Coverage Metrics for a Scenario Database for the Scenario-Based Assessment of Automated Driving Systems

Erwin de Gelder[1*], Maren Buermann[1], Olaf Op den Camp[1]

Abstract— **Automated Driving Systems (ADSs) have the potential to make mobility services available and safe for all. A multi-pillar Safety Assessment Framework (SAF) has been proposed for the type-approval process of ADSs. The SAF requires that the test scenarios for the ADS adequately covers the Operational Design Domain (ODD) of the ADS. A common method for generating test scenarios involves basing them on scenarios identified and characterized from driving data.**

This work addresses two questions when collecting scenarios from driving data. First, do the collected scenarios cover all relevant aspects of the ADS' ODD? Second, do the collected scenarios cover all relevant aspects that are in the driving data, such that no potentially important situations are missed? This work proposes coverage metrics that provide a quantitative answer to these questions.

The proposed coverage metrics are illustrated by means of an experiment in which over $200\,000$ scenarios from 10 different scenario categories are collected from the HighD data set. The experiment demonstrates that a coverage of $100\,\%$ can be achieved under certain conditions, and it also identifies which data and scenarios could be added to enhance the coverage outcomes in case a $100\,\%$ coverage has not been achieved. Whereas this work presents metrics for the quantification of the coverage of driving data and the identified scenarios, this paper concludes with future research directions, including the quantification of the completeness of driving data and the identified scenarios.

I. INTRODUCTION

The road traffic system is changing rapidly, due to changes in the existing mobility system (e.g., the increasing share of cycling), the introduction of new mobility systems such as Connected, Cooperative, and Automated Mobility (CCAM) systems, shared mobility concepts, and new enabling technologies such as artificial intelligence and wireless V2X-communication [1]. Simultaneously, vehicles are increasingly automated. The goal of automation is to make mobility services available and safe for all, including vulnerable road users. At the same time, automation can provide more comfort to drivers and passengers and increase the efficiency of the mobility system.

Authorities are being asked to allow vehicles equipped with new advanced communication and automation technologies onto public roads. To put a legal framework to the safe deployment of Automated Driving Systems (ADSs), regulations are being implemented by the United Nations Economic Commission for Europe (UNECE), e.g., for Automated Lane Keeping Systems (ALKSs) [2], and the European Commission, i.e., for CCAM systems in four use cases [3]. The UNECE WP.29 Working Party on Automated/Autonomous and Connected Vehicles (GRVA) has developed the New Assessment/Test Methods (NATM) Master Document [4], in which a multi-pillar Safety Assessment Framework (SAF) is proposed for the type-approval process of CCAM systems.

Though ADSs might be complex, and the assessment procedure of such systems might be complicated, the assessment results should be unambiguous, easily understood by experts in the field, and explainable to authorities and the public. This is one reason that scenarios, as a structured way to describe the large varieties of situations and conditions that an ADS may encounter on the road, form the most important source of information to generate test scenarios for the different ways of testing: virtual testing using computer models and simulation tools, track testing under realistic and reproducible conditions, and real-world tests on the road, e.g., by means of field operational tests.

A common way of collecting (naturalistic driving) data for the identification and characterization of scenarios is by the use of instrumented vehicles driving on public roads [5]. In such data collection efforts, the ego vehicle refers to the vehicle that is perceiving the world through its sensors or the vehicle that must perform a specific task. A scenario describes any situation on the road encountered by the ego vehicle and how this situation develops over time. A drive on the road is considered a continuous sequence of scenarios — which might overlap.

An important characteristic of an ADS is represented by its Operational Design Domain (ODD), which refers to the "operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics" [6]. In other words, the ODD is used to describe the external environment and conditions for which an ADS system is designed for. The importance of describing the ODD is underlined by the different initiatives for specifying an ODD, e.g., see [7]–[9]. Since scenarios are a description of the environment of the ADS-equipped vehicle, scenarios can be used to describe the ODD.

The trustworthiness of the safety assessment results of an ADS depends on the quality of the selection of test

The research presented in this work has been made possible by the SUNRISE project. This project is funded by the European Union's Horizon Europe Research & Innovation Actions under grant agreement No. 101069573. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

[1]TNO, Integrated Vehicle Safety, Helmond, The Netherlands
[*]Corresponding author: erwin.degelder@tno.nl

scenarios, and consequently on determining the range of scenarios that is relevant for the ADS under assessment. Hence, trustworthiness depends on how well the underlying data for scenario identification and collection as well as the selected set of scenarios cover the ODD. In this paper, we study metrics to quantify how well observed scenarios cover an ODD.

Coverage can be measured in multiple ways, which is why this work presents multiple coverage metrics. The first coverage metric quantifies whether all relevant aspects of an ODD — labeled with the use of tags — are captured. The other coverage metrics measure the coverage of the data containing scenarios with respect to time, actors, and a combination of time and actors. To illustrate the use of the proposed metrics, they have been implemented to determine the coverage for the StreetWise Scenario Categories (SCs) on the HighD data set [10]. We demonstrate how the resulting numbers can be used to determine what data and which SCs may need to be included for increasing the trustworthiness of the safety assessment results.

After a short overview of related works in Section II, the different coverage metrics are introduced in Section III. Results of the experiment are provided in Section IV. After discussion of the results in Section V, conclusions are drawn and recommendations for further research are given in Section VI.

II. Related Works

The term "coverage" is informally understood as the degree to which something deals with something else. For example, in statistics, the coverage of a confidence interval of p indicates the actual probability that the confidence interval contains p [11]. In the field of software engineering, coverage is "a measure of verification and completeness" [12] and, as described by Piziali [12], there is no single (best) way to define coverage. Coverage metrics can be tailored to the verification progress from the perspective of the functional requirements of the product (functional coverage), the part of the code that have been executed during the verification (code coverage), the fractions of assertions evaluated (assertion coverage), etc. To illustrate the many "coverage" metrics, consider the code coverage. The code coverage can be measured in terms of the lines of codes that have been executed, the branches or the paths, etc.

In [13], the importance of coverage metrics is emphasized for the testing of autonomous vehicles, as the authors argue that testing that inadequately covers the situations that an autonomous vehicle will encounter is similar to inadequate testing. Therefore, Alexander *et al.* [13] have proposed the use of a "situation coverage metric". They have suggested that such a metric should be tractable, which has two implications. First, a percentage should be calculated. Thus, when considering an ADS, the number of kilometers driven or the number of (simulated) scenarios are insufficient examples, because both can be infinite. Also, the number of failures found is not a good example, because the total number of failures is unknown. Second, a $100\,\%$ coverage should be achievable under realistic practical conditions.

In the field of testing of ADSs, the term "coverage" is often used as a measure used to decide the adequacy of a testing effort and as a stopping criteria for testing [14]. Riedmaier *et al.* [15] defined the term "scenario coverage" as the extent to which concrete scenarios used for testing cover the entire space, without further defining quantitative measures. In [16], this idea is further exploited as several metrics are proposed for measuring the coverage of concrete scenarios with respect to the ODD of an ADS. Note that given the fact that the number of concrete scenarios is virtually infinite [17], and following the aforementioned reasoning of Alexander *et al.* [13], using concrete scenarios will not provide a good coverage metric. As an alternative for concrete scenarios, the types of scenarios or the scenes may be considered, where a scene refers to the situation at a single time instant of a scenario. Although Hauer *et al.* [18] did not mention the term "coverage", the metric that they proposed estimates the number of types of scenarios that are not addressed during testing. In [19], a coverage metric is defined using scenes, although no practical results are presented.

When deriving test scenarios from scenarios observed in real-world data, the real-world data should provide good coverage. Compared with the amount of literature on coverage regarding the testing effort of ADSs, there is little literature available regarding the coverage of the real-world data. In [20], a criterion is proposed for the collection of naturalistic driving data. In [21], the asymptotic mean integrated squared error of an estimated probability density function is used as a metric to quantify the coverage of the collected data. A disadvantage of both these works is that a $100\,\%$ coverage cannot be achieved. In [22], a metric is proposed based on the number of distinct sequences of maneuvers of an observed object. A disadvantage of this metric is that the total number of distinct sequences is unknown, so a percentage cannot be calculated. Recently, Glasmacher *et al.* [23] defined coverage with respect to a set of scenarios as "the quantifiable extent to which a set of scenarios or parameters represent a defined ODD or predefined set of scenarios", but no metric has been proposed as [23] focused on completeness instead (more on that in the discussion of Section V). Glasmacher et al. proposed a coverage metric based on scenario parameter values in [24]. However, this approach requires selecting a parameterization and limiting the number of parameters, as achieving $100\,\%$ coverage could be impractical otherwise. Despite this drawback, their approach is promising and complements the method presented in this work.

III. Coverage Metrics

In this work and in line with [23], coverage is defined as the degree to which a set of scenarios observed in real-world data cover an ODD. To further distinguish the metrics that are proposed later in this section, two types of coverage are considered, both aiming to answer different questions:

- Type I: Do the collected scenarios cover all relevant aspects of an ODD?
- Type II: Do the collected scenarios cover all relevant aspects that are in the driving data?

Four different coverage metrics are proposed. The first metric is the tag-based coverage, which addresses coverage type I. The other three metrics, i.e., time-based coverage, actor-based coverage, and actor-over-time-based coverage, address coverage type II.

quantifiable extent to which a set of scenarios or parameters represent an ODD

A. Tag-based coverage

Before introducing the tag-based coverage metric, we need to distinguish scenarios from SCs [25]. Here, a scenario refers to a quantitative description of the relevant characteristics of the ego vehicle, its activities and/or goals, its static environment, and its dynamic environment. In contrast, an SC refers to a qualitative description of the ego vehicle, its activities and/or goals, its static environment, and its dynamic environment. For example, the SC "cut in" comprises all possible cut-in scenarios. Scenarios may further be enriched with tags, e.g., a scenario belonging to the SC "cut in" may have the tag "actor at left" to indicate that there is an actor at the left side of the ego vehicle that prevents the ego vehicle from changing lane to the left.

Let $\mathcal{L}$ denote a set of tags and let $\mathcal{C}$ denote a set of SCs. Note that the set of tags should be based on the relevant aspects of an ODD, whereas the set of SCs could be based on the coverage type II metrics presented in Sections III-B–III-D. For the tag-based coverage, we make use of the function $N(L, C)$, which returns the number of scenarios that belong to SC C and contain the tag L. Continuing the previous example, in case we have 10 cut-in scenarios with an actor at the left of the ego vehicle, we would have $N(\text{Actor at left}, \text{Cut-in}) = 10$. The tag-based coverage metric is defined as follows:

$$\mathrm{Coverage}_{\mathrm{Tag}}(n) = \frac{1}{n|\mathcal{L}||\mathcal{C}|} \sum_{L \in \mathcal{L}} \sum_{C \in \mathcal{C}} \min(n, N(L, C)), \quad (1)$$

where $n \in \mathbb{Z}^+$ and $|\cdot|$ denotes the cardinality, e.g., $|\mathcal{L}|$ equals the number of (distinct) tags. In case $\mathrm{Coverage}_{\mathrm{Tag}}(1) = 1$, each tag is associated to at least one scenario of each SC.

For this coverage metric, three choices need to be made:

1) The SCs belonging to $\mathcal{C}$. The SCs should cover the ODD. The set of SCs could be based on relevant literature [26], [27], though we suggest using other coverage metrics to justify that the set of SCs is complete. As mentioned before, the metrics presented in Sections III-B–III-D may be used.
2) The tags belonging to $\mathcal{L}$. The tags should follow from the ODD description. When defining the ODD in accordance with the ISO 34503 standard [7], the corresponding tags listed in the ISO 34504 standard [28] may be used.
3) The required number of tags per SC, n. Minimally, $n = 1$, but to achieve more accurate statistics, if may be required to choose a higher value for n. To determine n, other metrics be used, e.g., see [20]–[24].

To obtain more accurate statistics of the scenarios belonging to an SC, it may be desired to have at least several scenarios of each SC with a certain tag. In that case, a larger value of n may be chosen.

Note that different tag-based coverage metrics can be defined if different sets of tags are considered. For example, one may choose to calculate (1) with $\mathcal{L}$ consisting of tags related to environmental conditions, such as weather and lighting conditions, and with another set of tags consisting of scenery attributes, such as different types of roads.

B. Time-based coverage

The time-based coverage metric answers the question of whether all timestamp in the data is covered by one or more scenarios. Let $\mathcal{T}$ denote the set of all timestamps in the data set. For the time-based coverage, we introduce the function $M(t)$, which returns the number of scenarios at time t. Note that it may be possible that scenarios happen in parallel, e.g., a leading vehicle decelerating and another vehicle overtaking the ego vehicle. The time-based coverage metric is defined as follows:

$$\mathrm{Coverage}_{\mathrm{T}}(n) = \frac{1}{n|\mathcal{T}|} \sum_{t \in \mathcal{T}} \min(n, M(t)), \quad (2)$$

with $n \in \mathbb{Z}^+$. In case $\mathrm{Coverage}_{\mathrm{T}}(1) = 1$, all timestamps in the data are covered by at least one scenario. To account for the number of scenarios that can occur in parallel, one can increase the value of n.

C. Actor-based coverage

The actor-based coverage metric answers the question of whether every relevant actor is covered by at least one scenario. Let $\mathcal{A}$ denote the set of relevant actors. Here, the term "relevant" could be defined using some conditions. For example, $\mathcal{A}$ could contain all actors that are at some point in time within a certain distance of the ego vehicle. Alternatively, $\mathcal{A}$ could contain all emergency vehicles in the data set, etc. Let $\mathcal{B}$ denote the set of actors that are part of at least one scenario. Then, the actor-based coverage metric is defined as follows:

$$\mathrm{Coverage}_{\mathrm{A}}(\mathcal{A}) = \frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{A}|}. \quad (3)$$

D. Actor-over-time-based coverage

Achieving $\mathrm{Coverage}_{\mathrm{A}}(\mathcal{A}) = 1$ means that all actors of the set $\mathcal{A}$ are part of at least one scenario. However, it does not consider the temporal aspect of when these actors are part of a scenario. For example, it could be the case that an actor is near the ego vehicle — and thus part of $\mathcal{A}$ — but only part of a scenario once this vehicle is far away. To accommodate the time aspect, we introduce the fourth coverage metric; the actor-over-time-based coverage.

Let $\mathcal{T}_a$ denote the set of timestamps at which the actor $a \in \mathcal{A}$ satisfies the conditions that makes this actor part of $\mathcal{A}$. Furthermore, let $K(a, t)$ be the number of scenarios at time t that contain actor a. Then, the actor-over-time-based coverage is defined as follows:

$$\mathrm{Coverage}_{\mathrm{AT}}(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \sum_{a\in\mathcal{A}} \frac{1}{|\mathcal{T}_a|} \sum_{t\in\mathcal{T}_a} \min(1, K(a,t)). \quad (4)$$

IV. RESULTS

To illustrate the use of the coverage metrics that are presented in Section III, the metrics are evaluated based on scenarios from real-world data. The setup of the experimental results are presented in Section IV-A and the results are presented in the subsequent subsections.

A. Setup experiment

The HighD data set [10] is chosen for the experiment because of its size (more than $40\,000\,\mathrm{km}$ of naturalistic driving data) and high accuracy. The data consists of trajectories of cars and trucks at six different locations on German motorways obtained using video footage from drones.

To obtain the scenario data, each of the more than $100\,000$ vehicles is treated as an ego vehicle once. I.e., from the total data set, more than $100\,000$ smaller data sets are created, where each of the smaller data sets contains a single ego vehicle and trajectory data relative to the ego vehicle as if the other vehicles are perceived from the ego vehicle. It is assumed that the ego vehicle can see all of its surrounding vehicles within a distance of $100\,\mathrm{m}$. Each of the smaller data sets stops whenever the ego vehicle is $100\,\mathrm{m}$ from its final position; this is done to avoid the sudden disappearance of vehicles in front of the ego vehicle, as these vehicles would be out of view of the drone camera. Note that, as a result, vehicles with a trajectory less than $100\,\mathrm{m}$ are not considered as ego vehicles. In total, this resulted in $109\,986$ data sets with a single ego vehicle.

Table I lists the 10 scenario categories considered in this study. This table also summarizes the activities of the ego vehicle and the main actor(s). Here, the main actor(s) refers to the actor(s) that are necessary for the scenario to occur. That is, there may be other actors participating in the scenario as well, e.g., a vehicle overtaking the ego vehicle in the leading vehicle cruising scenario. Based on the activities of the ego vehicle and the main actors and the approach outlined in [29], the scenarios are automatically extracted. Table I also indicates the number of scenarios found for each SC.

B. Results tag-based coverage

For the tag-based coverage, 18 different tags are considered, see Table II. The first two tags apply to the two different types of vehicles that are considered in the HighD data set. Tags L_3 to L_{10} relate to the initial position of a vehicle with respect to the ego vehicle. The tags L_{11} and L_{12} apply if an actor is substantially slower or faster than the ego vehicle, respectively. The remaining tags describe the longitudinal (L_{13} to L_{15}) and lateral (L_{16} to L_{18}) activities of vehicles surrounding the ego vehicle. I.e., if there are 5 cars surrounding the ego vehicle, than the tag L_1 is applied only once.

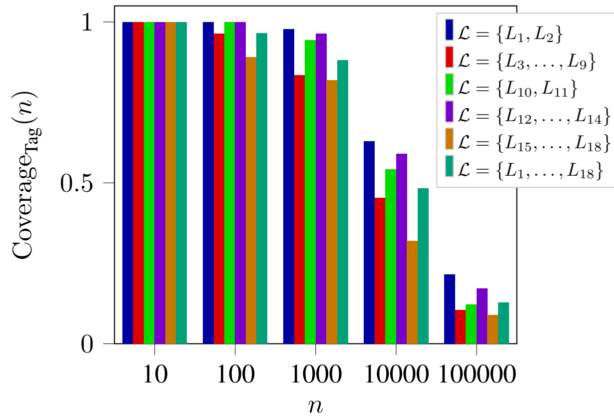

Fig. 1. Results of the tag-based coverage.

Table II lists the results of the number of scenarios that contain a certain tag, i.e., the function $N(L,C)$ that has been introduced in Section III-A. Note that some tags are by definition part of a scenario, e.g., scenarios belonging to C_1, C_2, C_3, or C_6 always contain the tag L_3.

Fig. 1 shows the tag-based coverage that results from the numbers listed in Table II. It also shows how the tag-based coverage depends on the set of tags that are considered. For any choice of $\mathcal{L}$, we have $\mathrm{Coverage}_{\mathrm{Tag}}(10) = 1$, meaning that for each tag there are at least 10 scenarios from each SC that contain that tag. We still have $\mathrm{Coverage}_{\mathrm{Tag}}(100) = 1$ when considering tags L_1, L_2, and L_{10} to L_{14}. For increasing values of n, $\mathrm{Coverage}_{\mathrm{Tag}}(n)$ starts to decrease. To further investigate why $\mathrm{Coverage}_{\mathrm{Tag}}(n) < 1$, the numbers in Table II can be studied. For example, given the relatively low occurrence of scenarios from SCs C_7 and C_8, the counts of the tags for these SCs are also relatively low.

C. Results time-based coverage

When calculating the time-based coverage, the time instants of each dataset containing an ego vehicle are treated separately, after which the results are combined and shown in Fig. 2. Fig. 2 shows that about $75\,\%$ of the time instants are covered by at least a single scenario. This indicates that still a substantial portion of the time instants are not covered by a scenario. It requires further investigation to determine if important scenario categories are missed. In this case, in the remaining $25\,\%$ of the time instants, there is no actor that complies with any of the descriptions listed in Table I, or simply no other actor at all. Indeed, if we would add the scenario category "ego vehicle has no leading vehicle", then $\mathrm{Coverage}_{\mathrm{T}}(1) = 1$.

D. Results actor-based coverage

Fig. 3 shows the result of the actor-based coverage for different sets of actors. Consider an imaginary box with a certain size around the ego vehicle. Then, $\mathcal{A}$ contains all actors that are at some point in time within this box. To obtain multiple values of $\mathrm{Coverage}_{\mathrm{A}}(\mathcal{A})$, the size of this imaginary box has been varied, as indicated in Fig. 3. For $\mathcal{B}$ (the set of actors that are part of a scenario) only the main

TABLE I

DESCRIPTION OF THE 10 SCENARIO CATEGORIES THAT ARE CONSIDERED IN THIS WORK'S EXPERIMENT.

Symbol	Name	Ego vehicle activity	Main actor(s) activity	Count
C_1	Leading vehicle cruising	Keeping lane	Keeping lane and cruising	102308
C_2	Leading vehicle accelerating	Keeping lane	Keeping lane and accelerating	22296
C_3	Leading vehicle decelerating	Keeping lane	Keeping lane and decelerating	20351
C_4	Approaching slower vehicle	Keeping lane	Keeping lane and driving slower than ego vehicle	5052
C_5	Cut-in in front of ego vehicle	Keeping lane	Changing lane to become leading vehicle	2992
C_6	Cut-out in front of ego vehicle	Keeping lane	Leading ego vehicle and then changing lane	3069
C_7	Changing lane with vehicle behind	Changing lane	Behind ego vehicle on adjacent lane	2156
C_8	Merging into an occupied lane	Changing lane	Both main actors stay in lane and become leading and following vehicles after ego vehicle lane change	819
C_9	Ego vehicle overtaking vehicle	Keeping lane	Keeping lane on overtaken by ego vehicle on adjacent lane	38147
C_{10}	Vehicle overtaking ego vehicle	Keeping lane	Keeping lane and overtaking ego vehicle on adjacent lane	40307

TABLE II

$N(L, C)$ FOR VARIOUS TAGS AND SCENARIO CATEGORIES, WITH THE CORRESPONDING SCENARIO CATEGORIES (SCS) NAMES LISTED IN TABLE I.

Symbol	Tag	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
L_1	Car	102 111	22 292	20 341	5050	2992	3067	2147	819	37 996	40 305
L_2	Truck	81 475	19 406	17 454	4234	2273	2624	1915	734	34 652	31 999
L_3	Same lane in front	102 308	22 296	20 351	5052	1188	3069	834	480	29 295	29 171
L_4	Same lane rear	37 281	11 248	14 377	2386	1339	1597	1006	351	23 666	24 913
L_5	In front left lane	70 385	17 664	15 934	3860	1857	2377	980	578	37 850	14 773
L_6	In front right lane	49 139	9190	8871	2443	2208	1187	820	476	12 388	32 625
L_7	At side left lane	4952	2052	1552	228	161	205	40	17	870	1151
L_8	At side right lane	4850	1284	1162	201	166	95	44	20	1216	1283
L_9	Rear left lane	32 394	12 730	13 183	2243	1245	1750	1205	366	24 979	12 760
L_{10}	Rear right lane	31 462	8052	8741	1777	1387	807	721	275	11 063	37 005
L_{11}	Slower ($\Delta v < -5\,\mathrm{m/s}$)	54 750	13 873	14 138	4021	1348	2369	1528	591	35 107	7480
L_{12}	Faster ($\Delta v > 5\,\mathrm{m/s}$)	41 061	9032	8046	1798	1831	957	931	403	8124	37 569
L_{13}	Cruising	102 308	22 296	20 351	5043	2964	3051	2142	816	37 935	39 660
L_{14}	Accelerating	57 081	22 296	7652	2554	1610	1481	1260	516	21 270	24 039
L_{15}	Decelerating	58 144	9107	20 351	3419	1794	1833	1287	607	21 132	24 804
L_{16}	Keeping lane	102 308	22 296	20 351	5052	2992	3068	2156	819	38 147	40 307
L_{17}	Changing lane left	6771	1405	1759	384	2090	2101	32	13	2545	2668
L_{18}	Changing lane right	4154	1127	982	339	987	1073	12	15	1794	1741

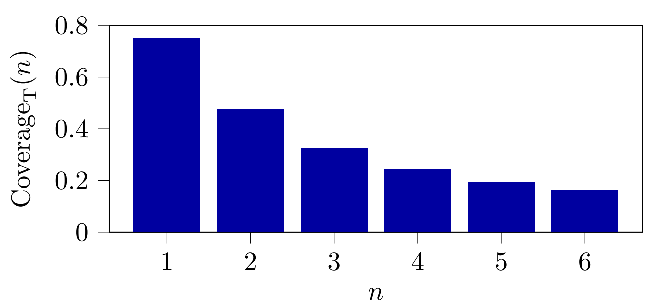

Fig. 2. Results of the time-based coverage.

actors of a scenario, as described in Table I, are considered. Note that if all actors part of a scenario would be considered, the result of the actor-based coverage would be practically similar to the time-based coverage.

When considering only actors that are ahead of the ego and in the ego vehicle's lane (blue solid line in Fig. 3), then $\mathrm{Coverage}_\mathrm{A}(\mathcal{A}) = 1$ if only vehicles that are within $10\,\mathrm{m}$ are considered. In all other cases, there is no full coverage. As with the other coverage metrics, a further investigation into the data is needed to find out why some actors are not a main actor of a scenario, even if vehicles are as near as $15\,\mathrm{m}$. In this study, these non-main actors are vehicles that are in front of the vehicle that the ego vehicle is following. Especially in a traffic jam, vehicles that are ahead of the leading vehicle can still be relatively close to the ego vehicle.

When changing the width of the imaginary box that contains the actors of $\mathcal{A}$, the value of $\mathrm{Coverage}_\mathrm{A}(\mathcal{A})$ drops substantially. This can be explained by the fact that vehicles in the adjacent lane are only considered in several occasions, namely if such a vehicle is overtaken by the ego vehicle (C_9) or if such a vehicle overtakes the ego vehicle (C_{10}). Vehicles that are two lanes away from the ego vehicle are not considered as main actors for any SC, which is why the green lines in Fig. 3 are lower than the corresponding red lines. It can also be noted that a substantially lower actor-based coverage is obtained if the imaginary box extends towards the back of the ego vehicle. This can be explained from the fact that there is only one SC (C_8) that considers a main actor that is or could be behind the ego for the whole duration of the scenario.

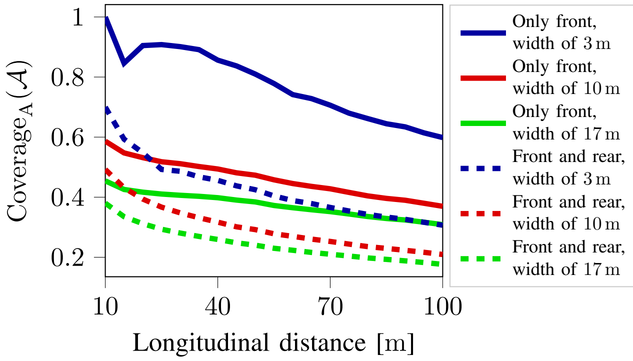

Fig. 3. Results of the actor-based coverage. For the actor set $\mathcal{A}$, every actor is considered that is at some point in time within a certain longitudinal distance of the ego vehicle, varying front 10 m to 100 m (x-axis), and within a lateral distance, varying between 1.5 m (blue), 5.0 m (red), and 8.5 m (green). For the solid lines, $\mathcal{A}$ only contain actors in front, while for the dashed lines, $\mathcal{A}$ also contain rear actors within the specified longitudinal distance.

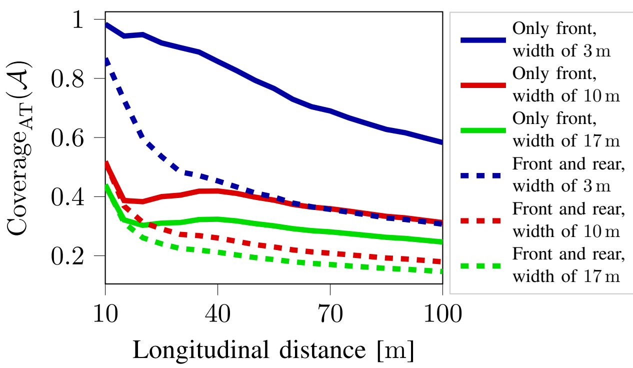

Fig. 4. Results of the actor-over-time-based coverage. See Fig. 3 for a further explanation.

E. Results actor-over-time-based coverage

The results of the actor-over-time-based coverage are shown in Fig. 4. The lines show a similar pattern as seen in Fig. 3. Especially for actors in the same lane and in front of the ego vehicle (blue lines in Figs. 3 and 4), the results are nearly the same. This indicates that those actors are generally part of a scenario as a main actor whenever they are within the imaginary box. In the other cases (i.e., all lines except the blue solid line) the actor-over-time-based coverage is generally a bit lower, which indicates that even if the relevant actors are covered, they are not covered for the entire duration that they are within the boundaries of the imaginary box.

V. DISCUSSION

This work presented four different types of coverage metrics that can be used to determine whether the collected scenarios cover all relevant aspects of an ODD (coverage type I) and whether the collected scenarios cover all relevant aspects that are in the data (coverage type II). The presented coverage metrics can be helpful to decide to collect more data, start dedicated data campaigns to address parts of the ODD that are insufficiently covered, include more SCs, or decide that enough data and SCs are considered. To decide if any of these actions is appropriate, a further investigation is needed while considering the following aspects:

- As described in Section IV, it is not always possible to achieve $\mathrm{Coverage}_\mathrm{Tag}(n) = 1$. However, it may neither be necessary. When considering tags related to the road layout at which the scenario is taking place, it might be the case that some scenario categories do not contain hardly any scenarios with certain tags because certain type of scenarios only occur with a certain road layout. For example, a cut-in of another vehicle far less likely on a single-lane road.
- Even if $\mathrm{Coverage}_\mathrm{T}(n) = 1$, it could mean that important information is not captured. For example, if generic SCs are considered, such as "driving on motorway" and "driving on a non-motorway road", one can already obtain $\mathrm{Coverage}_\mathrm{T}(1) = 1$. Still, it could be the case that important information is not captured by the scenarios, e.g., if the trajectory of a relevant actor is not captured.
- A high coverage type II could be a result of lots of false positives of the detected scenarios. The coverage type II metrics presented in this work only check if some time instants and/or actors are covered by a scenario. It might be possible that a high coverage is obtained due to the false detection of scenarios.
- On the other hand, a low coverage score could be the result of many false negatives. I.e., if all scenarios would have been correctly detected, a higher coverage would be obtained.

Note that achieving both high coverage for type I and type II is generally difficult. For achieving high coverage for type II, one may need to consider many scenario categories, including scenario categories that contain only few scenarios. As a result, it will generally be more difficult to achieve a high score for the tag-based coverage, which is a type I coverage metric. Given that a value of 1 for all coverage metrics is generally difficult, it remains future work to determine — depending on the context — what appropriate values of the coverage are. Note that the required coverage values may depend on the input parameters of those coverage metrics, e.g., the set of tags ($\mathcal{L}$), n for $\mathrm{Coverage}_\mathrm{Tag}(n)$ and $\mathrm{Coverage}_\mathrm{T}(n)$, and the conditions used to determine whether an actor is an element of $\mathcal{A}$ for $\mathrm{Coverage}_\mathrm{A}(\mathcal{A})$ and $\mathrm{Coverage}_\mathrm{AT}(\mathcal{A})$.

We have illustrated the presented coverage metrics by applying them on the scenarios obtained from the HighD data set. The experiment has shown how varying metrics can be obtained by using different tags, SCs, and values for n. While we have considered ten SCs, it is important to note that the number of SCs in a real-world application is likely to be substantially higher. Similarly, the number of tags relevant to an an ODD is expected to be higher in practice. Next to involving more SCs and tags, future work could consider the use of additional or alternative data sets.

For increasing the trustworthiness of the data-driven,

TABLE III
OVERVIEW OF THE DIFFERENT TYPES OF COVERAGE AND COMPLETENESS.

	Coverage	Completeness
Type I	Do the collected scenarios cover all relevant aspects of an ODD?	Do the driving data contain all relevant details of an ODD?
Type II	Do the collected scenarios cover all relevant aspects that are in the driving data?	Do the collected scenarios describe all relevant details that are in the driving data?

scenario-based safety assessment of an ADS, not only achieving a high coverage is important, but also achieving a high completeness is important. Here, "completeness" is not to be confused with "coverage". Where coverage refers to the extent to which data capture aspects of interest — in our case, the ODD — completeness refers to the extent to which the data is free from missing values. Similarly to coverage, two types of completeness could be considered, both aiming to answer different questions:

- Type I: Do the driving data contain all relevant details of an ODD?
- Type II: Do the collected scenarios describe all relevant details that are in the driving data?

An overview of the different types for coverage and completeness is shown in Table III. An aspect considered for type I completeness could be the trajectories of actors that are simply missing from the data. Note that in that case, a high coverage could still be obtained even though important information is missing. Therefore it is essential that the data is free from large omissions, i.e., high completeness type I. Also, a high coverage could be obtained even if the collected scenarios do not contain all relevant aspects that are in the data, which would indicate insufficient completeness for type II. In a recent work from Glasmacher *et al.* [23], a methodology is proposed to argue a sufficient completeness of a scenario concept, which is related to type II completeness. Future research is required to further address this and to develop metrics that can be used to quantify the extent of completeness.

Both the concepts of coverage and completeness as described in Table III refer to the ODD. However, for the safe deployment of an ADS, the ADS must be capable of dealing safely in all operating conditions it encounters during its deployment, rather than the operating conditions it is designed for. To describe the actual operating conditions of an ADS, the term Target Operational Domain (TOD) is coined [7]. Future research should address how to measure the extent to which the ODD covers the TOD.

VI. CONCLUSIONS

To realize the potential benefits of the deployment of Automated Driving Systems (ADSs), the safety should be adequately ensured. Assessing safety can be systematically approached through scenario-based evaluations. Data from (naturalistic) driving can serve as a source for characterizing scenarios. This work proposes two types of coverage metrics to quantify the extent to which the data and the scenarios derived from them cover the ADS' Operational Design Domain (ODD). Coverage type I measures the extent to which the scenarios extracted from driving data cover the relevant details of an ODD whereas coverage type II measures the extent to which the scenarios cover the relevant details in the driving data. The proposed coverage metrics can be used to identify missing data or scenarios that should also be considered for the safety assessment. Hence, the presented metrics may serve as part of the argumentation that adequate (test) scenarios have been accounted for in the safety assessment of an ADS, thus aiding the type-approval process in accordance with the multi-pillar Safety Assessment Framework (SAF) outlined by the United Nations Economic Commission for Europe (UNECE) [4]. Given that achieving $100\,\%$ coverage for all presented coverage metrics is not always practical or necessary, future work involves establishing suitable coverage thresholds. Where this work outlines coverage metrics, a topic for future research is the quantification of the completeness of driving data and the identified scenarios. Additionally, future research should be dedicated to the quantification of the degree to which the ODD encompasses the actual operating conditions of an ADS.

REFERENCES

[1] J. Piao and M. McDonald, "Advanced driver assistance systems from autonomous to cooperative approach," *Transport Reviews*, vol. 28, no. 5, pp. 659–684, 2008.

[2] ECE/TRANS/WP.29/2022/59/Rev.1, "Proposal for the 01 series of amendments to un regulation no. 157 (automated lane keeping systems)," World Forum for Harmonization of Vehicle Regulations, Standard, 2022. [Online]. Available: https://unece.org/sites/default/files/2022-05/ECE-TRANS-WP.29-2022-59r1e.pdf.

[3] European Commission, "Commission implementing regulation (EU) 2022/1426 for the application of regulation (EU) 2019/2144 as regards uniform procedures and technical specifications for the type-approval of the automated driving system (ads) of fully automated vehicles," *Official Journal of the European Union*, vol. L221, pp. 1–64, 2022. [Online]. Available: https://eur-lex.europa.eu/eli/reg_impl/2022/1426/oj.

[4] ECE/TRANS/WP.29/2021/61, "New Assessment/Test Method for Automated Driving (NATM) – Master Document," World Forum for Harmonization of Vehicle Regulations, Tech. Rep., 2021. [Online]. Available: https://unece.org/sites/default/files/2021-04/ECE-TRANS-WP29-2021-61e.pdf.

[5] H. Elrofai, J.-P. Paardekooper, E. de Gelder, S. Kalisvaart, and O. Op den Camp, "Scenario-based safety validation of connected and automated driving," Netherlands Organization for Applied Scientific Research, TNO, Tech. Rep., 2018. [Online]. Available: http://publications.tno.nl/publication/34626550/AyT8Zc/TNO-2018-streetwise.pdf.

[6] SAE J3016, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE International, Tech. Rep., 2021.

[7] ISO 34503, "Road Vehicles – Test scenarios for automated driving systems – Taxonomy for operational design domain for automated driving systems," International Organization for Standardization, Standard, 2023. [Online]. Available: https://www.iso.org/standard/78952.html.

[8] PAS 1883:2020, "Operational design domain (ODD) taxonomy for an automated driving system (ADS) — specification," The Britisch Standards Institution, Standard, 2020.

[9] Association for Standardization of Automation and Measuring Systems (ASAM), *ASAM OpenODD*, Accessed May, 2024, 2021. [Online]. Available: https://www.asam.net/standards/detail/openodd/.

[10] 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 *IEEE 21st International Conference on Intelligent Transportations Systems (ITSC)*, 2018, pp. 2118–2125.

[11] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

[12] A. Piziali, *Functional verification coverage measurement and analysis*. Springer Science & Business Media, 2007.

[13] R. Alexander, H. Hawkins, and D. Rae, "Situation coverage — a coverage criterion for testing autonomous robots," Department of Computer Science, University of York, Tech. Rep., 2015. [Online]. Available: https://eprints.whiterose.ac.uk/88736/.

[14] H. Araujo, M. R. Mousavi, and M. Varshosaz, "Testing, validation, and verification of robotic and autonomous systems: A systematic review," *ACM Transactions on Software Engineering and Methodology*, vol. 32, no. 2, pp. 1–61, 2023.

[15] 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. 87 456–87 477, 2020.

[16] P. Weissensteiner, G. Stettinger, S. Khastgir, and D. Watzenig, "Operational design domain-driven coverage for the safety argumentation of automated vehicles," *IEEE Access*, vol. 11, pp. 12 263–12 284, 2023.

[17] C. Amersbach and H. Winner, "Defining required and feasible test coverage for scenario-based validation of highly automated vehicles," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 425–430.

[18] F. Hauer, T. Schmidt, B. Holzmüller, and A. Pretschner, "Did we test all scenarios for automated and autonomous driving systems?" In *IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 2950–2955.

[19] A. J. Alnaser, A. Sargolzaei, and M. I. Akbaş, "Autonomous vehicles scenario testing framework and model of computation: On generation and coverage," *IEEE Access*, vol. 9, pp. 60 617–60 628, 2021.

[20] W. Wang, C. Liu, and D. Zhao, "How much data are enough? A statistical approach with case study on longitudinal driving behavior," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 2, pp. 85–98, 2017.

[21] E. de Gelder, J.-P. Paardekooper, O. Op den Camp, and B. De Schutter, "Safety assessment of automated vehicles: How to determine whether we have collected enough field data?" *Traffic Injury Prevention*, vol. 20, no. S1, pp. 162–170, 2019.

[22] L. Hartjen, R. Philipp, F. Schuldt, and B. Friedrich, "Saturation effects in recorded maneuver data for the test of automated driving," in *13. Uni-DAS eV Workshop Fahrerassistenz und automatisiertes Fahren*, 2020, pp. 74–83. [Online]. Available: https://www.uni-das.de/images/pdf/fas-workshop/2020/FAS_2020_HARTJEN.pdf.

[23] C. Glasmacher, H. Weber, and L. Eckstein, "Towards a completeness argumentation for scenario concepts," in *IEEE Intelligent Vehicles Symposium (IV)*, 2024.

[24] C. Glasmacher, M. Schuldes, H. Weber, N. Wagener, and L. Eckstein, "Acquire driving scenarios efficiently: A framework for prospective assessment of cost-optimal scenario acquisition," in *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 2023, pp. 1971–1976.

[25] E. de Gelder, J.-P. Paardekooper, A. Khabbaz Saberi, H. Elrofai, O. Op den Camp, S. Kraines, J. Ploeg, and B. De Schutter, "Towards an ontology for scenario definition for the assessment of automated vehicles: An object-oriented framework," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 300–314, 2022.

[26] W. G. Najm, J. D. Smith, and M. Yanagisawa, "Pre-crash scenario typology for crash avoidance research," U.S. Department of Transportation Research and Innovative Technology Administration, Tech. Rep. DOT HS 810 767, 2007. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/6281/dot_6281_DS1.pdf.

[27] E. de Gelder, O. Op den Camp, and N. de Boer, "Scenario categories for the assessment of automated vehicles," CETRAN, Tech. Rep., 2020, Version 1.7. [Online]. Available: http://cetran.sg/wp-content/uploads/2020/01/REP200121_Scenario_Categories_v1.7.pdf.

[28] ISO 34504, "Road Vehicles – Test scenarios for automated driving systems – Scenario categorization," International Organization for Standardization, Standard, 2024. [Online]. Available: https://www.iso.org/standard/78953.html.

[29] E. de Gelder, J. Manders, C. Grappiolo, J.-P. Paardekooper, O. Op den Camp, and B. De Schutter, "Real-world scenario mining for the assessment of automated vehicles," in *IEEE International Transportation Systems Conference (ITSC)*, 2020, pp. 1073–1080.