positions, maneuvers (e.g., go straight, turn right, and turn left), and speed at the start.

(Logical) test scenarios can be generated, for example, using expert knowledge (i.e., knowledge-driven) or road traffic data (i.e., data-driven) [1, 22]. In the data-driven approach, which is followed in this thesis and in the DD method [13], road traffic data sources can be RTA data recorded by the police (PD), in-depth investigation studies (IDSs), video-based traffic observations (VOs), or naturalistic driving studies (NDSs) [2,13,19,22,23]. To ensure a comprehensive ADS safety performance assessment, the test scenarios used should cover all scenarios – normal driving, critical, complex, and RTA scenarios [1, 8, 24] – that occur in the ODD of the ADS tested. The ODD specifies the domain and conditions for which the ADS should operate safely, including road layout, speed ranges, and environmental conditions [25]. Consequently, road traffic data sources should continuously monitor and record the traffic in the ODD of the ADS tested to obtain a catalog of logical test scenarios that covers the scenario space as thoroughly as possible, even before stochastic variation. In particular, the catalog of logical test scenarios obtained from the road traffic data sources monitoring the ODD should be representative, meaning that it "*represents reality at a given point in time and is thus representative of the ODD inherent to the SuT [(system under testing)] addressed by the catalog, if the distribution of the scenarios in the test scenario catalog and in real traffic matches at the given point in time*" [2, p. 9]. Thus, a representative test scenario catalog ensures that all relevant test scenarios of the ADS's ODD can be considered in the ADS assessment at a certain point in time.

1.2. Problem and objective

However, there are at least three major challenges (C1–C3), as indicated above, to overcome when creating representative test scenario catalogs using road traffic data sources [7, 14]:

(C1) Limited availability of road traffic data sources;

(C2) Divergent information between different types of road traffic data sources; and

(C3) Reliable identification of all scenarios in the road traffic data sources.

Concerning the first challenge (C1), the limited availability of road traffic data sources means that there is no standardized or continuous collection of real-world traffic data yet. While RTA data are often collected nationwide as a total survey and in a standardized and continuous manner by the police (i.e., PD) [26], such is not the case for VOs and NDSs, which often represent only samples from a superordinate population. To date, there are only temporally and spatially limited VOs [23, 27, 28] and NDSs [23, 29, 30], which often differ in how the data were collected (see Table 1.2). Owing to high financial and organizational costs, comprehensive and continuous real-world data collection, including drone-based VOs, will also be unlikely in the future [22].

Furthermore, road traffic data sources differ in the content and level of information they can provide (C2) to describe the layers of the 6LM (see Table 1.2) [2]. The disparities are most pronounced in Layer 4, which deals with dynamic objects. While the exemplary VO, "ListDB," can provide dynamic information about all visible objects in a video in principle, it cannot provide any data on objects not within the camera's or sensor's range (e.g., the driver of a car or the driver's visibility limitations) [2]. As a result, details about drivers or road users (e.g., age, gender, driving experience, and vehicle model) remain unknown [2]. By contrast, the police in Germany collect information about participants involved in an RTA and record the conflict situation as a variable called "$\mathrm{AccidentType}$" [31, 34], but they do not provide reconstructed RTAs for research purposes [2]. Therefore, the German PD do not include any information about the temporal development of dynamic objects [2]. By contrast, IDSs, including the German In-Depth Accident Study (GIDAS), provide information about participants as well as dynamic information through reconstruction [2, 32, 33]. However, GIDAS, for example, collects only RTAs with personal injuries and not ones with property damage [2, 32, 33]. Moreover, NDSs often have a limited view on the ego vehicle and cannot capture the temporal development of all surrounding objects, as VOs can [2]. Table 1.2 shows that there is no single road traffic data source that can perfectly describe all six layers of the 6LM. Even so, different road traffic data sources can complement each other based on the information required and the ODD addressed by the ADS [2].

Table 1.2.: **Qualitative comparison of four types of road traffic data sources [2, 27, 29, 32–34, 36] with respect to the 6LM [16]**. x = information is determinable directly or via post-processing using also additional data sources; X = delivers the best information. In the table, it is assumed that either no records from event data recorders (EDRs) are available for RTAs or that they are not available for all participants involved, for they will not be mandatory for all new cars registered in the EU until July 2024 [37]. In addition, German PD does not yet contain EDR data such as EDRs records (e.g., the vehicle speed and steering angle 5 seconds before a crash) [37]. Layer 6 was excluded from comparison due to missing information.

Road traffic data source	German police accident data	GIDAS	SHRP 2	ListDB (drone-based)
Data source type	PD	IDS	NDS	VO
Objective	All RTAs	RTAs with personal injury only	Normal-driving behavior	
Period	Continuous collection		Collection in fixed period	
Area or coverage	Nationwide	Regional or specified locations		
Study design	Total survey	Sample		
Perspective	Global: Post-event reconstruction		Ego vehicle view	Bird's-eye view
Layer 1: Road network and traffic guidance objects	x	x	x	x
Layer 2: Roadside structures	x	x	X	X
Layer 3: Temporary modifications of L1 and L2	x	x	x	X
Layer 4: Dynamic objects: Information regarding participants	x	X	X	Only visible information

Road traffic data source	German police accident data	GIDAS	SHRP 2	ListDB (drone-based)
Layer 4: Dynamic objects: Temporal development	None	Reconstructed	For all objects covered by ego-vehicle sensors	For all visible objects
Layer 5: Environmental conditions	x	x	x	x

Last, even if the first two challenges (C1 and C2) can be solved, the third challenge (C3), the reliable identification of all scenarios that occur in road traffic data sources, remains. Even if a "*high-performance scenario detector*" [22, p. 12], building on all current scenario identification approaches – that is, rule-based, supervised, or unsupervised approaches [22] – could be trained, it is impossible to verify whether all "*known or unknown scenarios*" [22, p. 12] were reliably detected given the volumes of road traffic data that have to be processed.

Independent from all three challenges (C1–C3), research has been conducted in data fusion, particularly in the areas of market research and social sciences, to merge different samples into one database [41,42], which is comparable to the idea of fusing scenario data sets emerging from different road traffic data samples. In the statistical context, *data fusion* refers to the process of combining statistically heterogeneous, incomplete, and nonconforming sample surveys [41]. Particularly in the case of statistical matching (SM), the aim "*is to create a single data set that can be regarded as a sample from the joint distribution of all variables of interest*" [41, p. 479]. Consequently, the data set created by fusion is assumed to be representative of the joint unobserved population of the fused data sets. Figure 1.2 illustrates the principle of asymmetrical SM by showing the combination of data set B, containing the random variables $X_1, ..., X_P$ and Z, with data set A, containing the random variables Y and $X_1, ..., X_P$. Last, the new data set A* is constructed by matching the variable Z from data set B into data set A with the help of the common variables $X_1, ..., X_P$. As displayed in Figure 1.2, asymmetric fusion leads to the fact that the resulting data set A* has only the size of the recipient A and not the size of A and B together.

When applied to C2, the concept of asymmetrical SM could help to create a representative test scenario catalog containing logical test scenarios by fusing scenarios identified in different road traffic data sets. For example, the missing dynamic information in PD can be supplemented by dynamic information from the observed traffic conflicts in a VO using scenario fusion.

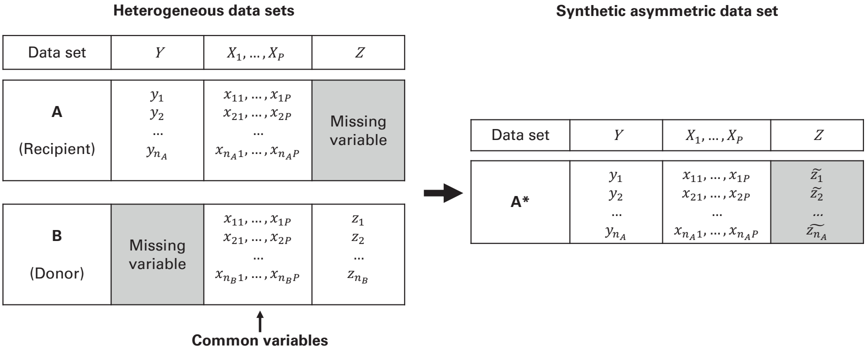

Figure 1.2.: Schematic representation of asymmetric data fusion using statistical matching [3].

To date, data-driven approaches to generate scenarios have not used scenario fusion [15,22,38,39]. Only Krause and Busch [40] matched the previously mentioned variable "$\mathrm{AccidentType}$" collected by the German police to NDS data using record linkage. However, Krause and Busch [40] did not aim to create a representative scenario catalog by fusing different scenario data sets. Moreover, the matching of, for instance, weather-related information to identified scenarios using timestamps as proposed by Ko et al. [35] corresponds to record linkage. With respect to C3, the stochastic variation in the stochastic traffic simulation during the ADS assessment could help to cover the entire continuum of scenarios from normal driving to critical and accident scenarios in the ODD's scenario space. Concerning C1, the exemplary fusion of PD scenarios with VO scenarios is also interesting insofar as PD and thus also German PD is often available to researchers on a large scale as a total survey. Moreover, VO can be performed cost-effectively using drones and/or stationary cameras. Beyond that, all types of RTAs, with or without personal injury, are considered when using the German PD.

Consequently, this thesis addresses the following research question (RQ):

Research question

How can a test scenario catalog
(being ODD representative in combination with a stochastic traffic simulation incorporating human driver behavior models)
be derived by fusing the scenarios identified in (German) PD and VOs?

To answer the research question, the following working hypothesis has been formulated:

Working hypothesis

Statistical matching enables the fusion of test scenarios identified in (German) PD and VOs.

Therefore, the objective of this thesis is to develop a process model for the fusion of scenarios identified from heterogeneous road traffic data sources using SM. The developed process model is then applied to the fusion of scenarios identified from German PD and a VO conducted for the thesis. A hypothetical intersection autonomous emergency braking system (AEB), which supports the driver in car–car conflicts and operates at selected intersections in Dresden, Germany, serves as an exemplary ADS for which the scenarios in PD and VOs are to be found.

To summarize all of the above, Figure 1.3 illustrates the relationships between the challenges addressed, the research question, and the working hypothesis. It should be emphasized that challenge C3 is addressed using the stochastic traffic simulation developed by Siebke et al. [8, 20].

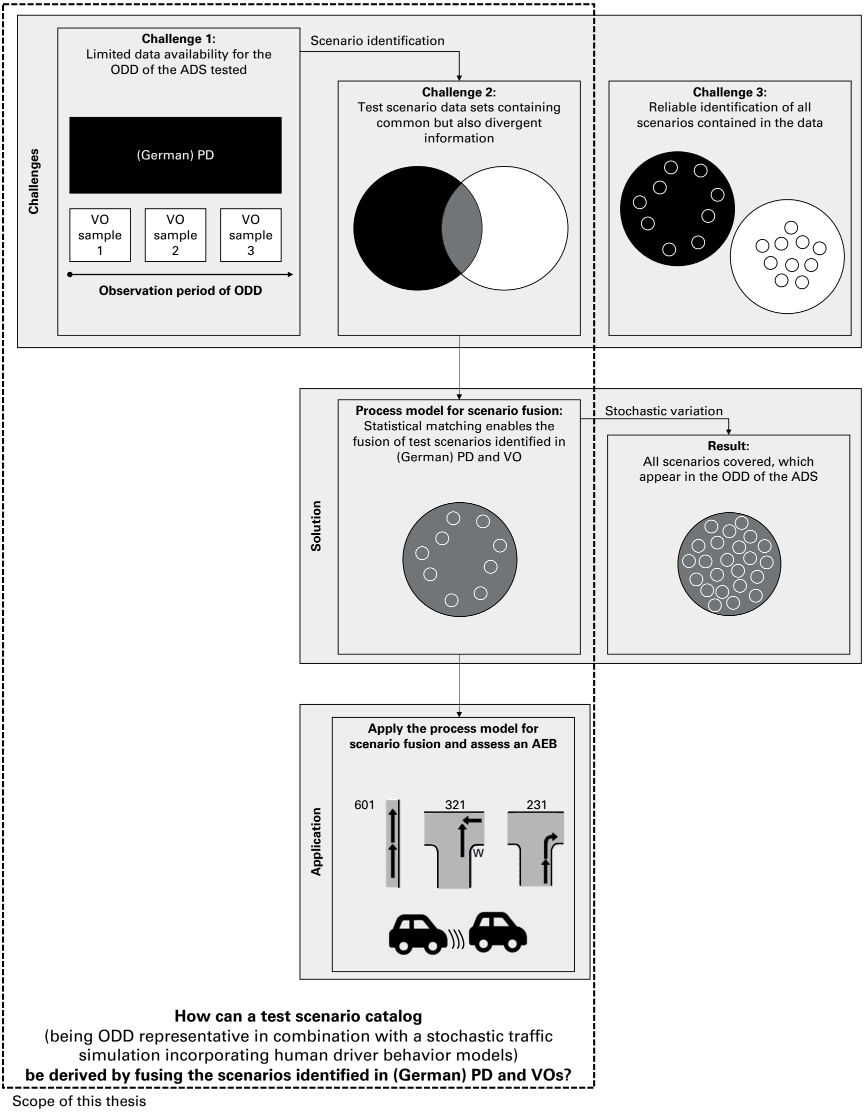

Figure 1.3.: Relationships between the challenges addressed, the research question, and the working hypothesis. Pictograms of accident type provided by Gesamtverband der deutschen Versicherungswirtschaft e.V. [31].