Quantifying the Impact of Automated Vehicles on Traffic

Sigl, Martin; Prins, Binnert; Schutz, Christoph; Wagner, Sebastian; Schulte, Frederik; Watzenig, Daniel

DOI
10.1109/IAVVC57316.2023.10328115

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the IEEE International Automated Vehicle Validation Conference, IAVVC 2023

Citation (APA)
Sigl, M., Prins, B., Schutz, C., Wagner, S., Schulte, F., & Watzenig, D. (2023). Quantifying the Impact of Automated Vehicles on Traffic. In *Proceedings of the IEEE International Automated Vehicle Validation Conference, IAVVC 2023* IEEE. https://doi.org/10.1109/IAVVC57316.2023.10328115

Quantifying the Impact of Automated Vehicles on Traffic

Martin Sigl[1] Binnert Prins[1] Christoph Schütz[1] Sebastian Wagner[1] Frederik Schulte[2] Daniel Watzenig[3]

***Abstract*— One of the major challenges in the development of Automated Driving is its assessment. It is expected that Automated Vehicles behave differently than human drivers. Therefore, mixed human-robot traffic will yield different and new driving situations as human-only traffic. It is important to know how this mixed traffic will change the composition of traffic situations to be able to quantify the impact Automated Vehicles will have on everyday traffic. This paper presents a methodology on how to find metrics that quantify traffic in order to detect changes in the traffic space that will come with the introduction of Automated Vehicles. Additionally, this methodology provides tools to help with the validation of virtual testing platforms such as simulation.**

***Index Terms*—Autonomous Vehicles, Automated Driving, Impact Analysis**

I. Introduction

After reaching a certain level of automation, testing automated vehicles (AVs) solely based on real-world test drives is infeasible [1]. With rising level of automation, more and more complex situations have to be assessed. Simulations are expected to reduce the effort to test automated driving (AD) functions [1] and are already in use [2], [3].

With the first market introduction of AVs, the fraction of AVs in the real-world traffic will rise from a small number of vehicles to more and more market penetration. Therefore, one can expect a hybrid traffic made out of human drivers and AVs. Since the perception, planning and action of AVs differ from the human driving process, AVs are expected to show a different behavior in real-world traffic scenarios. Therefore, in a hybrid traffic environment, previously unseen compositions of traffic scenarios will occur compared to human-only traffic (see fig. 1).

When developing AD functions, it cannot be expected that these new scenarios are present in any data set. Each change in a AD function may result in a different or new hybrid traffic space. Therefore it is necessary to detect these new scenarios as they occur. In principle, this is possible in both real-world test drives or with virtual testing. Next to checking

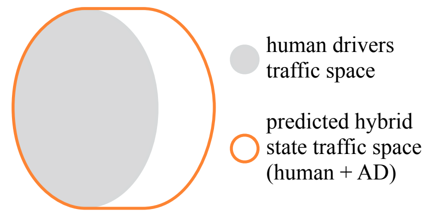


Fig. 1: Predicted traffic space change. With a mix of human drivers and AVs, new and previously unseen scenarios will occur.

if this new traffic space contains safety critical scenarios and offers a comfortable ride complying with traffic rules, it is also necessary to check how the hybrid state impacts the traffic space in general. Quantifying the shift in the traffic space helps in developing AD functions that are accepted by customers. For example, an AV providing a safe and comfortable ride might be unattractive to buy if it constantly travels significantly below the official speed limit, causing traffic jams.

Usually, previous investigations of the impact of AD functions only focus on single evaluation criteria to answer those questions, like ”What will be the change in traffic flow?” or ”What will be the impact on the fuel consumption?” [4]–[7]. This work aims to develop a holistic, more general approach describing and quantifying the changes in the traffic space as precise as possible. This quantification process will rely on virtual test environments because of their property to quickly generate new data.

In Section II, a short overview of virtual testing with simulations is given, as well as an introduction to metrics that describe traffic. It is followed by Section III which presents the used approach to quantify the change caused by the hybrid traffic space. This paper ends with the results in Section IV showing both the evaluation of the presented method to describe changes in traffic space and the impact AVs will have in a hybrid traffic scenarios, and concludes with a summary in Section V.

II. State of the Art

The first subsection states the key principles and requirements for virtual testing which are the foundation for the presented approach. The second subsection provides an overview of state-of-the-art methods to quantify traffic.

A. Simulation for AD Assessment

In the following, two approaches of AD assessment using simulations are presented. Coming from a test-case-centered

[1]Martin Sigl (martin.ms.sigl@bmw.de), Binnert Prins (binnert-mark.prins@bmw.de), Christoph Schütz (christoph.schuetz@bmw.de), Sebastian Wagner (sebastian.ws.wagner@bmw.de), are with BMW AG, Petuelring 130, 80788 München, Germany.

[2] Frederik Schulte (f.schulte@tudelft.nl) is with the Faculty of Mechanical, Maritime and Materials Engineering (3mE), Mekelweg 2, 2628 CD Delft

[3]Daniel Watzenig (daniel.watzenig@tugraz.at) is with the Institute of Automation and Control at Graz University of Technology, Inffeldgasse 21/B/I, 8010 Graz, Austria and with the Virtual Vehicle Research GmbH, Inffeldgasse 21A, 8010 Graz, Austria

2023 IEEE International Automated Vehicle Validation Conference (IAVVC) | 979-8-3503-2253-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/IAVVC57316.2023.10328115

automotive development process [8], data-driven simulation (DDS) tries to recreate real-world test cases or so called scenarios [3], [9], [10]. With a simulation framework which produces valid results, the scenario can be assessed and variations (e.g. regarding the environment or AD function) can be applied. Thus, the scenario space can be explored and covered, finally leading to an assessment of AD in the expected scenario space.

With stochastic simulations based on the Monte Carlo method (Monte-Carlo Simulation, MCS), parameters which describe real-world scenarios are used. Test cases for the simulations are created which try to recreate real-world test conditions within the virtual test bench by sampling over the distributions of those parameters. Both approaches focus on different aspects of the assessment process:

The strength of DDSs is the precise recreation of real-world test cases. This is especially useful for critical or potential dangerous scenarios, that can then be repeated and explored in-depth in a scalable and safe environment with adjustable conditions. Due to its focus on the exact recreation of real-world scenarios, it lacks the ability to naturally explore the scenario space in an unrestricted way [11]. If a scenario is varied too much, its difference to the underlying verified real-world scenario gets too big and therefore may become a unverifiable scenario without realistic foundation.

Stochastic MCSs have their strength in generating large quantities of traffic observations and scenarios independent from a direct representation in a real-world test case. After extracting environment-defining parameters and their distributions, no more variance in the underlying real-world data is needed. This facilitates capturing rare events without the need to first record them in real-world test cases.

Both approaches focus on generating virtual test cases from real world test cases or parameters. In contrast to using purely synthetic tests, these approaches facilitate the verification and validation process of simulations. DDS can be cross-verified by evaluating the real-world test case with the virtualized version of the test [3]. MCS can be verified by comparing the real-world parameters that define test spaces with the distribution of parameters in the virtual test space [12].

For the course of this work, the focus will lie on MCS due to their ability to generate much data with only a relatively small set of parameters compared to the large amount of real-world data that is needed for DDS.

B. Description and Classification of Traffic Quantification Metrics

In the past, many approaches evolved for quantifying traffic with regard to specific objectives. For example, road construction has been focusing on traffic rule based manipulation of traffic and traffic flow for many decades. For the validation if road planners reach their goal, they apply metrics to describe the traffic in a quantitative way [13]. Typical goals include increased safety and traffic flow, especially in bottleneck situations with many vehicles and restricted space. In the following, a classification of the many Traffic Quantification Metrics (TQMs) that exist nowadays will be performed.

TQMs can be assigned to either being *macroscopic* or *microscopic* [4], [14]. Generally speaking, microscopic measures can be linked to individual vehicles. This includes velocities, positions, relative distances, and also incorporates modeling desired speeds, intended destinations, and other parameters defining the individual behavior of traffic participants. Macroscopic TQMs aggregate individual measures and therefore lose their ability to describe individual behavior. Examples for macroscopic TQMs are traffic densities or traffic flows, mean velocities, or average travel times. Next to aggregating multiple traffic participants, macroscopic measures typically also cover a certain time span.

Also, TQMs can be divided into four major objectives which they are trying to fulfill [15], [16]. First, quantifying traffic can help to measure the safety of traffic. For example, calculating time-to-collision (TTC) values for all individual traffic participants gives a method to assess the danger that lies withing a traffic situation. Second, TQMs can also be used to implement traffic rules and evaluate their adherence [17]. In order to avoid dangerous situations, time-to-headway (THW) is used to provide means to comply to a safe style of driving. The THW TQM is easily verifiable by drivers and law enforcement and independent of variable traffic rules like speed limits. Third, after ensuring safe traffic, one of the main goals for road designers and city planners is generating a road network capable of handling the amount of expected traffic [14]. Therefore, TQMs are needed that describe traffic flow in order to provide means to detect or anticipate traffic congestion and maximum vehicle throughput of roads. Fourth, there exist TQMs aiming to describe the comfort of traffic participants. In general, microscopic values are considered to describe the comfort of vehicle passengers [18]–[20]. This includes for example measuring accelerations that feel uncomfortable or capturing movements that are associated with inducing motion sickness.

III. Methodology & Approach

The core idea of this work is to provide means for quantifying the change in traffic between different traffic observations. The primary criterion the chosen TQMs fulfill is their applicability to an arbitrary combination of real-world or simulated traffic observations. In the following, this feature is used to perform two things. First, we can cross-verify simulations that aim to recreate real-world traffic observations. For example, if a MCS extracts traffic-defining parameters from a real-world traffic observation, TQMs should deliver similar results when evaluating both the simulated traffic space and the real-world traffic space. Second, after assuming a valid simulation is available, the TQMs can be used to quantify the impact of automated vehicles. If comparing two different simulations, one recreating a real-world traffic observations with only human drivers, and another one that contains a certain amount of AVs, it is possible to quantify how these two traffic scenarios differ.

To achieve the above, several steps are necessary which will be described in detail in the next sections. With two traffic observations as start, dividing them up into sub-spaces is the first step. These sub-spaces can be similar to logical scenarios or more generally clusters in the scenario space. Then, TQMs are applied to these sub-spaces and deliver a quantitative description of the sub-space. The quantified sub-spaces then can be evaluated between the two traffic observation, delivering the actual difference between those two traffic observations regarding the specific sub-space. In the following, these steps will be explored in detail.

A. Traffic Observations

A traffic observation contains data of traffic objects over time, for example velocity and positions of vehicles. In general there are no diverging limitations in either using real-world, virtual or synthetic data as source for traffic observations. As long as they fulfill the requirements to provide all the necessary data to calculate TQMs, all data sources can be used to measure the difference of the traffic space.

In order to test the ability of the presented approach to help validating simulation, we use a real-world data set as well as traffic observations coming from simulation frameworks. For real-world traffic observations the highD data set was used, in which drones collect video data by hovering over German highway sections [21]. For the simulated traffic observations we use the MCS framework openPASS [22]. In general, there exist more influences on traffic than only driver behavior. To account for this, only traffic observations with similar environmental conditions are compared in this paper. This especially includes weather and road conditions, and therefore also visibility ranges. There may be additional influences that are not within the scope of this paper, e.g. time of day, day of the week, or season.

B. Traffic Sub-spaces

It is important to divide traffic observations into sub-spaces, because otherwise rare events are underrepresented in data. For example, if an AV acts very differently in a specific situation like being overtaken from a vehicle driving right of the AV, the effect would vanish in the vast amount of more common driving situations. The definitions of these sub-spaces can be chosen freely, as long as they contain only scenarios which qualify as similar. These can either be logical scenarios such as overtaking, emergency braking, lane changes, or defined in a more abstract way like clusters in the scenario space [23].

For this work, lane changes are used as exemplary traffic sub-space, as they are easily detectable and lateral movement is involved, with the latter being rare on highways. Lane changes are defined by a change of an object's lane ID. The time points where the lateral movement exceeds a certain threshold marks the beginning and end of the lane change. The duration of the lane change is limited to 7.5 seconds to avoid inconsistencies that might arise with aborted or double lane changes.

Although it is possible to aggregate sub-spaces back to a global traffic space, this is not part of this paper. It is sufficient to solely look onto sub-spaces, it provides information in which sub-space the AV causes major changes in the traffic space instead of describing minor impact on a global level. Even macroscopic TQMs such as traffic flow don't have to be aggregated over sub-spaces, because the evaluation of sub-spaces and their weight regarding the total traffic observation suffice to extrapolate the overall effect on the global traffic space.

C. Traffic Quantification Metrics

In this paper, several different TQMs are applied to the traffic sub-spaces. Since the focus is on traffic observations which always span a certain amount of time, the presented TQMs are always either macroscopic or aggregate microscopic TQMs over time.

Table I presents an overview over the used TQMs in this paper. The velocity v, acceleration a, and jerk j are aggregated over all object traces of a traffic observation. Also, both lateral and longitudinal characteristics are covered, using the subscripts $lat.$ and $long.$. The Time-To-Headway (THW) metrics measures the time that is needed for a vehicle to reach the position of its predecessing vehicle. THW is capped at 20 seconds because its unreasonable to assume an influence of a preceding vehicle on the following vehicle with a high distance [17]. Time-to-collision (TTC) calculates the time that will pass until a collision with the predecessing vehicle occurs [24], if the velocity of the following vehicle is higher than the velocity of the leading vehicle. The Quickness Q serves as description metric for lane change speed and also for measuring comfort, since lateral movements can directly be linked to the subjective comfort experience driving in vehicles. It divides the lateral velocity $v_{lat.}$ by the traveled lateral distance $d_{lat.}$.

D. TQM evaluation

In principle, it is expected that a TQM is an objective measure for the similarity. If two similar traffic observations are compared, the applied TQMs should distributed similarly. Analogously, two highly deviating observations should yield the TQMs to indicate different key values. In the following, we call this the *homogeneity* (similarity) and *heterogeneity* (difference) property.

To compare the TQM distributions between two traffic observations $TO1, TO2$, the following null hypothesis is used:

$$H_0 : \mu(TQM_{TO1}) = \mu(TQM_{TO2}) \tag{1}$$

The null hypothesis that the means μ of the TQM results of two distinct traffic observations are equal. If the null hypothesis is accepted, homogeneity between the two traffic observations is assumed, if the null hypothesis is rejected, heterogeneity is assumed. Regarding the TQMs to all be distributions, the Welch's test is applied to test for the null hypothesis. The p-value threshold of 0.03 is chosen as acceptance criteria to reject or accept the null hypothesis. The null hypothesis test is performed for all TQMs, and between all traffic observations. If a TQM fails to meet the expected homogeneity and heterogeneity properties, it can be considered to be unsuited for describing changes in traffic spaces.

TABLE I: Overview of the proposed TQMs that will be used throughout the approach. Velocity, acceleration, and jerk are further split between lateral and longitudinal values. The presented TQMs are applied to all vehicles or vehicle traces of an individual traffic observation if not stated otherwise.

Symbol	Description	Definition	Unit
v	Distribution of vehicle velocities	dx/dt	m/s
a	Distribution of vehicle accelerations	d^2x/dt^2	m/s^2
j	Distribution of vehicle jerks [25]	d^3x/dt^3	m/s^3
THW	Distribution of time gaps between vehicles and their predecessor [26]	$\min(\text{t-gap}, 20)$ for t-gap > 0 with t-gap $:= \lvert pos_{lead.} - pos_{foll.} \rvert / v_{foll.}$	s
TTC	Distribution of time-to-collision values between vehicles and their predecessor [24]	$\min(\text{ttc}, 20)$ for ttc > 0 with ttc $:= \lvert pos_{lead.} - pos_{foll.} \rvert / \Delta v$ with $\Delta v = (v_{lead.} - v_{foll.})$ if $v_{lead.} < v_{foll.}$ else 0	s
Q	Distribution of the quickness of lane changes [25]	$v_{\text{lateral}}/d_{\text{lateral}}$ (during lane change)	$1/s$
TF	Traffic Flow in vehicles per hour and per lane	$\text{num(vehicles)}/(\text{hour} \cdot \text{lane})$	$1/s$

To test for homogeneity and heterogeneity, data sets have to be established that contain either similar or dissimilar overall traffic situations. To have a most realistic data base, the highD data set is used to provide test data for the homogeneity and heterogeneity analysis. First, basic parameters like the road architecture are fixed to contain only sets with three lanes per road. Second, clusters in the highD traffic observations are investigated with the k-means method for the parameters mean velocity, mean THW, and traffic flow (TF). Two clusters were identified that share strong inner-cluster similarity but are also distinct to the other cluster. With $\pm$ depicting standard deviation, the cluster properties are:

$$
\begin{aligned}
C_{1,\text{highD}} : & \mu\text{TF} = 3338 \pm 178[\text{veh/hour}] \\
& \mu\text{THW} = 0.65 \pm 0.25[\frac{1}{s}] \\
& \mu v_{long.} = 25.7 \pm 2.9[\frac{m}{s}] \\
C_{2,\text{highD}} : & \mu\text{TF} = 1740 \pm 235[\text{veh/hour}] \\
& \mu\text{THW} = 0.55 \pm 0.23[\frac{1}{s}] \\
& \mu v_{long.} = 25.75 \pm 3.6[\frac{m}{s}]
\end{aligned}
$$

The data retrieved from openPASS is also placed into two clusters with the condition $C_{1,\text{openPASS}} \cong C_{1,\text{HighD}}$. The clusters are not perfectly similar which is caused by the limitations of the internal methods of recreating real-world traffic in openPASS. An arbitrary combination of the cluster parameters could not be achieved in openPASS. Regarding the second cluster, the goal is to achieve heterogeneity between the clusters. Therefore, the following properties are true:

$$
\begin{aligned}
C_{1,\text{openPASS}} : & \mu\text{TF} = 3068 \pm 220[\text{veh/hour}] \\
& \mu\text{THW} = 1.3 \pm 0.05[\frac{1}{s}] \\
& \mu v_{long.} = 32.78 \pm 0.68[\frac{m}{s}] \\
C_{2,\text{openPASS}} : & \mu\text{TF} = 1691 \pm 400[\text{veh/hour}] \\
& \mu\text{THW} = 1.28 \pm 0.14[\frac{1}{s}] \\
& \mu v_{long.} = 33.37 \pm 1.03[\frac{m}{s}]
\end{aligned}
$$

In a final step to evaluate TQMs, a correlation analysis is performed in order to find TQMs that can be omitted in the overall TSA, due to mutual redundancy.

E. AD Function

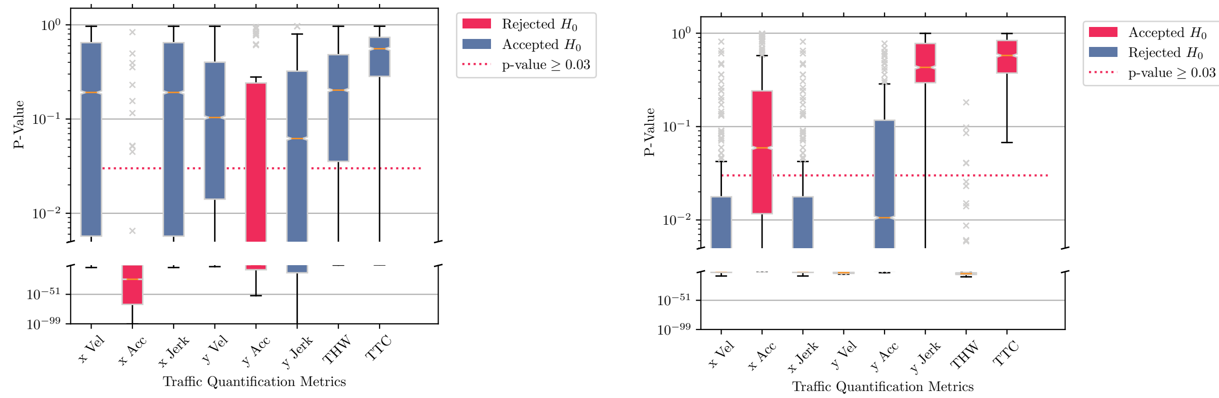
In principle, the presented approach is able to show any kind of traffic space changes, independent of their origin. Therefore, it is applicable to provide means to quantify the impact of AVs and AD on traffic. This can also be extended to cover other influences on traffic like weather, time of day, and so on.
Since this paper uses a MCS, no specific real-world test cases to base the simulation on exist. Therefore, also the human behavior must be generated stochastically. This is performed with the Stochastic Cognitive Driver Model (SCM) [12]. As AD function which controls the AVs in the simulation, a simple lane keeping assistant is used. For the proof of concept, an elaborated, complex AD function is not necessary.

IV. Results

This chapter presents results after applying the methodology to traffic observations from the real world and simulations. It starts with TQM evaluation, showing the homogeneity and heterogeneity analysis, as well as a correlation analysis. Afterwards, the evaluated TQMs demonstrate the necessity of dividing traffic observation into sub-spaces. Additionally, the evaluated TQMs show their use in validating simulation. The last subsection applies the TQMs to simulated traffic observations with different ratios of AVs in the traffic, showing the AV's influence on the traffic observation.
As mentioned earlier, several global parameters have to be fixed in order to make traffic observations comparable. The results in this section are limited to road sections with three lanes. The boxplots in this section are standard boxplots with 1.5 inter-quartile range and showing outliers as well as the median. Their y-axes are logarithmic to account for boxes whose range is close to zero.

A. TQM Evaluation

In this subsection, the presented TQMs are evaluated regarding their ability to detect changes in different traffic observations. In fig. 2, the homogeneity analysis is depicted to check if traces within a cluster in a traffic observation are similar, and a heterogeneity analysis to check if traces from distinct classes



(a) **Homogeneity** between traces within $C_{1,\text{HighD}}$ (b) **Heterogeneity** of traces between $C_{1,\text{openPASS}}$ and $C_{2,\text{openPASS}}$

Fig. 2: Homogeneity and heterogeneity tests for traffic observation from either real-world HighD data or synthetic openPASS data.

are diverging. The p-value level line at 0.03 shows if the null hypothesis is accepted or rejected. The boxes for the individual TQMs are colored in a way to show if the expected result is achieved. In the case of the homogeneity plots, this means blue boxes if the null hypothesis is accepted and the median of p-values is below the p-value threshold, and red if it is above. For heterogeneity, dissimilarity is expected, therefore the colors are swapped.

In fig. 2a, the homogeneity between traces within the HighD cluster $C_{1,HighD}$ is shown and most of the TQMs are similar. There are two notable results in the homogeneity analysis within clusters. First, lateral and longitudinal acceleration seem to be diverse within a real-world traffic observation, probably indicating a wide variety of driver behavior. Second, a similar homogeneity analysis shows the simulated openPASS traffic observation is much more homogeneous than compared to the real world traffic. Again, this can also be explained by a driver behavior, but in this case with a much more narrow bandwidth.

Fig. 2b shows the TQM's ability of capturing heterogeneous traffic by comparing the two openPASS clusters $C_{1,\text{openPASS}}$ and $C_{2,\text{openPASS}}$. While the comparison of the two HighD clusters rejects all null hypotheses (not depicted), the comparison between the two openPASS clusters does not show as many rejected null hypotheses. This can again be explained by less diverse driver behavior models in the simulation.

In summary, the homogeneity and heterogeneity analysis prove the ability of TQMs to differ between similar and dissimilar traffic observations. The evaluation of TQMs also provides insights like showing more diverse traffic in real-world than within simulations, and that not every aspect of driver behavior is influenced equally by traffic conditions.

As final step in the evaluation process, the correlation between the TQMs is analysed (see table II) with the Spearman rank coefficient correlation, to avoid using TQMs that express similar or equal results as other TQMs. The highest correlation within this table is between TF and THW, causing TF to be dropped as a TQM in the used TQM set. There are no other strong correlations within the TQM list, therefore the best way to proceed is to keep all of the remaining TQMs. Moderate correlations such as between a_{long} and v_{lat} do not necessarily contain useful information, especially since lateral movement on highways is generally very rare and restricted. The quickness Q does not seem to be correlated to any of the other TQMs, indicating significant knowledge gain by calculating and using this TQM. In conclusion, there are some medium correlations between individual TQMs, but they are not strong enough to fully omit one or the other. TF and THW are the only exception with very strong correlation. Since THW also contains information about driving safety, TF is being dropped in the remainder of this paper.

B. Sub-Space validation

Applying homogeneity and heterogeneity analysis as presented in the last subsection to different sub-spaces delivers further insight on the need of dividing traffic observations into smaller parts. When testing the whole traffic observation and the sub-space for homogeneity, individual TQMs may indicate significant dissimilarities. Fig. 3 applies the homogeneity analysis for sub-space lane changes and the whole traffic observation. It shows similarity of some of the TQMs, while mostly the longitudinal variables of the observed vehicles differ significantly from the original whole traffic observation. A similar effect occurs when comparing the same sub-spaces of different traffic observations. Most of the TQMs still fail the heterogeneity test, indicating that driver behavior in the lane change sub-space share similarities over otherwise very dissimilar traffic observations.

In conclusion, dividing traffic observations into sub-spaces is essential. The presented approach can be used to evaluate if sub-spaces such as rare events in traffic observations show a significant deviation and therefore are necessary to investigate in order to detect critical changes in overall traffic.

	v_{long}	a_{long}	THW	TTC	Q	v_{lat}	a_{lat}	TF
v_{long}	1.00	0.10	0.43	-0.44	0.23	0.14	-0.16	-0.46
a_{long}	0.10	1.00	0.07	0.59	0.22	0.66	0.01	-0.03
THW	0.43	0.07	1.00	-0.55	0.32	0.12	-0.46	-0.94
TTC	-0.44	0.59	-0.55	1.00	-0.13	0.34	0.28	0.62
Q	0.23	0.22	0.32	-0.13	1.00	0.26	-0.29	-0.32
v_{lat}	0.14	0.66	0.12	0.34	0.26	1.00	0.01	-0.09
a_{lat}	-0.16	0.01	-0.46	0.28	-0.29	0.01	1.00	0.38
TF	-0.46	-0.03	-0.94	0.62	-0.32	-0.09	0.38	1.00

TABLE II: Spearman's rank coefficient correlation table of a selection of TQMs.

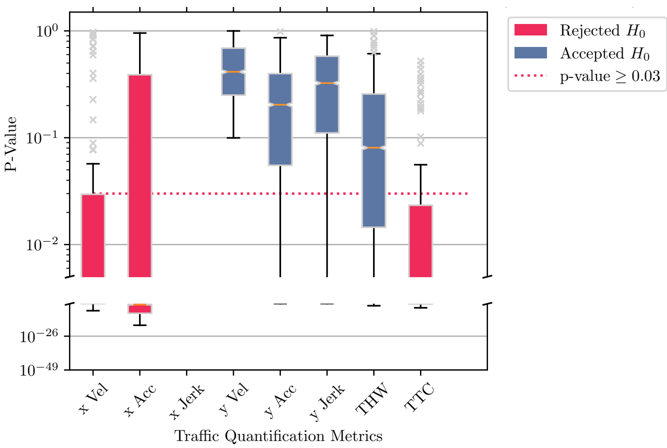


Fig. 3: The homogeneity test compares the whole traffic observation of $C_{1,\text{HighD}}$ with its lane change sub-space. Some of the sub-space's TQMs fail the similarity test, since lane changes do not share all the same properties as the whole traffic observation.

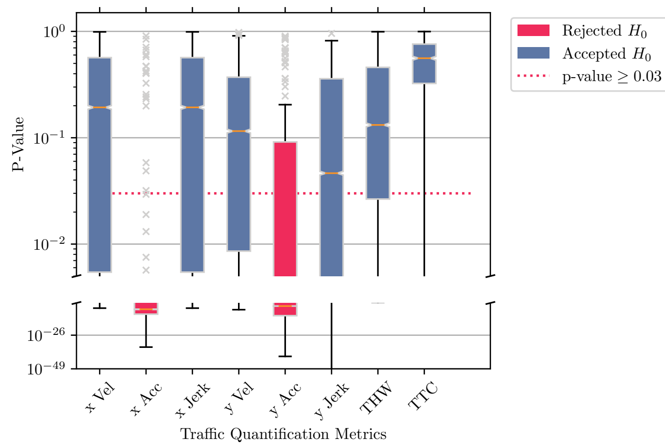


Fig. 4: Homogeneity between $C_{1,HighD}$ and $C_{1,openPASS}$, expressing the quality of the simulation's recreation of a real-world traffic observation.

C. Simulation cross-verification

The presented approach can be used as a tool that helps to validate simulation. In principle, it can be applied when trying to virtually recreate a real-world traffic observation. A homogeneity test can provide insight which aspects of a virtual twin of a traffic observation meet the requirements and which ones have to be improved. Under the assumption that all necessary values to parameterize a simulation can be extracted from a real-world traffic observation and the simulation framework being able to provide all tools for realistic traffic simulation, both the real-world and the simulated traffic should not differ. The openPASS cluster $C_{1,openPASS}$ uses parameterization directly from the HighD cluster $C_{1,HighD}$. Fig. 4 shows the result of the homogeneity analysis between the real world and the re-simulated traffic. Not all TQMs fulfill the expected similarity between the two traffic observations. In order to improve the simulation framework one could focus on the missed homogeneity criteria and look into the modeling of acceleration models of the virtual vehicles. Another goal to optimize the simulation is to increase homogeneity by applying measures that make the accepted H_0 in blue more narrow or move their median p-value closer to 1.

While this methodology does not yield an absolute evaluation about the validity of a simulation framework, it can provide insight which traffic sub-spaces cannot be simulated sufficiently. Also, single TQMs are able to show shortcomings, e.g. a vehicle model that generates unrealistic vehicle accelerations throughout a simulated traffic observation. Additionally, measures to improve simulation tools can be evaluated qualitatively and therefore speed up the development of simulation frameworks.

D. Impact of AVs

The described process is able to provide expressive TQMs and valid simulation frameworks. This allows to use this methodology to evaluate the impact of AVs to quantify the change that will come in hybrid traffic (compare fig. 1). After applying the previously described steps, AD function developers can be sure to use a quantification metric that is expressive and able to capture significant changes, and that the simulation framework is sufficiently precise and valid.

V. Conclusion and Outlook

The aim of this work is to provide a methodology to holistically describe changes between traffic observations. A literature research resulted in a set of TQMs that are able to quantify different aspects of real-world and simulated traffic. First, these TQMs have been evaluated and it has been proven that they are sensitive to changes in traffic and can correctly identify similarities and dissimilarities. For this task, the homogeneity and heterogeneity tests of Section III-D are used.

Second, the overall methodology's ability to provide insight to simulation framework validity is presented, again using the homogeneity and heterogeneity test approach. Finally, with expressive TQMs and valid simulation, the impact of AVs is quantified. This provides the necessary tools to quantitatively analyse the hybrid traffic space as depicted in fig. 1.

The goal to develop a method to holistically quantify the impact of AVs on traffic has been achieved. Furthermore, the presented methodology is able to comprise any additional TQM since it provides means to evaluate the any metric regarding its ability to capture changes in traffic spaces.

Unfortunately, it is not possible to say how small changes in traffic can be to still be detected. One possible solution is to use methods to divide the whole traffic observation into sub-spaces and to investigate them. Only if the traffic space as a whole, divided into sub-spaces, is considered, new scenarios can be detected and quantified. Next to this traffic space decomposition, a promising approach to evaluate the results of this paper is to investigate if and how the TQMs as objective traffic description metrics are subjectively experienced by human traffic participants. For example, an averagely attentive driver might not perceive changes in the mean velocity while traveling a long distance, but might react negatively to slight changes in the acceleration behavior of the AV and even develop nausea.

References

[1] W. Wachenfeld and H. Winner, "Die Freigabe des autonomen Fahrens," in *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*, M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, Eds. Springer Berlin Heidelberg, 2015, pp. 439–464.

[2] T. Helmer, L. Wang, K. Kompass, and R. Kates, "Safety Performance Assessment of Assisted and Automated Driving by Virtual Experiments: Stochastic Microscopic Traffic Simulation as Knowledge Synthesis," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 2019–2023.

[3] K. Groh, S. Wagner, T. Kuehbeck, and A. Knoll, "Simulation and Its Contribution to Evaluate Highly Automated Driving Functions," *SAE International Journal of Advances and Current Practices in Mobility*, vol. 1, no. 2019-01-0140, pp. 539–549, 2019.

[4] K. Li and P. Ioannou, "Modeling of traffic flow of automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 2, pp. 99–113, 2004.

[5] R. Hoogendoorn, B. van Arerm, and S. Hoogendoom. (2014) Automated Driving, Traffic Flow Efficiency, and Human Factors: Literature Review.

[6] S. C. Calvert, W. J. Schakel, and J. W. C. van Lint, "Will Automated Vehicles Negatively Impact Traffic Flow?" *Journal of Advanced Transportation*, vol. 2017, 2017.

[7] T. Ard, L. Guo, R. A. Dollar, A. Fayazi, N. Goulet, Y. Jia, B. Ayalew, and A. Vahidi, "Energy and flow effects of optimal automated driving in mixed traffic: Vehicle-in-the-loop experimental results," *Transportation Research Part C: Emerging Technologies*, vol. 130.

[8] H. Winner, K. Lemmer, T. Form, and J. Mazzega, "Pegasus—first steps for the safe introduction of automated driving," in *Road Vehicle Automation 5*. Springer, 2019, pp. 185–195.

[9] K. Groh, T. Kuehbeck, B. Fleischmann, M. Schiementz, and C. C. Chibelushi, "Towards a Scenario-Based Assessment Method for Highly Automated Driving Functions," 2017.

[10] P. Minnerup, T. Kessler, and A. Knoll, "Collecting Simulation Scenarios by Analyzing Physical Test Drives," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015.

[11] S. Wagner, K. Groh, T. Kühbeck, and A. Knoll, "Towards Cross-Verification and Use of Simulation in the Assessment of Automated Driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019.

[12] A. Fries and F. Fahrenkrog, "Validation and Verification of the Stochastic Cognitive Driver Model," 2021.

[13] R. J. Smeed and G. T. Bennett, "Research on Road Safety and Traffic Flow. Road Engineering Division." *The Institution of Civil Engineers Engineering Division Papers*, 2015.

[14] M. S. Grewal and H. J. Payne, "Identification of Parameters in a Freeway Traffic Model," vol. SMC-6, no. 3, pp. 176–185, 1976.

[15] S. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS Research*, vol. 41, no. 4, pp. 153–163, 2017.

[16] C. Hyden, "The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique," *Bulletin Lund Institute of Technology, Department*, no. 70, 1987.

[17] K. Vogel, "A comparison of headway and time to collision as safety indicators," *Accident Analysis & Prevention*, vol. 35, no. 3, pp. 427–433, 2003.

[18] S. P. Hess, "Automobile Riding-Comfort," *SAE Transactions*, vol. 19, pp. 250–272, 1924.

[19] I. Bae, J. Moon, J. Jhung, H. Suk, T. Kim, H. Park, J. Cha, J. Kim, D. Kim, and S. Kim. (2020) Self-Driving like a Human driver instead of a Robocar: Personalized comfortable driving experience for autonomous vehicles.

[20] A. L. Higgins, *The Transition Spiral and Its Introduction to Railway Curves with Field Exercises in Construction and Alignment*. New York, Van Nostrand Company, 1922.

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

[22] C. Guindon. OpenPASS Working Group — The Eclipse Foundation. OpenPASS. [Online]. Available: https://openpass.eclipse.org/

[23] J. Kerber, S. Wagner, K. Groh, D. Notz, T. Kühbeck, D. Watzenig, and A. Knoll, "Clustering of the Scenario Space for the Assessment of Automated Driving," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 578–583.

[24] J. C. Hayward, "Near-Miss Determination though Use of a Scale of Danger," *Pennsylvania State University - University Park*, 1972.

[25] H. Bellem, T. Schönenberg, J. F. Krems, and M. Schrauf, "Objective metrics of comfort: Developing a driving style for highly automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 41, pp. 45–54, 2016.

[26] L. Evans, *Traffic safety and the driver*. Science Serving Society, 1991.