2021 IEEE Intelligent Transportation Systems Conference (ITSC)
Indianapolis, USA. September 19-21, 2021

2021 IEEE International Intelligent Transportation Systems Conference (ITSC) | 978-1-7281-9142-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITSC48978.2021.9564743

Calibration and Assessment of Urban Microscopic Traffic Simulation as an Environment for Testing of Automated Driving

Marcel Langer, Michael Harth, Lena Preitschaft, Ronald Kates and Klaus Bogenberger

***Abstract* — The development and testing of automated driving systems currently pose a challenge for the automotive industry. Since real world tests of automated driving systems are time consuming, often infeasible and potentially dangerous, especially the evaluation of urban automated driving must begin with simulation in a realistic virtual environment. Therefore, it has been proposed to employ traffic simulation to provide an environment of traffic participants for an automated driving system to interact with. For the achievement of this goal, this work presents two contributions. Firstly, we provide an optimal parameterization of existing driver behavior models for urban traffic based on real-world vehicle trajectory data gathered in Ingolstadt, Germany. Secondly, the result of the calibration is assessed for potential improvements and central design aspects of future models for environment simulation. Using a simulation of the city in SUMO, the best calibration result is achieved with the Krauss model with slightly modified parameters compared to the default parameters. Deviations between the calibrated simulation and the real-world data are evaluated for individual roads. Assessment of residuals between real-world data and calibrated simulation suggests that behavior could be dependent on characteristics not yet represented in the simulation models, such as number of lanes or speed limit.**

I. Introduction

Traffic simulation can be employed to analyze both traffic efficiency and traffic safety of automated vehicles. To evaluate safety, traffic simulation is used to create an environment in which the influence of various traffic participants on the performance of an automated driving function is analyzed. In contrast, simulating the influence of an automated driving function on surrounding traffic is used for an assessment of its impact on traffic efficiency. While various tools for traffic simulation are available, the implemented models are designed mostly for evaluations of traffic efficiency. In order to make the existing tools and models suitable for the application as an environment simulation, they must not only represent traffic flow characteristics but also individual human behavior accurately.

There are multiple aspects of human behavior that are currently being researched including the longitudinal and lateral behavior within the lane, lane changing behavior, behavior at intersections and various traffic violations and human errors. This publication focusses on longitudinal behavior of the vehicles as well as decision making at signalized intersections. Because of the importance of vehicle speeds for urban road safety, an emphasis is placed on an accurate representation of the velocities in the simulation. As a basis for the evaluation of the simulated behavior, a simulation of the city of Ingolstadt in SUMO [1] is compared to real-world vehicle trajectory data. In Fig. 1, a part of the downtown area of the simulation is depicted. To evaluate the quality of the simulation, a fitness function is defined. This fitness function is utilized to find an optimal parameterization for driver behavior in Ingolstadt. Furthermore, the fitness function provides a basis for the assessment of residuals between simulation and reality.

The rest of this paper is structured as follows. First, the state of the art in calibration of driver behavior in microscopic traffic simulations is reviewed. The most promising approaches are used as a foundation for the design of the method introduced in section III. The simulation setup, fitness function as well as the parameters selected for calibration are described in section IV. The results of the calibration for different behavior models are presented in section V together with a discussion of these results in the context of an environment simulation.

II. State of the Art

A review of the literature on calibration of behavior models in microscopic traffic simulation reveals a multitude of approaches towards the challenge at hand. An overview of the

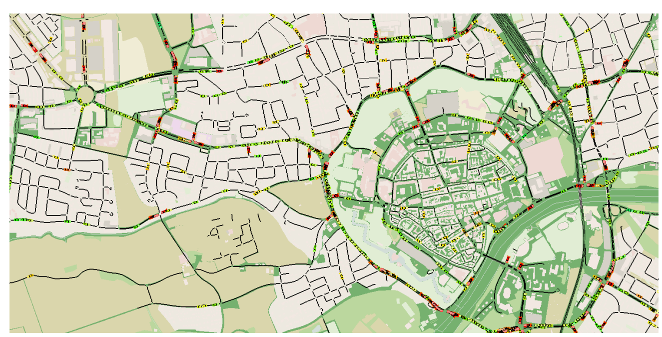

Figure 1: Simulation of Ingolstadt in SUMO

This work is supported by the German Federal Ministry of Transport and Digital Infrastructure (BMVI) within the *Automated and Connected Driving* funding program under grant no. 16AVF2145A (SAVe).

Marcel Langer, Michael Harth and Lena Preitschaft are with AUDI AG, 85045 Ingolstadt, Germany (+49 841 89 986125; e-mail: marcel.langer@audi.de; michael.harth@audi.de, lena.preitschaft@audi.de).

Ronald Kates is with REK Consulting, 83624 Otterfing, Germany, (e-mail: ronald.kates@t-online.de).

Klaus Bogenberger is with the Chair of Traffic Engineering and Control, Technical University of Munich, 80333 Munich, Germany (e-mail: klaus.bogenberger@tum.de).

existing literature is provided by Yu and Fan [2]. The approaches differ in 1. the algorithms applied for optimization, 2. whether entire simulations or standalone models are calibrated, 3. the type and amount of data used and 4. the street type selected for calibration. The existing methods are similar in the way, that they compare real-world traffic data to simulated data and then iterate through various parameter combinations in order to find an optimal parameterization. For the optimization process the Nelder-Mead Simplex (downhill simplex) Method [3], Simultaneous Perturbation for Stochastic Approximation Algorithms (SPSA) [4] or Genetic Algorithms (GA) [5]–[8] are applied. Abdalhaq and Baker [9] compare multiple methods and suggest that Particle Swarm Optimization (PSO) and Tabu Search (TS) are best suited for the optimization. However, the authors also conclude that a GA performs well and point out that it is the most commonly used and therefore most thoroughly tested algorithm for the calibration process. Yu and Fan [2] compare the GA to TS and achieve similar results with both algorithms. They also suggest the use of a "warm" start for the optimization, which refers to utilizing a previously obtained solution as a starting point for further optimization runs.

Considering the objective of the optimization procedure, the existing literature can be split into two groups. In the first approach described in [2], [6], [9]–[11], a traffic simulation is calibrated based on local observations of the traffic state. The parameters of the simulation are optimized to reproduce one or multiple characteristics of traffic flow, traffic density, travel times and vehicle speeds observed in reality. The data required for calibration is collected for specific edges of the simulation network by locally installed cameras or loop detectors.

In the second approach, it is not the aim to calibrate a traffic simulation but a standalone behavior model instead. Car-following-models are parameterized to reproduce real-world car following situations as accurately as possible. Kesting and Treiber [5] calibrate the Intelligent Driver Model (IDM) as well as the Velocity Difference Model based on trajectory data from a single vehicle. In a later publication, Treiber and Kesting [12] calibrate behavior models using trajectory data collected by individual vehicles as well as trajectories from local observations in the Next Generation SIMulation (NGSIM) datasets [13]. Treiber and Kesting evaluate the influence of data quality, sampling rate and smoothing on the calibration procedure. Individual trajectories from the NGSIM datasets are also used by Rahman et al. [14] and Li et al. [15] to find optimal parameters for different car-following models. A different application of the NGSIM dataset is provided by Kim et al. [16], who follow the first approach and calculate traffic flow, traffic density and vehicle speeds from the trajectories as a baseline for their calibration. A large amount of trajectory data collected by Volvo employees is available to Pourabdollah et al. [7]. From these trajectories Pourabdollah et al. select 200 car following situations of up to two minutes and utilize these for the parameterization of three car following models with the aim of evaluating the influence of driving behavior on energy consumption.

The existing approaches focus either on the calibration of the traffic characteristics inside a simulation or of a standalone behavior model. However, it is the aim of this contribution to find a calibration procedure for a traffic simulation that reproduces the real-world traffic state, while accurately modelling driver behavior simultaneously. This is made possible by using an accurate traffic simulation as a basis and taking into account the influence of the behavior parameterization on the simulation. Furthermore, although three of the provided publications [8], [9], [11] consider urban road networks, most of the existing work focusses on the parameterization of highway traffic. While simulation of highway traffic mainly requires calibration of car following and lane changing models, in an urban environment there are further aspects of driver behavior that must be accurately represented. Especially the driver behavior at intersections has a large influence on urban traffic and must therefore be modelled correctly.

In addition, utilizing traffic simulation as an environment simulation once again increases the requirements towards the accuracy of the behavior models. According to the Highway Capacity Manual (HCM) [17], there are various influences on driver behavior that are not yet modelled in state-of-the-art simulation tools. For example, the HCM states that a driver's desired velocity will be up to 7.3 km/h faster on roads with multiple lanes or up to 5.8 km/h faster dependent on the lateral clearance beside the road. While these influences may not be crucial for evaluations of traffic flow, they must be accounted for in an environment simulation, as they may result in unique interactions or cause faulty predictions in an automated driving system. Modelling these influences requires taking into account different road or intersection characteristics, which would again need to be based on exhaustive data. Although e.g. Pourabdollah et al. [7] appear to have access to data that would be required to evaluate driver behavior across a large spectrum of roads and intersections, no evaluations of the data with regard to these influences are provided.

III. METHOD

Based on the existing state of the art, we propose a method for the calibration of microscopic driver behavior within a traffic simulation of the city of Ingolstadt in SUMO. This calibration is based on a comparison between simulated and real-world driving data. The simulated behavior is adjusted based on parameters selected from both the car-following and the junction model in SUMO. Most previous publications focus on a calibration of either traffic demand, traffic regulation, car-following behavior or junction behavior. We provide a method for calibrating both car-following behavior and junction behavior simultaneously inside a simulation in which traffic regulation and traffic demand are set up beforehand. This is necessary because of the strong interdependency between the individual behavior and the traffic state in a large simulation network. As an example, we consider the vehicle speeds on an arbitrary road in reality and on its corresponding edge in the simulation network. A lower speed of the vehicles in simulation compared to reality could indicate an underestimation of the desired speed in free-flowing traffic. However, lower speeds could also result from an interfering leading vehicle or a bound traffic state originating from a junction ahead. This consideration holds true vice-versa for the case of overrepresented speeds in the simulation. In order to calibrate the vehicle speeds and accelerations in the simulation according to observed values in reality, driver behavior must be accurately reproduced in combination with the traffic state in simulation. This requires a calibration of the behavior both on the edges as well as within the junctions of the simulation network.

To achieve this, vehicle movement within the simulation is compared to evaluations from real-world vehicle trajectories gathered throughout the city. The data are obtained from data logging devices installed into vehicles in personal use throughout the city of Ingolstadt. The position, heading, speed and acceleration are measured with great accuracy by the sensors of the vehicles and collected together with a timestamp directly from the vehicle data bus. Since all vehicles are equipped with an Adaptive Cruise Control (ACC), information regarding traffic ahead is obtained by the built-in object detection of the vehicles. The result of the sensor data fusion from camera and radar sensors, which contains the relative speed and position of a leading vehicle in car-following situations, is also collected directly from the vehicle data bus. After pre-processing, around 40 hours of vehicle trajectories are available for calibration within the simulation network of the city of Ingolstadt. All measurements are processed with a time step of 0.2 s, which is sufficiently accurate according to the results of Treiber and Kesting [12].

For evaluation of the vehicle data, the trajectories are matched to the simulation map based on the vehicle's location and heading at each data point. The trajectories are split up into segments (sub-trajectories) for each edge of the simulation network, which enables edge-based evaluations of the data. For each segment, the following evaluations are calculated:

- Vehicle speeds per time step [m/s]
- Vehicle accelerations per time step [m/s²]
- Time gap between ego and leading vehicle per time step [s]
- Time-to-traverse the edge per segment [s]

Since the traffic state may vary according to different hours of different days of the week, the real-world segments of each edge are split into groups by their timestamp. Firstly, the segments are split up in six groups by the day of the week into weekdays (Mo – Thu), Fridays, Saturdays, Sundays, public holidays and weekdays during school holidays. Within each day, the segments are separated again into eight time groups each containing the segments from three hours of the day. This results in a total of 48 time groups. For the calibration of the simulation, the traffic demand and traffic light control is set up for one three-hour time group of one day and all real-world segments of the time group are selected. From the aforementioned evaluations, distributions are calculated for each edge for which a sufficient number of segments is available in a selected time group. The same distributions are obtained from the SUMO simulation using outputs from the FCD-device (Floating Car Data). The quality of the simulation is assessed by a fitness function based on an edge-wise comparison between the real-world distributions and the distributions resulting from simulation. Following the state of the art, a GA is applied for the optimization of the simulation parameters. The described calibration method is depicted in Fig. 2.

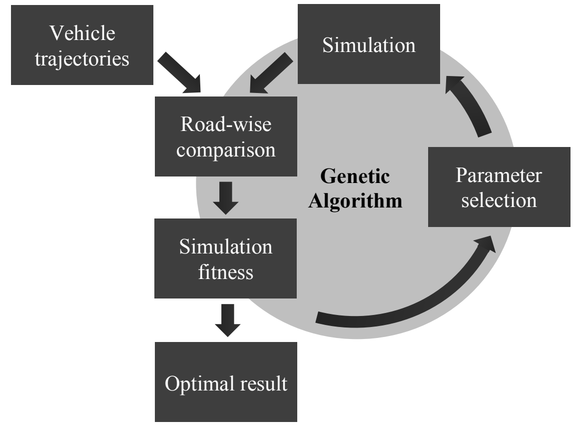

Figure 2: A GA is applied to calculate an optimal value for selected parameters in simulation.

For the optimization, the following parameters are selected from the car-following and the junction model in SUMO:

- Minimum gap in standstill – d_{min} [m]
- Time gap – τ [s]
- Driver imperfection – σ [-]
- Mean desired speed relative to speed limit – $\bar{v}_{rel}$ [-]
- Standard deviation of desired speed relative to speed limit – v_{sd} [-]
- Junction drive-after-yellow-time – t_y [s]

A detailed explanation of the chosen parameters and their influence within the simulation is provided in the following section. While the parameters τ, $\bar{v}_{rel}$ and v_{sd} have direct counterparts in the distributions of the vehicle speeds and time-gaps, the other parameters have an indirect influence on the distributions, which are compared in the fitness calculation. All of the parameters selected for optimization influence the capacity of edges or junction-maneuvers in the simulation, which reflects in the distributions of vehicle speeds and accelerations as well as the time-to-traverse. The acceleration itself is not set directly via an individual parameter, but is instead influenced by the parameters defining the car-following behavior and the desired speed.

It is the aim of the calibration procedure to find one optimal set of the selected parameters for the entire simulation. Although it is theoretically possible to calculate an optimal set of parameters for each junction and edge individually, this is not advisable, since it would lead to overfitting. Inaccuracies in the simulation setup or individualities in the vehicle trajectories of a specific edge strongly influence the fitness calculation of the edge. By calculating one set of parameters for the entire simulation, an overfitting of the models to these particularities is prevented.

IV. IMPLEMENTATION

In this section, the simulation setup, the fitness function used for optimization as well as the simulation parameters are described in detail.

A. Simulation setup

Before the calibration procedure, the simulation is set up with an accurate road network, traffic demand and traffic regulation. The road network is imported directly from OpenStreetMap using SUMO and manually improved based on aerial images. The challenging representation of the traffic demand is carried out using both statistical data as well as traffic counts. The statistical data provided by the city of Ingolstadt and Audi as the major employer in the city contain detailed information on the distribution of population, workplaces as well as shopping and leisure opportunities across the city. Traffic counts are obtained from induction loops at all traffic lights in the city, and the traffic demand in the simulation is tuned to match both the statistical data as well as the measured traffic volumes. Heavy-vehicle traffic is also added to the simulation based on manual counts provided by the city of Ingolstadt. However, the parameter optimization is carried out only for passenger vehicles and not for heavy-vehicles. Together with the induction loop data, the traffic light states are recorded every second for each traffic signal group at each signalized intersection in the city. These data are then applied for a reconstruction of the traffic light cycles within the simulation. The result of the described process is a simulation in which the road network, traffic regulation and traffic demand are modeled as close to reality as possible. This simulation with 12.000 edges, 5600 nodes and 120 traffic lights is utilized for all further evaluations of driver behavior.

B. Fitness function

The quality of the simulation is calculated from a comparison of four distributions from real-world vehicle trajectories with the corresponding distributions obtained from vehicles in SUMO for each edge in the network. For the comparison of two individual distributions, the Kolmogorov-Smirnov-Test [18] is applied. The Kolmogorov-Smirnov-Test estimates the similarity between two distributions by calculating the maximum difference between the cumulative distributions. The results of the Kolmogorov-Smirnov-Test are always in the interval [0, 1]. A test result of zero is achieved with two identical distributions, which in this application would indicate a perfect match between simulation and reality in the observed metric. In contrast, if there is no overlap between the two distributions, the test results in the value one. Fig. 3 shows the result of this test for the distributions of the time gaps for one edge. These time gap distributions are cut off above 3 s to minimize the influence of sensor errors in the real-world data.

For each edge and for each of the distributions of speed, acceleration, time headway and time-to-traverse, the similarity between simulation and reality is calculated using the Kolmogorov-Smirnov-Test. Before the test results are aggregated to assess the overall quality of the simulation, two weights are introduced into the calculation. Firstly, the result of the Kolmogorov-Smirnov-Test of the speed distributions is weighted twice as high as all other comparisons. This is done because the distribution of the vehicle speed has both a strong influence on the calibration and is obtained with maximum accuracy from the real-world trajectories. As described above, the comparison of the vehicle speeds is the main optimization factor for the two parameters $\overline{v}_{rel}$ and v_{sd}. However, all the other parameters also indirectly influence the speed

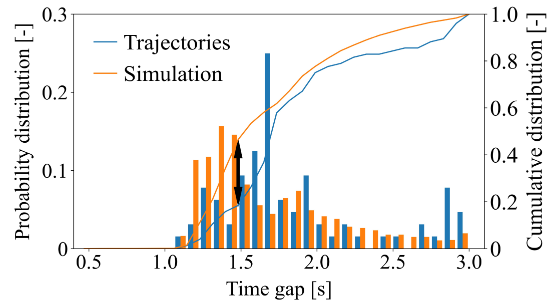

Figure 3: Exemplary result of the Kolmogorov-Smirnov-Test applied on the time gaps in real-world data and simulation indicated by the black arrow.

distribution in simulation, as they adjust traffic flow or intersection capacity, which can affect queueing and waiting times. The fitness of each edge f_e is calculated as the weighted sum of the comparisons of the four distributions of speed $f_{e,s}$, acceleration $f_{e,a}$, time gap $f_{e,tg}$ and time-to-traverse $f_{e,c}$, as shown in (1).

$$f_e = 2f_{e,s} + f_{e,a} + f_{e,tg} + f_{e,c} \tag{1}$$

The second weight is added in the aggregation of the edges. Since a different number of sub-trajectories is collected for each individual edge, the real-world calibration reference is not equally reliable for each edge. Although edges with less than three sub-trajectories are not factored into the fitness calculation, edges with a large amount of reference data should still be weighted higher than edges with few sub-trajectories. The square root of the number of sub-trajectories per edge c_e is therefore factored into the fitness calculation. The overall fitness of the simulation with a given set of parameters is calculated as the weighted sum of all edges e with more than three sub-trajectories following (2).

$$f = \frac{1}{f_{max}} \sum_e \sqrt{c_e} * f_e \tag{2}$$

The overall fitness of the simulation is divided by the maximum possible fitness value f_{max}, which would indicate all Kolmogorov-Smirnov-Tests of all edges resulting in one. The final fitness of the simulation is therefore always in the interval [0, 1] with zero being the desired optimal result.

C. Simulation parameters

A systematic evaluation of the parameters of car-following and junction models in SUMO is conducted to create a list of potentially relevant parameters for the calibration. The two commonly used car-following models Krauss [19] and IDM [20] are considered in the calibration procedure. In order to select the most influential parameters that are likely to influence the optimization criteria, each parameter in this list is varied individually within a defined interval while all other parameters are kept at their default value. The largest difference between the simulation runs is evaluated via the previously described fitness function. All temporal parameters are varied in the optimization in steps of 0.2 s, which results from the equivalent simulation step length. A comparison between the fitness results as well as the default value of each parameter selected for optimization is depicted in Table I.

The six parameters that are adjusted in the calibration procedure are selected for their strong influence on the fitness

TABLE I. INFLUENCE OF SELECTED PARAMETERS

Parameter	Default	Interval	Step	Change in fitness
d_{min}	2.5 m	[0.1, 3.9] m	0.2 m	0.028
τ	1 s	[0.2, 3] s	0.2 s	0.084
σ	0.5	[0, 1]	0.05	0.038
$\bar{v}_{rel}$	1	[0.9, 1.3]	0.05	0.064
v_{sd}	0.1	[0, 0.3]	0.05	0.06
t_y	0 s	[0, 3] s	0.2 s	0.006

of the simulation. The minimum gap d_{min} defines the distance a vehicle will maintain to a leading vehicle in standstill. Similarly, the time gap τ sets the minimum time headway a vehicle will keep in car-following situations. The desired speeds of the vehicles are expressed relative to the speed limit by the ratio v_{rel}. The individual ratios per vehicle are sampled from a Gaussian distribution with mean value $\bar{v}_{rel}$ and standard deviation v_{sd}. To avoid extreme behavior in the simulation, SUMO cuts off the speed distribution at [0.6, 1.4] relative to the speed limit independent of the parameter input. When using the Krauss model, further stochastic driving behavior is introduced into the simulation using the driver imperfection parameter σ. This parameter causes random variations of the vehicle speeds in car-following and free-flowing situations. The acceleration ability of vehicles in simulation is also evaluated as a possible parameter. However it is not selected for optimization, since a variation in the interval [2, 4] m/s² shows only a minor influence on the fitness result. The behavior of the vehicles at signalized intersections is adjusted by setting the drive-after-yellow-time t_y. This parameter sets the duration for which vehicles will pass a traffic light in simulation after it switches from green to yellow. This parameter does not cause vehicles to violate red lights.

While Table I shows that each selected parameter has a measureable influence on the simulation, it is noticeable, that the drive-after-yellow-time t_y has by far the smallest impact on the simulation fitness. The default value $t_y = 0$ causes vehicles in SUMO to stop at yellow lights whenever possible. However, the vehicles will always pass traffic lights at yellow if they are traveling too fast to be able to stop. Due to this, the parameter t_y has the strongest influence on the simulation at intersection maneuvers with a short green time or when vehicles are queueing for a maneuver. Since in this particular situation, t_y has a crucial influence on the capacity of a maneuver (e.g. right turn, left turn, straight) of a junction in simulation, it is selected for optimization despite the low change in fitness observed during the variation. The maximum value of the optimization interval of t_y is set to 3 s since this is the longest yellow time implemented in the simulation.

While the drive-after-yellow-time is calibrated by its influence on speed and travel time on edges leading up to signalized intersections, the same cannot be done for other junction parameters. SUMO allows the user to vary the way vehicles cross or merge with prioritized traffic by using the junction-minor-time-gap and the impatience parameters. However, only the drive-after-yellow-time can be calibrated with the available data for two reasons. Firstly, an observation of the Ingolstadt simulation shows queues forming only at signalized but not at priority-controlled intersections. Secondly, the majority of vehicle trajectories is collected for large roads connected by signal-controlled intersections. The influence of the behavior in priority situations on the fitness function and thus on the calibration process is therefore negligible in this sample. This also shows in variations of the priority parameters, which lead to changes in the fitness calculation below 0.002.

V. RESULTS AND DISCUSSION

For the calibration, the Ingolstadt simulation is set up with a simulation step of 0.2 s, which is equal to the sample rate of the trajectory data. Separating the trajectory data into the defined time groups reveals that the largest amount of data is available between 9 am – 12 am on weekdays. Out of the total 40 h of driving data, 10 h are collected within this time interval and matched to 1500 edges in the simulation network. After filtering edges with a minimum length of 30 m and three or more sub-trajectories, a total of 8 h and 303 edges are used for calibration. The demand in the simulation is set up for an average weekday between 9 am – 12 am with 45.000 vehicle routes. By equipping only some of the SUMO vehicles with a FCD-device, the simulated FCD are generated only for 15% of the simulated vehicles. This provides a representative set of trajectories from the simulation, while keeping both output data and evaluation time to a minimum. All simulation parameters that are not explicitly mentioned are left set to their default value in SUMO. The optimization of the selected simulation parameters is carried out with a GA in 15 iterations and a population size of 20. In each generation, 40 % of genomes are selected as parent genomes, mutation probability is set to 30 % and crossover probability is set to 50 %. In order to investigate the global and local stability of the optimization result, multiple optimization runs with different random initial populations as well as a sensitivity analysis are performed. All optimization runs converge to an identical optimal solution for each car-following model described in the following section.

A. Optimization results

The calibration is applied to find optimal parameters for the two commonly used car-following models Krauss and IDM individually. The calibration results as well as the fitness values of the default settings for each model are displayed in Table II. As highlighted in Table II, the best calibration result with the given setup is achieved with the Krauss model. This aligns with the assessment of Krajzewicz et al. [21], who suggest that the Krauss model works best in SUMO because the simulation software is developed and tested based on it. Parameters that are not part of the optimization, e.g. from the lane-changing model, or built-in assumptions in SUMO may influence the calibration in favor of the Krauss model.

The fitness of the calibrated simulation is improved by 1.8 % compared to a simulation using the default settings of SUMO. The two parameters d_{min} and τ, which define the distance to a leading vehicle, are left set to the default values n the optimized simulation. An improvement of the simulation is achieved by setting t_y to 1.8 s, which increases the number of vehicles that are able to pass a signal-controlled intersection per traffic light phase. Furthermore, the driver imperfection σ is increased to 0.55.

TABLE II. CALIBRATION RESULTS

	Krauss	**IDM**
d_{min} [m]	2.5	2.0
τ [s]	1.0	0.8
σ [-]	0.55	n.a.
$\bar{v}_{rel}$ [-]	1.15	1.15
v_{sd} [-]	0.25	0.2
t_y [s]	1.8	1.8
Fitness (Default)	**0.411** (0.429)	0.421 (0.472)

The largest difference between the default and the optimized parameters is obtained for the desired speeds of vehicles in the simulation. The desired speeds follow a Gaussian distribution with mean value 1.15 and standard deviation 0.25 relative to the speed limit. As previously mentioned, the speed distribution is cut off at [0.6, 1.4] relative to the speed limit to avoid extreme behavior in the simulation. This parameterization sets the desired speed of more than 2/3 of the vehicles in simulation to a value above the speed limit. However, simulated vehicles often drive slower due to the imperfection σ or are prevented from traveling at their desired speeds by traffic regulation and slower leading vehicles. With regard to the following discussion of the vehicle speeds, it must be pointed out, that around 3/4 of the roads for which a sufficient number of sub-trajectories is collected are large, primary roads with two lanes per direction. This suggests that the speed calibration is biased towards increased desired speeds.

B. Assessment and implications for environment simulation

Although the calibration method returns an improved parameterization compared to the SUMO default parameters, the calibrated simulation differs in multiple ways from the real-world data. It is clear that the simulation can never perfectly reproduce the real-world trajectories due to the individuality of human behavior as well as variations between the traffic states of the days on which the trajectories were collected. The fitness of the optimal parameter set is calculated to 0.411, which indicates that the cumulative distributions of simulation and reality differ on average by this value.

Some of this final deviation results from the amount of data available for calibration: On roads where few real-world sub-trajectories are collected, the calculated distributions will differ due to statistical variance in the real-world data. This influence shows in the average fitness 0.35 of roads with more than 20 sub-trajectories, which is significantly lower than the overall fitness of the optimization. Furthermore, due to the properties of the Kolmogorov-Smirnov statistic, differences will always remain when comparing two statistics with low sample size. This is particularly influential in the comparison of the distributions of the time-to-traverse, since these distributions are made up of only one value per sub-trajectory.

However, further residuals between real-world data and simulation result from selecting one set of parameters for the entire simulation. Since the parameters are optimized for the entire simulation and not for each individual edge, the calibration result defines an average behavior over all observed roads.

Taking a closer look at the vehicle speeds shows that this averaging is another reason for deviations between simulation and reality in the final calibration. The HCM [17] specifies that a driver's desired speed depends on environmental factors such as e.g. the width of the road, the number of lanes or the layout of the shoulder of the road. In Fig. 4, the share of speeding events in the real-world trajectories and in the simulation is evaluated for roads with different characteristics. Since the absolute values cannot be published due to data privacy, the provided normalized values are calculated following (3) and (4).

$$\beta(s, n) = \frac{m_{speeding}(s, n)}{m_{total}(s, n)} \tag{3}$$

$$\nu = \frac{\beta(s, n)}{\beta_0} \quad \text{with} \quad \beta_0 = \beta_{Traj}(50, 1) \tag{4}$$

First, the share of observed speeding events β is calculated for each group of roads with speed limit s and number of lanes per direction n. As shown in (3), this is done by dividing the number of measurements above the speed limit $m_{speeding}$ by the total number of measurements m_{total} for all s and n. The normalized values ν plotted in Fig. 4 are calculated for real-world trajectories and simulation following (4). The share of all speeding events per dataset and per road characteristic is compared to the share, which is observed in the real-world trajectories on one-lane roads with a speed limit of 50 km/h.

Comparing the real-world measurements between roads with different characteristics shows, that speeding events occur about 3.5 times more often on roads with two lanes or with a speed limit of 30 km/h compared to one-lane roads with a speed limit of 50 km/h. Since this difference is not modelled within the simulation, the calibration results in an overrepresentation of the vehicle speeds on 50 km/h one-lane edges and an underrepresentation on the other edges. Although more speeding events are measured in the simulation on two-lane edges compared to one-lane edges, this results from the possibility to overtake slower leading vehicles and does not fully represent the difference observed in reality. Similar distributions are obtained when comparing the average speeds of the edges instead of the speeding events.

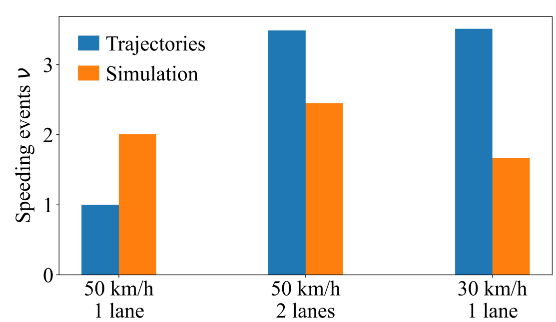

Figure 4: Comparison between the share of measurements above the speed limit in the real-world trajectories and in simulation. Measurements are grouped by the number of lanes and the speed limit of the edge. The values are provided relative to the real-world measurements for one-lane roads with a speed limit of 50 km/h.

Fig. 4 visualizes the limitation of the employed simulation setup with regard to the described influences. The fitness of the simulation cannot be improved in one location without reducing the accuracy in another. While SUMO allows calibrating the speeding behavior on different roads explicitly, this would require large amounts of data to be collected for each road to prevent overfitting to single trajectories. Instead, we suggest systematically modelling the influence of the road layout and road periphery in upcoming behavior models for better generalizability.

Although no further evaluations can be presented at this point, visual inspection of the available data suggests that various other aspects of driver behavior are influenced by the traffic and road environment. E.g., the time gap in car-following situations is likely to be lower at signalized intersections because of drivers trying to cross the intersection before the end of the green phase. Other influencing factors on the desired speeds could include the proximity to sidewalks or vehicles parked at the side of the road.

VI. Conclusion and Outlook

This work presents a contribution towards creating a traffic environment for testing of automated driving based on an existing microscopic traffic simulation framework. An optimal set of parameters for a simulation in SUMO is obtained by comparing the simulation to real-world vehicle trajectories. While the presented results are generated for the morning hours of an average weekday, the method could be applied in future to create multiple optimal parameter sets for different times of day and for varying environmental conditions. It is expected, that driver behavior will be different during daytime or nighttime as well as in rainy or sunny weather.

Although the accuracy of the simulation is improved compared to the default setup, residuals between the simulation and the real-world trajectories remain. These residuals can likely be reduced further by integrating more parameters into the calibration procedure or collecting more real-world data for the calibration. However, an evaluation of the simulation brings forth influences that are not represented in existing behavior models. Existing behavior models are designed for traffic flow simulation. The focus of these models lies in efficiently simulating a large number of vehicles with sufficient accuracy based on few input variables. For the application in an environment simulation, the requirements shift towards simulating fewer vehicles with greater accuracy and increased individuality in the microscopic behavior. Furthermore, virtual testing of automated driving functions requires detailed models of the road geometry and road periphery. While this type of environmental information is rarely available in a simulation set-up for an evaluation of traffic flow, it could be taken into account in the design of behavior models for the environment simulation.

References

[1] P. A. Lopez *et al.*, “Microscopic Traffic Simulation using SUMO,” in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2018-Novem, 2018, pp. 2575–2582.

[2] M. Yu and W. Fan, “Calibration of microscopic traffic simulation models using metaheuristic algorithms,” in *International Journal of Transportation Science and Technology*, vol. 6, no. 1, 2017, pp. 63–77.

[3] E. Brockfeld and P. Wagner, “Calibration and validation of microscopic traffic flow models,” in *Traffic and Granular Flow*, no. 750, 2004, pp. 67–72.

[4] A. Paz, V. Molano, and C. Gaviria, “Calibration of CORSIM models considering all model parameters simultaneously,” in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, no. September, 2012, pp. 1417–1422.

[5] A. Kesting and M. Treiber, “Calibrating car-following models by using trajectory data: methodological study,” in *Transportation Research Record*, no. 2088, 2008, pp. 148–156.

[6] S. Chiappone, O. Giuffrè, A. Granà, R. Mauro, and A. Sferlazza, “Traffic simulation models calibration using speed-density relationship: An automated procedure based on genetic algorithm,” in *Expert Systems with Applications*, vol. 44, 2016, pp. 147–155.

[7] M. Pourabdollah, E. Bjarkvik, F. Furer, B. Lindenberg, and K. Burgdorf, “Calibration and evaluation of car following models using real-world driving data,” in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2017.

[8] L. Zhang, D. Sun, L. Wang, and H. Zhang, “Parameter Calibration of Microscopic Traffic Simulation Based on Floating Car Data,” in *Data Driven Control and Leanring Systems Conference*, 2019, pp. 1219–1224.

[9] B. K. Abdalhaq and M. A. Baker, “Using Meta Heuristic Algorithms to Improve Traffic Simulation,” in *Journal of Algorithms and Optimization*, vol. 2, no. 4, 2014, pp. 110–128.

[10] A. Paz, V. Molano, E. Martinez, C. Gaviria, and C. Arteaga, “Calibration of traffic flow models using a memetic algorithm,” in *Transportation Research Part C: Emerging Technologies*, vol. 55, no. 4, 2015, pp. 432–443.

[11] A. Paz, K. Shrestha, C. Arteaga, and D. Baker, “Calibration of Microscopic Traffic Flow Simulation Models considering Subsets of Links and Parameters,” in *Journal of Advanced Transportation*, vol. 2020, 2020.

[12] M. Treiber and A. Kesting, “Microscopic Calibration and Validation of Car-Following Models – A Systematic Approach,” in *Procedia - Social and Behavioral Sciences*, vol. 80, 2013, pp. 922–939.

[13] J. Halkias and J. Colyar, “NGSIM Overview,” 2006. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

[14] M. Rahman, M. Chowdhury, T. Khan, and P. Bhavsar, “Improving the Efficacy of Car-Following Models With a New Stochastic Parameter Estimation and Calibration Method,” in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, 2015, pp. 2687–2699.

[15] L. Li, X. M. Chen, and L. Zhang, “A global optimization algorithm for trajectory data based car-following model calibration,” in *Transportation Research Part C: Emerging Technologies*, vol. 68, 2016, pp. 311–332.

[16] J. Kim, J. H. Kim, G. Lee, H. J. Shin, and J. H. Park, “Microscopic Traffic Simulation Calibration Level for Reliable Estimation of Vehicle Emissions,” in *Journal of Advanced Transportation*, vol. 2020, 2020.

[17] National Research Council, “Highway Capacity Manual,” 2000.

[18] Frank J. Massey Jr., “The Kolmogorov-Smirnov Test for Goodness of Fit,” in *Journal of the American Statistical Association*, 1951, pp. 68–78.

[19] S. Krauß, “Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics.” German Aerospace Center, Dissertation, 1998.

[20] M. Treiber, A. Hennecke, and D. Helbing, “Congested traffic states in empirical observations and microscopic simulations,” in *Physical Review E*, vol. 62, no. 2, 2000, pp. 1805–1824.

[21] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, “SUMO - Recent Development and Applications of {SUMO - Simulation of Urban MObility},” in *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3, 2012, pp. 128–138.