Fuzzy Modeling and Inference for Physics-aware Road Vehicle Driver Behavior Model Calibration

Cristian Axenie[a](cristian.axenie@huawei.com),
Wolfgang Scherr[b](wolfgang.scherr@moventes.net), Alexander Wieder[a](alexander.wieder@huawei.com),
Anibal Siguenza Torres[a,c](anibal.siguenza.torres@huawei.com),
Zhuoxiao Meng[a,c](zhuoxiao.meng@huawei.com), Xiaorui Du[a,c](xiaorui.du@huawei.com),
Paolo Sottovia[a](paolo.sottovia@huawei.com), Daniele Foroni[a](daniele.foroni@huawei.com),
Margherita Grossi[a](margherita.grossi@huawei.com), Stefano Bortoli[a](stefano.bortoli@huawei.com),
Götz Brasche[a](goetz.brasche@huawei.com)

[a] Huawei Munich Research Center, Riesstraße 25, 80992 Munich, Germany
[b] MOVENTES GmbH, Egghölzlistrasse 69, 3006 Bern, Switzerland
[c] Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany

Corresponding Author:
Cristian Axenie
Intelligent Cloud Technologies Laboratory, Huawei Munich Research Center, Riesstraße 25, 80992 Munich, Germany
Phone: +49(0) 174 321 1689
Email: cristian.axenie@huawei.com

Fuzzy Modeling and Inference for Physics-aware Road Vehicle Driver Behavior Model Calibration

Cristian Axenie[a,*], Wolfgang Scherr[b], Alexander Wieder[a], Anibal Siguenza Torres[a,c], Zhuoxiao Meng[a,c], Xiaorui Du[a,c], Paolo Sottovia[a], Daniele Foroni[a], Margherita Grossi[a], Stefano Bortoli[a], Götz Brasche[a]

[a]Intelligent Cloud Technologies Laboratory, Huawei Munich Research Center, Riesstraße 25, 80992 Munich, Germany
[b]MOVENTES GmbH, Egghölzlistrasse 69, 3006 Bern, Switzerland
[c]Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany

Abstract

Traffic simulation is a crucial tool for testing and evaluating hypotheses in both the design of autonomous driving systems and the optimization of urban transportation, where modeling and calibrating driver behavior is key. In this article, we introduce a novel approach to calibrating driver models by combining fuzzy logic and inference to produce a plausible parametrization that, simultaneously, reproduces drivers' behavior peculiarities and road traffic characteristics. The calibration system uses solely vehicle trajectories that describe cars' movements through the road network, with a resolution of time and space enough to allow for determining the lane of the road used, speed, and acceleration at every second of the survey period. This data is processed by a lightweight modeling and inference system capable of interpreting the input data through the lens of domain knowledge and physics laws in order to learn plausible driver model parameters using fuzzy computation. The outcome is a set of parameters for mathematical models which emulate the social and physical features of the behavior of a driver-vehicle unit. Albeit this study limits itself to determining parameters for car-following and lane-change models, the systematic approach for calibration can be seen as a framework where also other driver model components can be calibrated (e.g. gap-acceptance). We evaluate our system on both synthetic and real data while comparing it with other approaches for calibration. We summarize our findings using a thorough analysis where we point out the most relevant aspects to consider when calibrating driver models.

Keywords: Driver Model, Calibration, Fuzzy Systems, Machine Learning, Traffic Simulation

1. Introduction

Drivers' behavior is very rich and diverse across the variety of road traffic situations encountered daily. Fundamentally, drivers' behavior describes a continuous negotiation of prevailing circumstances that guide decision-making. Computer-based road-traffic simulation is used to estimate the benefits of different actions and solutions before they are rolled out onto the road and, hence, support the decision on investments and technological improvements in traffic management and traffic control. An important part of traffic simulation is the so called "driver model", which is a mathematical formulation of vehicle movements and drivers' actions and reactions. Both the commercial and academic fronts pushed very good candidates for simulation software that support such driver models evaluation and testing, such as AIMSUN from Barceló

*Corresponding author.
Email addresses: cristian.axenie@huawei.com (Cristian Axenie), wolfgang.scherr@moventes.net (Wolfgang Scherr), alexander.wieder@huawei.com (Alexander Wieder), anibal.siguenza.torres@huawei.com (Anibal Siguenza Torres), zhuoxiao.meng@huawei.com (Zhuoxiao Meng), xiaorui.du@huawei.com (Xiaorui Du), paolo.sottovia@huawei.com (Paolo Sottovia), daniele.foroni@huawei.com (Daniele Foroni), margherita.grossi@huawei.com (Margherita Grossi), stefano.bortoli@huawei.com (Stefano Bortoli), goetz.brasche@huawei.com (Götz Brasche)

& Casas (2005), VISSIM from Lownes & Machemehl (2006), and PARAMICS from Cameron & Duncan (1996), on the commercial side, and SUMO fromKrajzewicz (2010) and CityMoS from Zehe et al. (2017) on the academic side. Above all, such tools enable capturing driver dynamics and its implications at micro- and macro-scale, a more and more relevant aspect for both the design of autonomous driving systems, such as the ones introduced in the work of Makridis et al. (2021); Hang et al. (2021) and Guo et al. (2020) and the optimization of urban mobility, described in the work of Genser et al. (2021); Massera Filho et al. (2017), and Guo et al. (2019).

But, models come in any size and any disguise! From a reductionist perspective, among driver behavior models, we distinguish input-output models, or mechanistic models, and psychological (motivation) models, following the taxonomy of Michon (1985). Practically, driver behavior models incorporate both mechanistic and motivational components. For instance, car-following models, such as the Intelligent Driver Model (IDM) - first introduced in the work of Treiber et al. (2000), the Gazis-Herman-Rothery (GHR) model of Chandler et al. (1958), and the Wiedemann Model - introduced in the seminal work of Wiedemann (1974), the Safe Distance Model of Yang et al. (2013), the Gipps' model of Gipps (1981), or the Optimal Velocity Model (OVM) of Sugiyama (1999) epitomize driver's "longitudinal dynamics" as a system of ordinary differential equations (ODE). Expanding to the "lateral dynamics" captured by lane-change models, such as the Gap Acceptance Models (GAM), introduced by Mahmassani & Sheffi (1981), Game Theory Models (GTM) developed by Liu et al. (2007), or the Minimizing Overall Braking Induced by Lane Changes (MOBIL) introduced in the excellent work of Kesting et al. (2007), the equated components induce a broad behavior spectrum, from purely egoistic to more cooperative driving.

Overall, parsimony is what describes such models and enables their adoption into practice, a place where a trade-off must be done in selecting the most relevant aspects of the behavior to be modeled. As previously mentioned, driver behavior models used in practice incorporate both mechanistic and motivational (psychological) components in order to capture and reproduce human driving behavior. For instance, the incentive to change lane and the safety constraint can be fused and described through an acceleration function of the underlying car-following model, such as described in the work of Markkula et al. (2012), which, in turn, allows for an efficient and parsimonious formulation of the lane-changing model, as postulated in the work of Kesting et al. (2007).

Reproducing such realistic driver behaviors is fundamental in both traffic simulations and autonomous driving systems - as described in the work of Xu et al. (2021). Yet, calibration, the essential procedure that brings simulation close to reality, is not a trivial task as demonstrated by Najmi et al. (2020). The computational burden of typical calibration approaches plays a key role in choosing a calibration approach, and usually forces traffic engineers to calibrate the driving behavior model components separately, as suggested in the great work of Treiber & Kesting (2013).

Also, trial-and-error methods and constrained optimization algorithms are currently the backbones of driver model calibration, which leaves room for error in the calibrated parameters in the face of intra- and inter-driver variability. This study addresses these challenges and suggests a structured approach for determining calibrated driver model parameters by fusing the physics of the driver model, expert knowledge, and machine learning.

To set the frame of our study, the next two sub-sections introduce the specifics and limitations of driver model calibration techniques (i.e. Section 1.1) and the need for physics-aware learning models when capturing driver dynamics in road traffic (i.e. Section 1.2).

1.1. Driver behavior models calibration

Independent of the underlying parameter inference approach, calibration procedures for driver models are typically comprised of five ingredients, as suggested in Sharma et al. (2019a): data acquired in multiple real traffic situations (i.e. trajectory), a calibration approach (global or local), a set of performance metrics (M), a set of metrics for the goodness-of-fit (G), and an optimization algorithm. A first, and crucial, question is what input data is available for calibration? Second-level resolution trajectory data were non-existent until ca. 2005 when the data of the NGSIM study of U.S. Department of Transportation Federal Highway Administration (2016) was published. But today, new technologies, such as intelligent video detectors, allow for more accurate calibrations.

Second, the discrepancy in calibration is, typically, measured by the goodness of fit of the simulated parameters to the real parameters and a series of calculated quantities by the metrics of performance in both real world and simulation. Every goodness of fit evaluation can be performed either globally (e.g. using Least-Squared Errors - extensively used by Treiber & Kesting (2013) and Berghaus et al. (2021)) or locally (e.g. using Maximum-Likelihood employed in the study of Treiber & Kesting (2013) and Alhariqi et al. (2021)), and since at its core the calibration is basically an optimization problem, the underlying algorithm aims to converge to a solution that is close to the global minimum of the goodness of fit metric G, as elaborated in the study of Berghaus et al. (2021), while obeying imposed constraints on parameters' values. A formulation of the calibration procedure can be synthetically described as:

$$\begin{aligned} & \min_{M_{sim}, M_{real}} \quad f(M_{real}, M_{sim}) \\ \text{with} \quad & f(.) = G(.),\ M_{sim} = F(\boldsymbol{x}|\boldsymbol{\beta}),\ M_{real} = F(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{\beta_{min}} \leq \boldsymbol{\beta} \leq \boldsymbol{\beta_{max}}, \end{aligned} \tag{1}$$

where: F is a mapping function of driver trajectory data $\boldsymbol{x}$ given driver model parameters $\boldsymbol{\beta}$ calculated after simulating the driver model to extract the simulated values of the performance metrics M_{sim}, M_{real} are the observed values of the performance metrics calculated by F from data $\boldsymbol{x}$ (i.e. speed, acceleration, headway), f is a function representing the goodness-of-fit G (i.e. realism of the simulation or closeness of M_{real} to M_{sim}), and $f(M_{real}, M_{sim})$ is the objective function to be optimized calculated from $\boldsymbol{x}$ - in principle a function describing the discrepancy between simulation and reality (i.e. $M_{sim} \approx M_{real}$).

Most approaches often treat the aforementioned driver model calibration process as a non-linear optimization problem, where the landscape of solutions is rather exotic. From heuristics which employ Scatter Search to initialize a Generalized Reduced Gradient Nonlinear Programming solvers used by Ciuffo et al. (2008), to Sensitivity Analysis of Analyses of Variance (SANOVA) developed by Punzo & Ciuffo (2009), up to Simultaneous Perturbation Stochastic Approximation (SPSA) employed by Ma et al. (2007), and finally computational intelligence and machine learning algorithms (e.g. Genetic Algorithms used in both the work of Kesting & Treiber (2008), and Cascan et al. (2019)).

To summarize, current approaches to calibrate driver models are based on time-consuming (optimization) approaches that use aggregated data (e.g. minute's aggregates of driver/traffic state – to form a spatio-temporal trajectory). But such aggregated data is typically insufficient to calibrate (simultaneously) the major components that comprise a simulation model (i.e. driver lane change, car-following, merge/yield in an urban traffic context). Unintentionally, and in search of the best fit, the modeler may modify model's parameters in a way that unrealistic behavior at the physical level is being produced involving, for instance, too many transitions from acceleration to deceleration or excessively high number of lane changes for a driver in a traffic context, hence a non-compliant behavior described and analyzed in the work of Sharma et al. (2019b). Additionally, without a proper spatio-temporal analysis tool and the proper validation metrics at this level, modelers may not have a way to measure the impact of the calibration process on system's behavior, as emphasized by Treiber & Kesting (2012). Of course, if there are no trajectory data of NGSIM quality, then try-and-error engineering might be the only viable approach.

Finally, despite their success in finding (numerically) optimal solutions, existing calibration approaches typically treat the "longitudinal dynamics" (i.e. car-following) and "lateral dynamics" (i.e. lane change, merge-yield) separately, hence discarding important interactions and relationships among variables describing driving dynamics. This is further enhanced by the strict focus of the calibration that excludes validation, where calibrated parameters that provide lower fitting errors do not necessarily better represent the actual driving behavior - as shown by Ossen & Hoogendoorn (2008). There is a strong need to fuse human driver factors in the calibration procedure together with the kinematic aspects of a driver's response - as proposed by Zheng (2014) - but unfortunately there is only limited research addressing such a perspective, for instance the study of Monteil et al. (2014) and Mirman (2019). The system proposed in this manuscript tries to fill in this gap and push forward a systematic approach for calibration which already demonstrates high potential.

1.2. Physics-aware learning models

All of the approaches discussed above focus on either the calibration of the ODE models for car-following (e.g. IDM) or the calibration of ODE models for lane-change (e.g. MOBIL). Such mechanistic models make assumptions about the dynamics of the driver within the lane (i.e. car-following) or between the lanes (i.e. lane-change) with respect to the other traffic participants. Advanced numerical methods and high-performance computing enable high-fidelity simulations of such calibrated mechanistic driver models to run at scale, as shown by the excellent works of Zehe et al. (2017) and Kučera & Chocholáč (2021) and resolve for microscopic performance goals - see the work of Treiber & Kesting (2013) and macroscopic performance goals - see the work of Sharma et al. (2019b).

As pointed in the systematic analysis of Treiber & Kesting (2013), most approaches for calibrating driver behavior models focus on the fit quality. Their authors argue that the calibration error (which varies depending on the optimization approach) reaches an insurmountable barrier (of around 20% as reported by Brockfeld et al. (2004)), resulting in a standstill in selecting the "optimal" model, as shown in the work of Zhu et al. (2018). Additionally, in order to capture corner cases in the spectrum of driver behaviors, the current approaches look at modification of the models themselves with adaptive features, as explored by Alhariqi et al. (2021) or the analysis of the influence of driver characteristics on driver behavior in extreme situations, described in the work of Berghaus et al. (2021).

Learning-based approaches offer an attractive alternative to optimization-based calibration approaches. Such approaches tackle the realistic reproduction of driver behaviors by learning the function F described in the problem formulation in Equation 1. In other words, such systems regress the underlying mapping from data $\boldsymbol{x}$ and model parameters $\boldsymbol{\beta}$ to a performance metric M. In this space he approaches are rather diverse. From Hidden Markov Models that employ driving signals to build a hierarchical models that capture driver behavior peculiarities, developed by Zaky et al. (2015) up to Particle Filter online parameter estimation, using the system of Bhattacharyya et al. (2020). From hybrid Game Theoretic Reinforcement Learning of interactive driving pattern extraction and control, developed by Li et al. (2019) up to a combination of Bayesian Optimization and Deep Recurrent Neural Networks for continuous driver model parameter calibration from the work of Yang et al. (2021) or driver-imitating Generative Adversarial Networks introduced by Kuefler et al. (2017).

A formulation of the learning approach to model calibration can be synthetically described as:

$$M_{sim} \approx M_{real} \iff F_{sim}(\boldsymbol{x}|\boldsymbol{\beta}) \approx F_{real}(\boldsymbol{x}) \tag{2}$$

where: F_{sim} is the learning model approximating the function F_{real} from which the metric of performance is calculated based on example training data $\boldsymbol{x}$. In order to relieve the data dependence of such regression models the goal is to add some level of "physical or domain insight" that automatically satisfies some of the physical invariants for better accuracy, faster training and improved generalization, as suggested by the seminal study of Karniadakis et al. (2021). More precisely, the robust incorporation of, for instance, an intermediate additive model $\rho(\boldsymbol{x})$ describing some physical constraint can increase accuracy without overfitting or additional data dependence, which turns Equation 2 into

$$F_{sim}(\boldsymbol{x}|\boldsymbol{\beta}) + \rho(\boldsymbol{x})\boldsymbol{\beta}_{\boldsymbol{\rho}} \approx F_{real}(\boldsymbol{x}) \tag{3}$$

Such an additive model is an improvement over the F_{sim} model because the physics-based coefficient β_ρ is determined simultaneously with the generic coefficients β, as shown in Wood (2008). While such an additive model is a step in the right direction, it is too reliant on the F_{sim} and the accuracy of the function ρ. To tackle this, our solution is to expand the physics-based model so that it considers domain insights as attributes of the intermediate model ρ. In order to achieve this, we consider introducing inductive biases through a fuzzy rule-based inference system, previously described by Magdalena (2015) that can steer the calibration process towards identifying physically consistent solutions for $F_{sim}(\boldsymbol{x}|\boldsymbol{\beta})$. Our approach is unique, as it combines fuzzy sets and fuzzy logic to represent knowledge about driver behaviors, as well as modeling the interactions and relationships among the variables describing the behavior by capturing and handling uncertainty and variance in the represented knowledge. Our framework maps prior knowledge in the form of propositional rules into a feed-forward formulation thus providing an explicit inductive bias.

2. Materials and methods

This section introduces the models, tools, and techniques we use in our approach to driver model calibration in road traffic simulations. We start with a short overview of driver models, formulate the calibration problem in the framework of fuzzy modeling and inference, and, finally, show how physics can be incorporated at every level of modeling and inference in the system in order to determine the computation of plausible parameter sets. The calibration framework we propose is generic and can accommodate different models for within-lane dynamics (i.e. car-following behavior) and between-lanes dynamics (i.e. lane-change behavior). In the following, we consider an instantiation of the framework which couples the IDM car-following model with the MOBIL lane-change model to describe the complete driver behavior in a highway scenario.

2.1. Driver models analysis

For our current instantiation of the calibration system, we use the Intelligent Driver Model (IDM) to capture the car-following dynamics and the MOBIL to capture the lane-change dynamics of a driver. The IDM was first introduced by Treiber et al. (2000) as a deterministic time-continuous model that describes the dynamics of each vehicle's positions and velocities within the lane. At the core it assumes that drivers control their vehicles to react to the changes in the state of preceding vehicles. It fundamentally aims to keep a balance between the need to maintain a safe space gap from the vehicle ahead (i.e. leader) and the goal to achieve "free flow" speed, as shown in Figure 1.

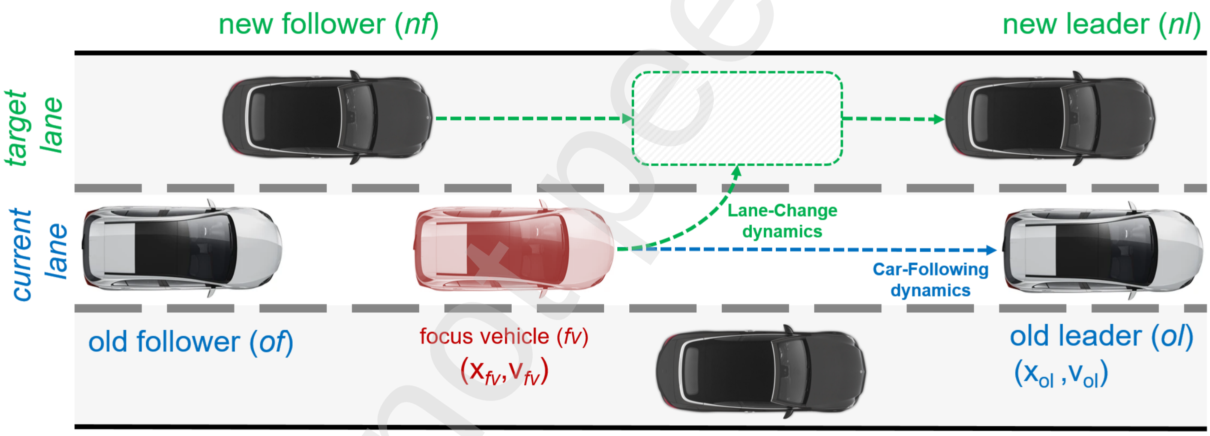

Figure 1: Car-following and lane-change dynamics.

2.1.1. Car-following dynamics with the Intelligent Driver Model

The basic IDM formulation of the model is given by the following set of ODE in pairs of leader and focus vehicle position $\boldsymbol{x} = (x_{ol}, x_{fv})$ and velocity $\boldsymbol{v} = (v_{ol}, v_{fv})$ over time t:

$$\begin{aligned} \dot{x}_{ol}(t) = v_{ol}(t), \dot{x}_{fv}(t) = v_{fv}(t) \\ \dot{v}_{fv}(t) = Acc(x_{fv}(t), v_{fv}(t), x_{ol}(t), v_{ol}(t)) \end{aligned} \tag{4}$$

where $Acc(x_{fv}(t), v_{fv}(t), x_{ol}(t), v_{ol}(t))$ is the IDM acceleration computed as

$$a\left(1 - \left(\frac{|v_{fv}|}{v_{fvfree}}\right)^{\delta}\right) - a\left(\frac{2\sqrt{ab}(s_0 + v_{fv}\tau) + v_{fv}(v_{fv} - v_{ol})}{2\sqrt{ab}(x_{ol} - x_{fv} - l)}\right)^2 \tag{5}$$

and a is the maximum acceleration of the focus vehicle, b is the comfortable deceleration, v_{free} is the desired focus vehicle velocity, τ is the desired headway, s_0 is the minimal net distance to leading vehicle, l is the focus vehicle length, δ is the acceleration exponent (i.e. describing how acceleration decreases when approaching v_{free}). A specific lane change, e.g., from the center lane to the median lane, as shown in Figure 1, depends on the leading and following vehicles on the present and the target lane, respectively. This maneuver is defined as a combined incentive to change lanes and safety constraints when changing lanes. More precisely, the safety criterion checks the possibility of performing a lane change by considering the effect on the upstream vehicle on the target lane.

2.1.2. Lane-change dynamics with MOBIL

In our current instantiation of the calibration framework, we consider the simultaneous calibration of the lane-change and car-following dynamics. The procedure takes into account both the incentive (i.e. utility) and the safety criteria of the accelerations for the MOBIL model on the current and the target new lanes, as calculated with the longitudinal model, i.e., the IDM in our case. To be more precise, the safety criterion is satisfied, if the calculated IDM braking deceleration b_{IDM} imposed on the new follower nf of the target lane after a possible lane-change does not exceed a certain deceleration threshold b_{safe}, in other words, the safety criterion in Equation 6 is fulfilled.

$$b_{IDM}(nf) > b_{safe} \tag{6}$$

In order to evaluate the incentive or utility criterion, we assess the focus vehicle advantage on the target lane, measured by the increased acceleration (or conversely reduced braking deceleration), against the disadvantage imposed to other drivers, again measured by the decrease acceleration or increased braking deceleration for these drivers. Since human drivers have the tendency to be egoistic, the disadvantage imposed on other drivers is weighted with a sub-unitary politeness factor p and regularized by a physical threshold acceleration $a_{threshold}$, resulting in the incentive criterion in Equation 7.

$$acc_{fv}^{target}(t) - acc_{fv}^{current}(t) > p(acc_{of}^{current} + acc_{nf}^{current}) - p(acc_{of}^{target} - acc_{fv}^{current}) + a_{threshold} \tag{7}$$

Expressing the lane-change utility and safety in terms of acceleration of the car-following dynamics enables for an efficient and compact (i.e. in the number of parameters) formulation, that allows for transferring properties of the car-following to the lane-change dynamics (i.e. parametrizing thresholds for acceleration, space gap, and time headway). This coupling is very important in calibration as it fuses strategic, tactical, and operational behaviors of the other cars on the same lane and on the target lane Wiedemann (1974) and captures both discretionary and mandatory lane-changes. While the analytical dependence between the IDM model and the MOBIL is obvious from Equations 5, 6, and 7, there is a more subtle temporal connection which is fundamental when simultaneously calibrating both the longitudinal (i.e. car-following model) and the lateral dynamics (i.e. lane-change model) of a driver behavior. A very important issues that our framework tackles is that very little research considered the combination of car-following and lane-change models that supports realistic behavioral representation of the driving task. Here, we can emphasize the work of Toledo et al. (2007), that captures drivers' planning capabilities and allows decisions to be based on anticipated future conditions, as also proposed by Wiedemann (1974).

A notable characteristic of the IDM model is that its parameters are orthogonal, with each parameter describing a particular aspect of driver behavior. Depending on the configuration of the trajectory data quantities (i.e. $[v(t), a(t), s(t), t(t)]$ - velocity, acceleration, space gap, and time gap, respectively) one can identify a certain regime/state the driver is in, as shown in Table 1. This systematic description provided by the very thorough analysis and study of Treiber & Kesting (2013), can be easily embedded in our system to facilitate the plausible calibration of the driver model.

Hence, one can extract rules to identify specific driving regimes which only depend on the modification of a single parameter, as depicted in Table 1. The interactions among its parameters describe also the number of traffic states/regimes in Table 1, in which a driver can transit during both car-following and lane-change while trying to maximize utility and keep safety. In our framework capturing such within lane and between lane transitions are vital for fully calibrating the driver model against a given trajectory. Plausible transitions in both space-velocity $(\Delta x, \Delta v)$ and space-velocity-time $(\Delta x, \Delta v, \Delta t)$ dimensions benefit from the fact that initial and final states are described by the orthogonal IDM dynamics. This condenses the

IDM Parameter	Driving regime/state	Identification criteria
Desired speed v_{free}	Cruising	$\tau > T_c$,$a > a_c$, $v > v_c$
Maximum acceleration a	Free-flow	$\tau > T_c$,$a > a_c$ and not cruising
Minimum space gap s_0	Braking	$s < s_c$
Maximum deceleration b	Approaching	$v > v_{ol}$, $\frac{s}{v} > T_c$, $\frac{2(v-v_{leader})}{2s} > a_c$
Desired time gap τ	Following	$\tau < T_c$ and no other regime condition

Table 1: Driving regimes detection and connection to driver model parameters.

driver model dynamics in a framework where discrete transitions are conditionally executed (i.e. between lanes) whereas the longitudinal dynamics obey a continuous transition among driving states/regimes (i.e. collision state, braking state, approaching state, free-following state, and free-flow state). Driver states are tiling the $(\Delta x, \Delta v)$ space, such that the ideal state trajectory enables only for plausible transitions among regimes within the lane. This constraint is also reflected when considering the $(\Delta x, \Delta v, \Delta t)$ space. Here the transitions can only be done towards inferior states (e.g. gaining minimal utility: from following to approaching, following to braking) or equal and superior states (e.g. gaining maximal utility: following to following, following to free-flow). The interplay between states and dynamics across the three dimensions of driver models are depicted in Figure 2, where the Wiedemann diagram, first introduced by Wiedemann (1974), is extended to cover the continuum space-time-acceleration.

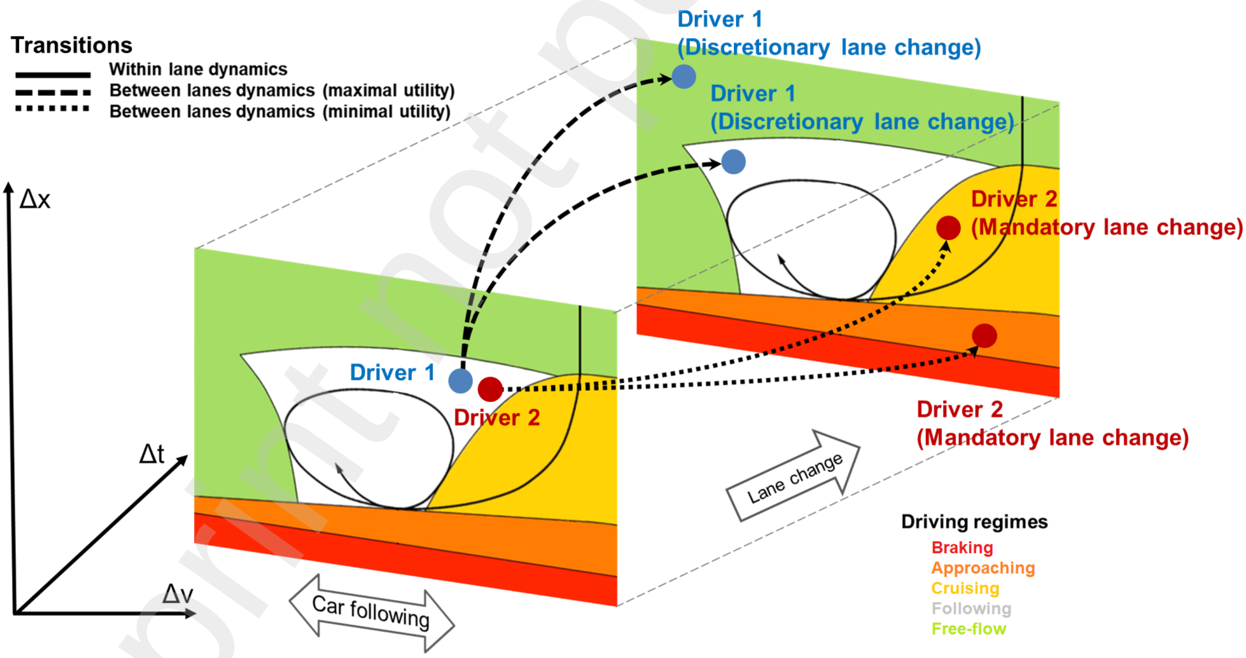

Figure 2: Coupled temporal and behavioral car-following and lane-change dynamics.

2.2. Fuzzy modeling and inference system

2.2.1. Basics

Turning the calibration problem from a computational heavy optimization to a feed-forward inference and regression problem, we need to capture both the spatial and temporal dynamics of the driver and the interactions with the other vehicles in traffic. But, this problems is a multi-factorial one, each driver being characterized by longitudinal and lateral dynamics as well as their temporal evolution, as proposed by Mirman (2019). In order to model the interactions and relationships existing between the physical quantities describing driver's trajectory data (i.e. the evolution of the velocity, acceleration, the space gap and the time headway, $[v(t), a(t), s(t), t(t)]$) and driver models (i.e. IDM, MOBIL) parameters we employ the framework of fuzzy logic, fuzzy sets and fuzzy inference.

The motivation to use fuzzy modeling and inference framework stems from the robust and flexible representation of uncertainty in the driver (noisy) data and its fast rule-based inference that exploits traffic engineering expertise. This alleviates the need of formulating the driver models calibration as an optimization, by converting the problem in the fuzzy framework where we can learn the function F in Equation 2 and use it to map the inferred driver regimes (see Figure 2), accounting for $\rho(x)$ in Eq. 3, to plausible models parameter sets for the coupled IDM and MOBIL, accounting for β_ρ in Eq. 3. Additionally, the fuzzy framework allows the problem to embed physical laws or constraints in the inference process by simply consolidating the fuzzy rule base or the fuzzy regression loss, as shown by the work of Shapiro (2005) through the $\rho(x)\beta_\rho$ term in Eq. 3. Such a coupling of mechanistic modeling with fuzzy inference has been employed to analyze complex systems when quantitative information describing the system is limited, as well as in describing interactions of the system that are not fully characterized, as described by Spolaor et al. (2019).

As we saw in the previous section, mechanistic driver models are generally formalized with a high level of detail (see Equations 4 - 7) and are characterized by parameters that control the dynamic behavior of the driver in order to come closer to the physical reality being modeled. On the other side, the definition of qualitative models, such as fuzzy rule-based systems, is based on uncertain or imprecise information, so that phenomena are described in an approximate way with respect to the physical reality, as explained by Spolaor et al. (2019). However, a significant advantage of qualitative models is that they are more closely related to human understanding and natural language, making them easier to comprehend for traffic engineers or practitioners who are not experts in mathematical modeling.

2.2.2. Fuzzy systems formalism

In our study, we build a repository of problem specific knowledge that models the relationship between driver trajectory data and driving regimes (see Figure 2), through the $\rho(x)\beta_\rho$ term in Eq. 3, and their subsequent mapping to driver models parameters. The inference and learning process reason upon the knowledge base to obtain a plausible driver model parameter set, as shown in Figure 5. The knowledge base (i.e. traffic engineering insights) is in the form of rules that include linguistic variables:

$$\begin{aligned} IF\ x_1\ is\ lt_{i1}\ AND\ x_2\ is\ lt_{i2}\ ...\ AND\ x_n\ is\ lt_{in} \\ THEN\ y_1\ is\ lt_{o1}\ AND\ y_2\ is\ lt_{o2}\ ...\ AND\ y_m\ is\ lt_{om} \end{aligned} \tag{8}$$

where x_i, y_i are the input and output variables, in our case the trajectory data $[v(t), a(t), s(t), t(t)]$ and the regimes $[approaching, braking, cruising, free-flow, following]$; lt_{in} are the input fuzzy sets and $lt_{o,}$ are the output fuzzy sets, respectively. The traffic insights use fuzzy rule semantics describing fuzzy sets partitions in the input data space of the driver trajectory, as shown in Figure 3, where the structure of the overall system in Figure 5 is now given an inside perspective in the Fuzzy Inference block.

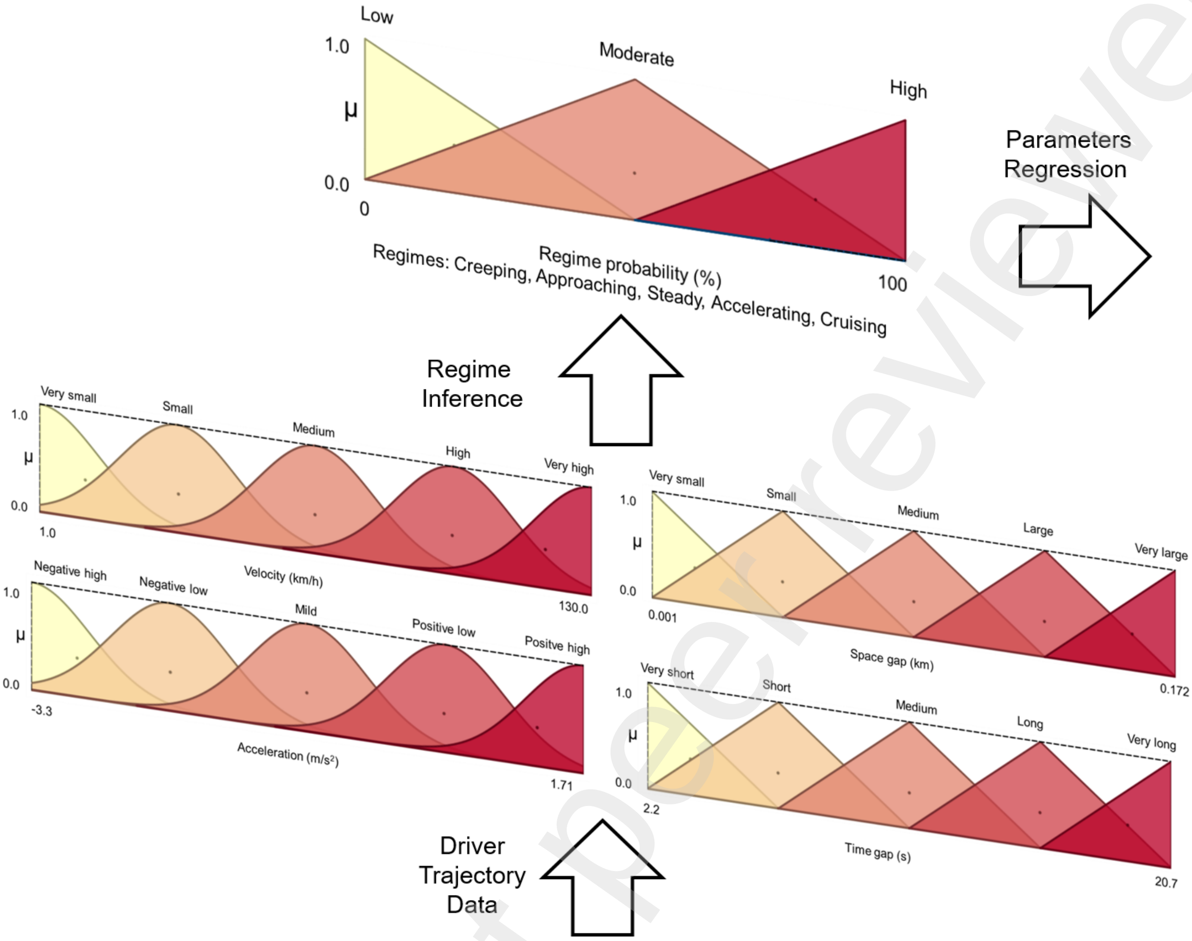

Figure 3: Fuzzy partitions for driver regime inference from trajectory data to model parameters.

As one case see in Figure 3, we split the input data space of trajectory (i.e. the universe of discourse) into partitions using Gaussian and Triangular membership functions that overlap to capture the structure of the patterns in the input data to be mapped into rules that model the interactions among the 4 dimensions of the input data $[v(t), a(t), s(t), t(t)]$. The core component of our calibration approach is a Mamdani fuzzy rule-based system. This type of fuzzy inference system converts crisp driver trajectory data items into fuzzy sets A through Gaussian and Triangular membership functions centered on c_A and with a spread s_A over the universe of discourse defined and partitioned over the trajectory data x

$$\mu_A(x) = e^{\frac{-(x-c_A)^2}{2s_A^2}}, \ x = [v(t), a(t), s(t), t(t)]^T \tag{9}$$

The choice of Gaussian membership functions is motivated by their capacity to capture measurement imprecision and noise. The inference system mapping the fuzzy sets inputs (e.g. high velocity, low acceleration, large space gap, moderate time headway) uses fuzzy conditional statements (i.e. expert rules) to derive the output driver regime:

```
if velocity is verysmall and acceleration is mild and spacegap is verysmall and timegap is veryshort then creepingregime is high
if velocity is small and acceleration is negativelow and spacegap is small and timegap is short then creepingregime is moderate
if velocity is medium and acceleration is negativelow and spacegap is medium and timegap is medium then creepingregime is low
if velocity is high and acceleration is mild and spacegap is medium and timegap is medium then steadyregime is moderate
if velocity is veryhigh and acceleration is mild and spacegap is large and timegap is large then steadyregime is high
if velocity is high and acceleration is positivelow and spacegap is large and timegap is large then steadyregime is low
if velocity is high and acceleration is negativelow and spacegap is medium and timegap is medium then approachingregime is low
if velocity is high and acceleration is negativehigh and spacegap is small and timegap is medium then approachingregime is moderate
if velocity is veryhigh and acceleration is negativehigh and spacegap is medium and timegap is short then approachingregime is high
if velocity is high and acceleration is positivelow and spacegap is large and timegap is large then acceleratingregime is low
```

Figure 4: Fuzzy rules to map driver trajectory data to driver regimes.

In the sample rules we use uniform fuzzy sets of the inputs (e.g. very small, medium, high) across all input variables and appropriate fuzzy sets for the outputs (e.g. low, moderate, high) that capture the expert insights connecting the input linguistic variables (i.e. velocity, acceleration, space gap, time gap) to the output linguistic variables (i.e. creeping regime, approaching regime, accelerating regime, steady regime) according to Figure 2.

From a generic perspective, the fuzzy rule evaluation is a fuzzy conditional statement that defines the relation between the input fuzzy set A (e.g. velocity, acceleration, space gap, or time gap) and the output fuzzy set B (e.g. creeping, accelerating, free-flow, steady-state driving regime) defined over the combined universe of discourse of the input U (i.e. trajectory data) and output V (i.e. driving regimes), $U \times V$. This relation is also a fuzzy set, R given by

$$\mu_R(x, y) = I(\mu_A(x), \mu_B(x))) \tag{10}$$

Here, I is a fuzzy implication operation which is implemented through the T-conjunction operator (i.e. t-norm or product).

In order to recover the regime weighting from the inference we use a defuzzification method in the first-aggregate-then-infer method, hence allowing the system to couple the information from all rules before converting the output to crisp. The individual rules firing output is aggregated using maximum T-conorm operator and subsequently passed through a center of gravity (CoG) operator to obtain the regime weighting:

$$y = \frac{\Sigma_Y y \mu_B(y) dy}{\Sigma_Y \mu_B(y) dy} \tag{11}$$

where B is the output fuzzy set (e.g. creeping, accelerating, free-flow, steady-state driving regime) induced by the A input fuzzy set (e.g. velocity, acceleration, space gap, or time gap) and represents the composition of R and A. The complete pipeline of modeling and inference is depicted in Figure 5.

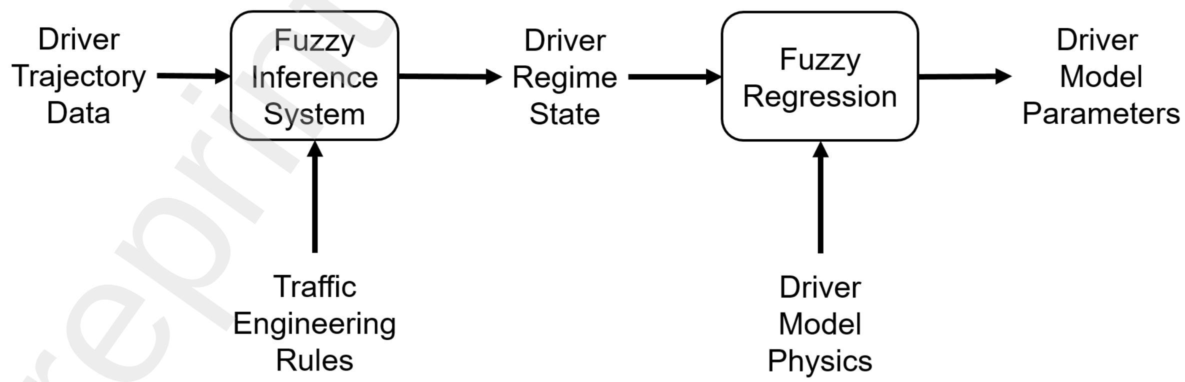

Figure 5: Fuzzy modeling and inference for driver model calibration.

In our framework, the fuzzy rule based system responsible with the intermediate inference of the driver regime acts as nonlinear dynamical system similar to the work of Wang (1995), in that it captures through traffic engineering rules the trajectory evolution in the landscape of the regimes depicted in Figure 2. In order to reduce the vagueness in the mapping from driving regime to model parameters, we adjust the parameters of a fuzzy regressor developed by Shapiro (2005) that exploits the physics of the driver model to push the model parameters towards a plausible solution. The fuzzy regression models the underlying relationships among the driving regimes and the model parameters, capturing effects such as parameter orthogonality, parsimony and robustness - as postulated by Treiber & Kesting (2013). More precisely, in our framework, the fuzzy regression models the ambiguity in describing a crisp measure of the dependent variables, i.e. how to adjust the driver model parameters based on regimes and trajectories. Unlike conventional regression where derivations between observed and predicted values reflect a measurement error, deviation in fuzzy regression reflects the vagueness of the system expressed by the fuzzy parameters of the regression model, as shown in Chang & Ayyub (2001). The fuzzy parameters are in our case possibility distributions that describe the fuzziness of the driver model (see Figure 4).

2.2.3. Framework instantiation

Starting from Equations 2 and 3, the goal of the proposed framework is to compute the parameters β in a way that combines expert knowledge from traffic engineers, the physics of the driver models, and model the dynamics in the fuzzy sets and inference formalism

$$\beta = f_{FR}(f_{FI}(x)),\ \boldsymbol{x} = [\boldsymbol{v}, \boldsymbol{a}, \boldsymbol{s}, \boldsymbol{t}]^T \tag{12}$$

where f_{FR} is the map corresponding to the fuzzy regression, f_{FI} is the map corresponding to the fuzzy inference, and $\boldsymbol{x}$ is the input data vector (i.e. driver trajectory data).

In order to describe the full end-to-end processing within our framework, we begin by decomposing Equation 12. The function f_{FI} is responsible to mapping the driver trajectory $\boldsymbol{x}$ to fuzzy sets A_i (e.g. "*very low*", "*low*", "*medium*", "*high*", "*very high*", "*moderate*") that transform the crisp quantities describing the driver motion to a fuzzy representation, i.e. a value of μ_{A_i}. After mapping all input's components to corresponding fuzzy sets, rule evaluation is performed. The rule evaluation assumes that the output vector $\boldsymbol{y} = [r_{approaching}, r_{braking}, r_{following}, r_{free-flow}, r_{cruising}]^T$ is mapped also to a corresponding fuzzy output set B_j (i.e. "*low*", "*moderate*", "*high*") and assumes the computation of a new fuzzy set R_{ij} that describes the implication operation I inducing the rule connecting A_i with B_j, where B_j is one of braking, approaching, free-flow, following, cruising driving regimes and A_i is one of the (e.g. velocity, acceleration, space gap, or time gap), respectively. Putting all together in the fuzzy inference formalism, for an instance of driver trajectory x_0 the inferred regime y is computed as:

FUZZYFICATION (GAUSSIAN MEMBERSHIP, CENTERED IN c_{A_i} AND SPREAD σ_{A_i})

$$\begin{aligned} \mu_{A_i}(x_0) &= e^{\frac{-(x_0 - c_{A_i})^2}{2\sigma^2_{A_i}}} \\ \mu_{B_j}(y_0) &= e^{\frac{-(y_0 - c_{B_j})^2}{2\sigma^2_{B_j}}} \end{aligned} \tag{13}$$

RULE EVALUATION (T-NORM PRODUCT)

$$\mu_{R_{ij}}(x_0, y_0) = \mu_{A_i}(x_0)\mu_{B_j}(y_0) \tag{14}$$

DEFUZZYFICATION (CENTER OF GRAVITY)

$$y = \frac{\Sigma_{B_i}\ y_0\ \mu_{R_{ij}}(x_0, y_0)dy}{\Sigma_{B_i}\ \mu_{R_{ij}}(x_0, y_0)\ dy}$$

$$y = \frac{\Sigma_{B_i}\ y_0\ \mu_{A_i}(x_0)\ \mu_{B_j}(y_0)dy}{\Sigma_{B_i}\ \mu_{A_i}(x_0)\ \mu_{B_j}(y_0)\ dy} \tag{15}$$

$$y = \frac{\Sigma_{B_i}\ y_0\ e^{\frac{-(x_0-c_{A_i})^2}{2\sigma_{A_i}^2}}\ e^{\frac{-(y_0-c_{B_j})^2}{2\sigma_{B_j}^2}}\ dy}{\Sigma_{B_i}\ e^{\frac{-(x_0-c_{A_i})^2}{2\sigma_{A_i}^2}}\ e^{\frac{-(y_0-c_{B_j})^2}{2\sigma_{B_j}^2}}\ dy}$$

Next, in order to infer the driver parameters β (i.e. IDM and MOBIL parameters) we employ fuzzy regression. Given a set of regime values and configurations of driver models parameters, the goal is to minimize the total vagueness resulting from the (fuzzy) regression through the tuning of its parameters. Fuzzy regression, as developed by Shapiro (2005) proves to be very useful in estimating the relationships among variables where the available data are imprecise (i.e. regime overlap and interference at high-frequency driving style changes), and variables are interacting in an uncertain, qualitative, and fuzzy way (i.e. describing the impact driver models parameters have upon the overall realism of the simulation).

Although our framework is not limited to a certain fuzzy regression model, we considered the approach by Kao & Chyu (2003) to model casual relationships in driver calibration where there is a high ambiguity in describing a crisp measure of the dependent variable (i.e. how to modify the IDM parameter of interest from trajectory data).

In traditional regression analysis, deviations S between observed y and predicted values $\theta_1 + \theta_2 x$ reflect the measurement error which needs to be penalized

$$\begin{aligned} y_i &= \theta_1 + \theta_2 x_i + \epsilon_i \\ S^2 &= \Sigma_i [y_i - (\theta_1 + \theta_2 x_i)]^2 \end{aligned} \tag{16}$$

deviations in fuzzy regression reflect the vagueness of the system structure expressed by the fuzzy parameters of the regression model

$$\begin{aligned} \tilde{y_i} &= \theta_1 + \theta_2 \tilde{x_i} + \tilde{\epsilon_i} \\ \mathbb{E}(\epsilon_i) = 0,\ \sigma^2(\epsilon_i) &= \sigma^2,\ \sigma(\epsilon_i, \epsilon_j) = 0,\ \forall i, j,\ i \neq j \end{aligned} \tag{17}$$

where $\tilde{x_i}$ is the fuzzy set of the input (i.e. driving regime), $\tilde{y_i}$ is the fuzzy set of the output (i.e. IDM and MOBIL parameters) and

$$\tilde{\epsilon_i} = \tilde{y_i} - \theta_1 - \theta_2 \tilde{x_i} \tag{18}$$

Fundamentally, this is a probabilistic-based method (fuzzy least squares calculates the fuzzy regression coefficients using least squares) and relatively robust against outliers compared to the possibilistic-based methods as it allows to compute

$$min \Sigma_{i=1}^N (\tilde{y_i} - \theta_1 - \theta_2 \tilde{x_i})^2 \tag{19}$$

where $\tilde{x_i}$ and $\tilde{y_i}$ are Gaussian-shaped fuzzy sets and the parameters θ_1 and θ_2 are easily calculated using a the closed-form solution in the previous work of Axenie et al. (2019) as

$$\begin{aligned} \theta_2 &= \frac{cov_{x,y}}{\sigma_x^2},\ cov_{x,y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N-1} \\ \theta_1 &= \overline{y} - \theta_2 \overline{x},\ \overline{x} = \frac{\sum_{i=1}^N (x_i)}{N},\ \overline{y} = \frac{\sum_{i=1}^N (y_i)}{N} \end{aligned} \tag{20}$$

Model parameter	Constraints
Maximum acceleration a	$\tau > T_c$,$a > a_c$, $v < v_c$
Minimum space gap s_0	$s < s_c$
Maximum desired deceleration b	$v > v_{leader}$, $\frac{s}{v} > T_c$, $\frac{2(v-v_{leader})}{2s} > a_c$
Desired time headway τ	$\tau < T_c$, $\frac{2(v-v_{leader})}{2s} < a_c$, $s > s_c$,$v > v_c$
Maximum acceleration threshold $a_{threshold}$	$a_{threshold} \leq a_c$
Minimum space gap s_{min}	$s_{min} < s_c$
Minimum time gap t_{min}	$t_{min} \leq \tau$

Table 2: Physical constraints on the driver dynamics.

2.2.4. Introducing physics constraints

In order to ensure that the mapping from driving regime to driver model parameters is plausible, we inject physical constraints on the detection of the regimes from the trajectory data, the possible values of the parameter values, and the stability of the solutions. Starting with Table 1 where the mapping from trajectory to regime is done, we impose the physical constraints on the coupled dynamics depicted in Figure 2.

Considering the special case of our instantiation with IDM car-following and MOBIL lane-change models, we consider the calibration of 7 parameters, of which 4 are for the longitudinal, within-lane dynamics (i.e. maximum acceleration a, minimum space gap s_0, maximum desired deceleration b, and the desired time headway τ) and 3 are for the lateral, between-lane dynamics (i.e. maximum acceleration threshold $a_{threshold}$, minimum space gap s_{min}, and the minimum time gap t_{min}).

The threshold values for v_c, a_c, T_c, s_c are given by driving standards from the European Urban driving Cycle, New European Driving Cycle, Extra-urban driving Cycle norms and the Worldwide Harmonized Light Vehicles Test Procedure.

In the basic formulation of the IDM, velocities of specific vehicles might become negative at specific times, which is, of course, not be desirable from a modeling point of view. More precisely, in the ODE formulation in Equations 4, the velocities of specific vehicles might diverge to $-\infty$ in finite time, so that the solution of the system of ODE's ceases to exist, as demonstrated by Albeaik et al. (2021). In typical calibration routines, when there are inherent flaws leading to unbounded or undefined solutions of the ODE models, as in the case of unbounded acceleration, ad-hoc methods have been traditionally applied such that the unbounded quantities are clipped, leading to "acceptable" numerical solutions, for instance in AIMSUN or SUMO simulators. However, in the calibration and simulation process, the fidelity to the original model is compromised, and the numerical simulations may not represent any realistic behaviors anymore. By considering physics and traffic engineering insights at calibration time, we alleviate these problems by bounding the parameter search space to a plausible and robust region, by employing the regularizing term $\rho(\boldsymbol{x})\boldsymbol{\beta_\rho}$ (see Eq 3).

Another aspect motivating our approach considers the gap-based objective functions used in typical optimization-based calibrations. Such models do not capture neither the coupled longitudinal and lateral dynamics of the driver and his reaction time when changing across driving states. This unmodelled dynamics can, in turn, bring issues in the stability of the model. Formulating the problem as compound car-following and lane-change dynamics through rules modeling interactions among the input variables, driving states and parameters, our calibration results are stable. Following the analysis of Kurtc & Anufriev (2015), we calculated the Jacobian matrix of the ODEs describing the coupled IDM-MOBIL dynamics. We proved the stability of the coupled driver dynamics by showing that the real parts of all eigenvalues of the Jacobian are negative.

As we will see in the experimental section, the physical constraints we inject in the fuzzy inference and fuzzy regression are consistent with other studies focused on the detection of driver states from trajectory data, such as the work of Treiber & Kesting (2013), studies focused on handling the realism of the IDM models, such as the work ofAlbeaik et al. (2021), and studies on closed-loop stability analysis of inferred parameter values, such as the work of Kurtc & Anufriev (2015).

3. Experimental evaluation

In order to evaluate our framework, we have considered both artificially generated scenarios as well as the open source NGSIM real-world datasets. The comparison was performed in terms of a battery of metrics M typically used in traffic engineering, as well as a novel calculation of realism, accounting for the goodness-of-fit function G (i.e. realism score). The goal of the analysis and evaluation was to quantify how well the calibrated models can reproduce reality. In all experiments we have used the large-scale agent-based CityMos Simulator developed by Zehe et al. (2017).

Evaluation on real-world data

For the real-world experiments, we used the NGSIM open source dataset created in the study of U.S. Department of Transportation Federal Highway Administration (2016) that is well established in the community. The NGSIM project was an investment by the U.S. Federal Government to provide a data set which did not exist until then, making it unique. The dataset is characterized by very detailed space-time trajectories on various road segments where cars/drivers can only change speed and acceleration (i.e. obeying car following and lane change dynamics). The basic scenario and ground truth data, used in our experiments, are depicted in Figure 6. For our experiments, we considered a sample of the NGSIM dataset containing

Flow and density by aggregation time

Period	Total flow (all lanes) (vph)	Total density (all lanes) (veh/km)
4:05:00 PM	6612	175.27
4:10:00 PM	8364	234.84
4:15:00 PM	7800	231.34
Average	7592	213.81

Aggregated flow and density per lane

Lane	Flow (vph)	Density (veh/km)
1	1325	20.48
2	1317	43.5
3	1091	37.37
4	1191	110.67
5	1218	39.39
6	1450	49.47
Average	1265.3	50.14

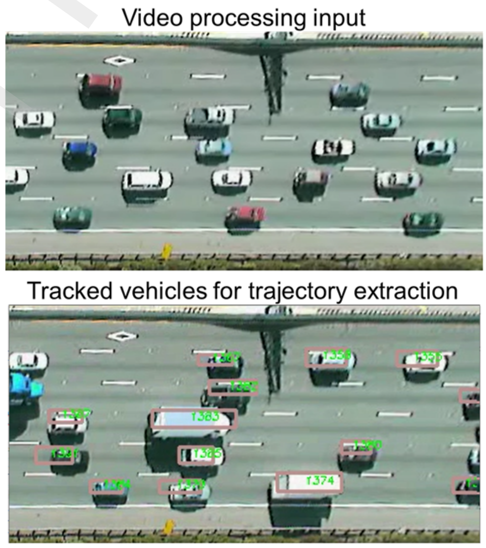

Figure 6: NGSIM open source real-world dataset characteristics.

detailed vehicle trajectory data on eastbound Interstate-80 in the San Francisco Bay area in Emeryville, CA, on April 13, 2005. The study area was approximately 500 meters long and consisted of six freeway lanes, including a high-occupancy vehicle (HOV) lane. A total of 45 minutes of data are available in the full dataset, segmented into three 15-minute periods: 4:00 p.m. to 4:15 p.m.; 5:00 p.m. to 5:15 p.m.; and 5:15 p.m. to 5:30 p.m. The full characteristics and a sneak preview of the scenario captured by the dataset are given in Figure 6. In order to thoroughly evaluate our calibration system (which from now on we will refers to as **Fuzzy**), we followed the excellent reviews of Sharma et al. (2019b), Ciuffo et al. (2008), Punzo & Ciuffo (2009), and of Punzo et al. (2021), and selected the best performing approaches as well as expert suggestions on parameter calibration of coupled lane-change and car-following models. As a result, we implemented for comparison: 1) a fuzzy clustering-based calibration approach inspired by the studies of Higgs & Abbas (2014) and of Ma & Andreasson (2007) which from now on we will refer to as **Clustering**; 2) a combination of genetic algorithms and Bayesian optimization inspired by the studies of Cascan et al. (2019) and of Punzo & Ciuffo (2009) which from now on we will refer to as **Optimization**; and 3) a baseline driver models parametrization from expert observations, termed **Baseline**.

	Driver model calibration approach			
Metric	Baseline	Optimization	**Fuzzy(ours)**	Clustering
Velocity-density characteristic, M_1	0.085	0.159	**0.080**	0.158
Joint velocity-acceleration-headway distribution, M_2	0.061	0.214	**0.006**	0.751
Number of lane changes, M_3	0.536	0.120	**0.777**	0.576
Flow percent deviation, M_4	0.637	0.481	**0.235**	0.382
Density percent deviation, M_5	0.233	0.090	**0.175**	0.483

Table 3: Comparative analysis of each calibration method over the metrics of performance. Values are unit normalized percentage, and lower is better.

In all our upcoming experiments on NGSIM data, we considered five measures of performance M_1,M_2, M_3, M_4, M_5 and one realism score G to quantify how good the simulations running the calibrated models reproduce reality (i.e. how close to ground-truth is the simulation). In our experiments: M_1 captures the velocity-density characteristic and is computed as shown in Appendix A, M_2 describes the joint velocity-acceleration-headway distribution and is computed using the algorithm in Appendix B, while M_3 is quantifying the number of lane changes computed using the algorithm in Appendix C. The battery of metrics is completed by the computation of the percent deviation of traffic flow, representing M_4, and the percent deviation of traffic density, representing M_5, respectively. Finally, the realism score is based on the combined impact of the measures of performance as shown in Appendix D.

We have evaluated the four calibration methods on the five measures of performance (see Appendix A, Appendix B, Appendix C, and Appendix D, respectively) designed to capture the specific characteristics of the NGSIM scenario, namely expressway with direct demand input from on-ramps.

The quantitative analysis of each of the competitive calibration methods provided in Table 3 emphasizes the strengths and limitations of each of the approaches in capturing both the microscopic traffic features (i.e. M_1 to M_3) and the macroscopic traffic characteristics (i.e. M_4 and M_5) to the point that there is a trade-off to be made between the two scales of analysis. For instance, our system outperforms the other methods in terms of reproducing the microscopic traffic features (i.e. velocity-density characteristic and velocity-acceleration-headway distribution) whereas it doesn't perform well when capturing lateral dynamics. This is due to the fact that the fuzzy system doesn't embed many rules to describe the lateral physics of the driver. When considering the reproduction of the macroscopic quantities, our method is on par with the optimization methods that ensure a close match of the road segment properties (i.e. multi-lane highway in the afternoon). Interested in understanding if the driver model calibration effects are also visible at the temporal road-level aggregation, we analyzed (in Figure 7) how each of the calibration methods can capture the space and time distribution of the vehicles as reflected in the total flow and density - with respect to ground truth (see Figure 6).

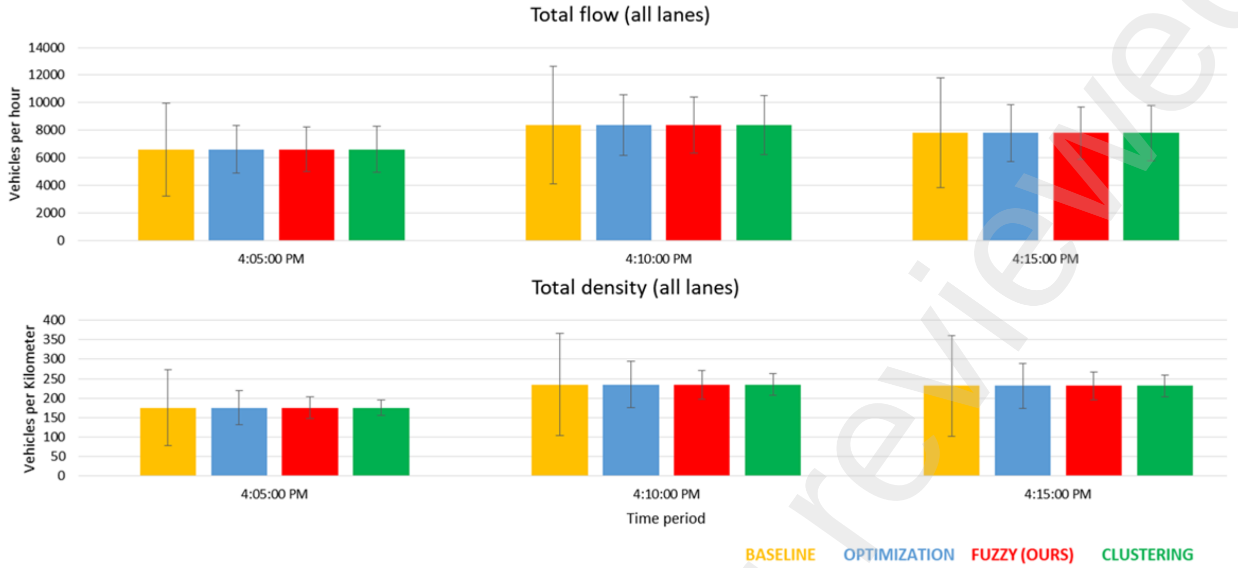

Figure 7: Full analysis of each method on metrics of performance.

Analysis on the accuracy and run-time
The purpose of calibration is to minimize the gap between the simulation and the real-world process the simulation models. In order to assess the performance (i.e. accuracy and run-time), we considered the capability of each of the competing methods not only to capture microscopic characteristics (i.e. M_1 to M_3 metrics) but also to capture the macroscopic traffic characteristics, namely density and flow, along their measures of performance. The full analysis is provided in Figure 8.

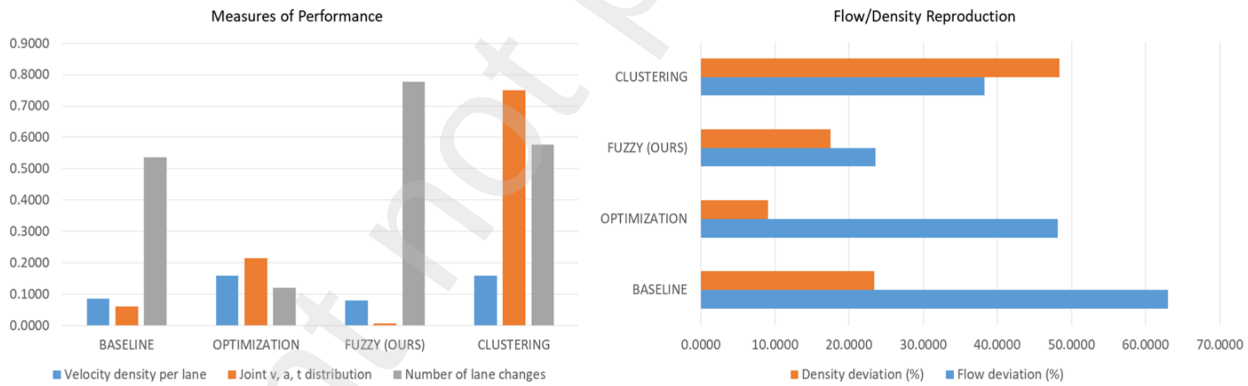

Figure 8: Reality reproduction capability and realism of the calibration approaches.

Our study proposes a novel, optimization-free, calibration method which alleviates the need for time-consuming multi-objective optimization routines that need iterative optimum searches. As we observed in our analysis, one needs to make a trade-off between the capability to reproduce reality and the time needed to find the optimal parameters achieving the best reproducibility. Our approach advocates the use of expert knowledge fused with known driver physics to compute lane-change and car-following parameters close to human behavior. Yet, for a ≈ 4 % loss in the realism (i.e. weighted measures of performance criterion accounting for the goodness-of-fit G) the proposed method offers a 2 order of magnitude improvement in run-time. This improvement scales with the realism requirements of the problem. An important note is that model re-calibration might be also integrated in on-line control loops for traffic control and, hence, need to provide a new driver parameters configuration on a short time window (i.e. 5 min, 10 min, or 15 min). With an extreme loss in realism of ≈ 10 % the proposed system provides a 4 order of magnitude improvement in run-time in the case of high-accuracy (i.e. high realism) driver parameters optimization. In order to evaluate the competing approaches at scale, we also considered three experiments sweeping through the number of optimization iterations (i.e. experiments) used to compute the optimal driver parameter configuration. The overall realism analysis, basically considering the degree of realism and the calibration duration, is provided in Figure 9 along with the parametrization details of our framework and the computing infrastructure.

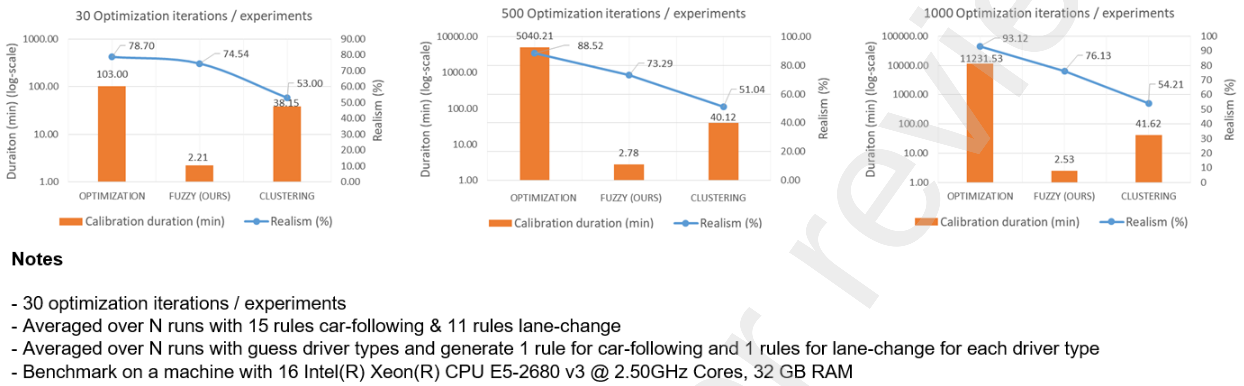

Figure 9: Calibration duration vs. realism analysis on multiple scales of experiments.

Synthetic data scenario In order to evaluate the capabilities of our system in extreme saturation scenarios, we carried a batch of experiments on synthetic simulation data. In the synthetic data experiments, we focused on evaluating the framework for extreme corner-case scenarios, that go beyond typical traffic situations, to see how well the framework is able to adapt/capture such scenario peculiarities.

The synthetic scenario consists of a one-way expressway loop with three travel lanes with traffic demand coming from three injection points (i.e. see direct input, left merge, and right merge, respectively in Figure 10) for a total of 1000 vehicles. The vehicles are injected incrementally such that each vehicle drives the same loop over and over while the traffic volume increases continuously. We measure fundamental diagrams (FD) at one detection point mounted on across the lanes (i.e. measured through an induction loop) (see Figure 10-left panel). As one can easily see, there are also unrealistic elements, such as the left merge from the north demand input origin. The motivation to include such elements was to understand if the calibration process can cope with the unusual driver dynamics induced by such geometry (i.e. lane change is disrupted).

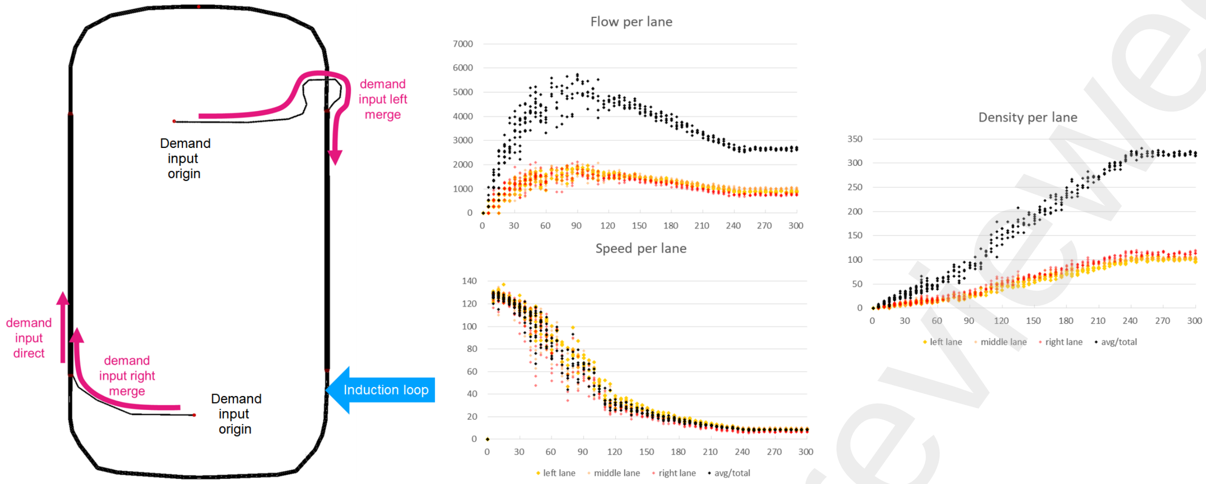

Figure 10: Synthetic scenario for calibration evaluation of our approach.

The length of the expressway loop is 2.75 km, the maximum speed on all three lanes of the loop is 130 km/h, and there are two 500-meter long merge lanes. The simulation step size is 10Hz, summing up to a simulation time of 5 hours. For the experimental analysis, we have performed 12 runs with variation of random seeds.

In order to analyze the capability of our framework to capture the peculiarities of the combination of car-following (i.e. IDM) and lane-change (i.e. MOBIL) in a rather untypical road geometry with high traffic demand under no discharge, we analyzed the extracted FDs at multiple granularities. Despite the high demand, the calibrated combination of IDM and MOBIL was able to reproduce a realistic flow and the breakdown point while the speed decreased sub-linearly with the injection of new vehicles up to the full congestion of two of the lanes (see Figure 10 - right panel).

As we observe in the per-lane analysis in Figure 10 - right panel, there is a realistic average flow of 2000 vehicles/h and a well defined breakdown point ($\approx$ 1.5h) given the continuous injection of vehicles. Additionally, once more vehicles accumulate in the network, there is a time decay speed decrease from the top permitted speed of 130 km/h towards a creeping speed of less than 20 km/h at congestion starting already after 2.5 h. The consistent behavior of the driver models is further supported by the density evolution per lane, where we see an increasing accumulation of cars on all lanes.

Going beyond the temporal evolution of the flow, speed, and density per lane, we analyzed the complete FDs (i.e. flow-speed, speed-density, and flow-density, respectively), to get insights on how the combined calibrated IDM-MOBIL dynamics of the vehicles reflect at per-lane and road-level dynamics. In Figure 11, we observed realistic shapes of both flow-speed and speed-density characteristics with a realistic breakdown at 35 km/h in the flow-density characteristic. These experiments, although exploring boundary situations for real-traffic, allowed us to get insights in how well our calibration framework can exploit expert rules and driver physics to learn driver model parameters which exhibit realistic macro-scale characteristics.

4. Discussion

In this study, we demonstrate that there is an alternative to the traditional approach to model calibration. Finding the optimal value of the model parameters β (see Eq 2) assumes a process where an optimum in the performance metrics is sought for in an iterative fashion. The ultimate goal is to optimize f (see Eq 1). But, in the case of driver model calibration, the choice of performance metrics is subjective and highly dependent on the scenario where the model is used. This, typically, assumes experience with modeling those relevant aspects of the real scenario which capture the impact the driver behavior, typically the $\rho(\boldsymbol{x})\boldsymbol{\beta_\rho}$ term.

Our framework proposes another perspective over calibration, which replaces the iterative optimization procedure with an explicit embedding of expert-rules that describe the driver dynamics in the considered

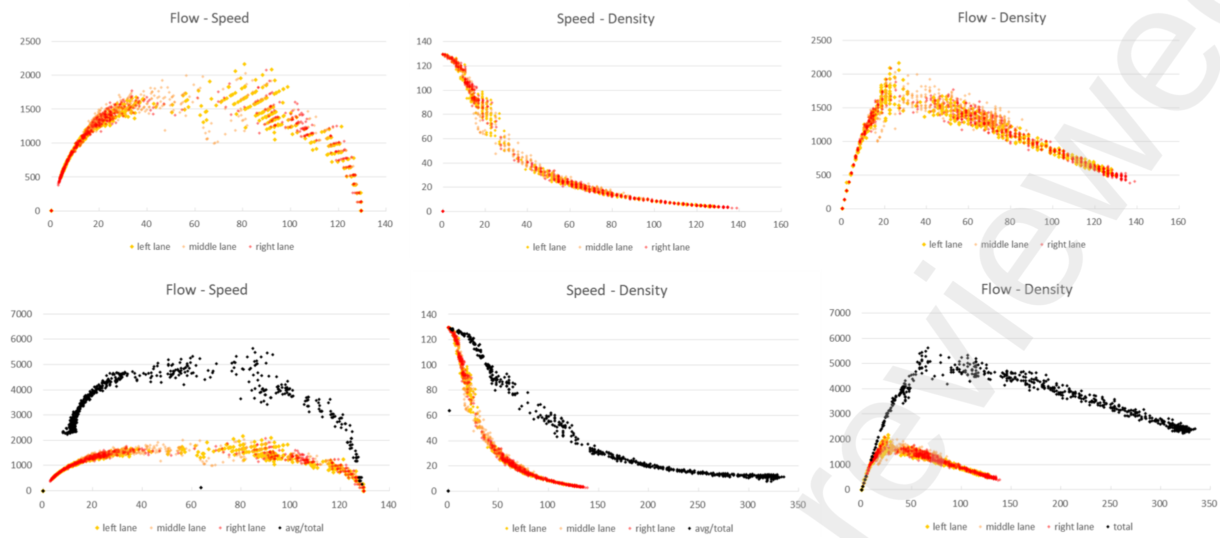

Figure 11: Complete fundamental diagrams analysis for our calibration system.

scenario (see Figure 5). Such rules are combined with the physical models of longitudinal (i.e. car-following) and lateral (i.e. lane-change) dynamics of drivers to fully describe the temporal evolution of driver behavior as depicted in Figure 2. The approach taken in the present instantiation of our framework utilizes fuzzy modeling, inference and rule-based systems for coupled IDM - MOBIL driver dynamics calibration.

Reproducing real traffic dynamics

In order to validate the approach on real traffic data, we conducted a series of experiments on the NGSIM dataset, a de-facto standard in driver model calibration. Here, we comparatively evaluated other state-of-the-art approaches for driver model calibration. For the analysis, we considered a batch of five measures of performance M and a realism score G to quantify the reproducibility capabilities of the competitive calibration approaches (see see Figure 9). Considering solely the capability of the calibrated models to reproduce the real world traffic dynamics, we considered how the per-lane velocity-density characteristic, the joint velocity-acceleration-headway distribution, and the number of lane changes are reproduces in the calibrated simulation (see Table 3). The analysis is completed by the per-road-segment flow and density computation (see Figure 7). When assessing the measures of performance, we observed that each of the four approaches excels in optimizing one of the measures (see Table 3). This behavior resides mainly in the capability of the calibration approach to capture the dynamics described by the measure of performance. More precisely, the baseline approach can capture the velocity-density characteristic and the velocity-acceleration-headway distribution but fails on capturing the lane-changes and the flow and density. This is due to the lack of coupling between the longitudinal and lateral behavior of the driver in the parametrization. On the other side of the spectrum, the Optimization method, employing a multi-objective optimization approach, does overall a good job in capturing the measures of performance and their covariance, but performs poorly while capturing the flow and the velocity-acceleration-headway distribution due to the iterative directed search of the parameter space under the constraints of the multiple measures of performance. A similar outcome is visible in the clustering-based approaches which, overall, provide a rather conservative but completely autonomous (i.a. data-driven) solution. This is due to the iterative splitting of the solution space by means of distance metrics which extract various driver profiles in the trajectory data. Finally, the current instantiation of our framework using fuzzy modeling, inference and rule-based systems for driver dynamics calibration, is capable of capturing very well the low-level criteria (i.e. velocity-density characteristic and the velocity-acceleration-headway distribution) and the macroscopic traffic characteristics (i.e. flow and density) but performs poorly when considering the lateral dynamics (see Figure 8). This is due to the fact that the system had only a limited set of rules (i.e. 11 lane-change MOBIL rules) that capture the lateral dynamics and the coupling with the longitudinal dynamics.

Capturing fundamental traffic dynamics

In a separate batch of experiments, we investigated if the fuzzy modeling, inference and rule-based systems for driver dynamics calibration can reproduce fundamental traffic diagrams (i.e. flow-speed, speed-density, and flow-density). We designed a synthetic scenario which exhibits a strong demand input on a circular path network in Figure 10. The purpose was to explore the capability of the proposed calibration method to capture the peculiarities of heavy traffic and the driver behaviors when a high demand was injected in the network.

Using a number of 15 rules for car-following together with the IDM dynamics and 11 rules for lane-change with MOBIL dynamics, the system was able to reproduce realistically the FDs (see Figure 11) at both fine granularity (i.e. per lane characteristics) and coarse granularity (i.e. per road segment). More precisely, as we observe in the per-lane analysis there is a realistic average flow of 2000 vehicles/h and a well defined breakdown point (≈ 1.5 h) given the continuous injection of vehicles. Additionally, once more vehicles accumulate in the network, there is a time decay speed decrease from the top permitted speed of 130 km/h towards a creeping speed of less than 20 km/h at congestion starting already after 2.5 h already.

Realism vs. calibration duration trade-off analysis

The final analysis of our study is concerned with the realism (i.e. accuracy) of the simulation using the calibrated driver models and the duration of the calibration routine and is depicted in Figure 9 and Figure 12. This aspect is very relevant in closed-loop systems where simulation plays an important role for traffic forecasting which is a very important components of the traffic control system. In principle, such a system can simulate the current state of the traffic based on initial conditions (i.e. measured at street level) and provide a possible evolution of traffic given the driver models used to describe the dynamics of real drivers. For the current analysis, we designed a protocol where we study how the driver model calibration works at scale. We centered the experiment design on the iterative calibration methods, basically considering an increasing number of optimization iterations. Intuitively, a larger number of iterations will ensure better realism but also a slower convergence to the optimal solutions. We then computed the time needed by the optimization-based, clustering-based and our fuzzy modeling, inference and rule-based systems for driver dynamics calibration and the realism value. An important note is that for the clustering-based and the fuzzy modeling, inference and rule-based system, we ran the system the same number of times (i.e. as equal to the iterations of the multi-objective optimization) and took the average run-time and realism, as shown in see Figure 9. When calibration duration is a critical aspect the optimization-based calibration provide a good realism but with a duration two order of magnitude slower than our fuzzy modeling, inference and rule-based system and one order of magnitude with respect to the data-driver clustering-based calibration. At the other end of the spectrum, the optimization-based calibration provides an excellent realism which overcomes with $\approx 14\%$ the fuzzy modeling, inference and rule-based calibration system. This is again demonstrating that a trade-off should be made when considering driver model calibration. This demonstrates that in order to accommodate the requirements of the closed-loop traffic control and the expected accuracy of the individual type of driver a trade-off must be made between realism and calibration duration.

The experiments we performed allowed us to gain valuable insights in the suitability of such calibration approaches in typical traffic simulations. Extrapolating from the previous analysis, we added another dimension to the the analysis, namely human/expert involvement in the system parametrization. We synthesized our findings in Figure 12. Note that the human involvement was empirically measured as the time needed for a traffic expert to parametrize the system. As one would expect, a high simulation realism is achieved by time-consuming methods (i.e. optimization-based calibration) which also involve human expertise in designing the mathematics of the performance metrics to optimize for. For the cases in which lower realism is also acceptable, but no human/expert knowledge is available, data-driven methods (e.g. clustering-based calibration) excel, but, of course, with a considerable calibration duration penalty. Finally, when a minimal set of explicit human/expert rules are available to describe the driver dynamics, the approach we propose in this study is a very fast calibration procedure with good capabilities to reproduce both low-level driver behavior and macro-scale traffic characteristics.

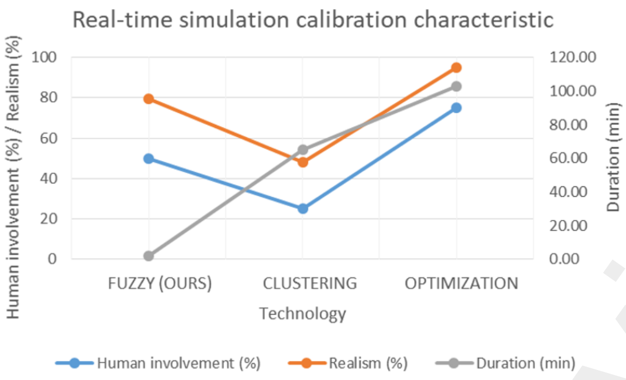

Figure 12: Real-time simulation calibration characteristic. Human involvement versus duration versus realism of the calibrated model simulated.

5. Conclusion

Driver model calibration is a fundamental component in large-scale traffic simulation tools. Such tools are typically used to test mobility hypotheses and traffic control procedures before deploying them in real world. To benefit from simulations, the gap between the reality and the simulated world should be minimized. This is performed through the calibration of the models used in the simulation. State-of-the-art approaches formulate calibration as a constraint optimization methods where a time-consuming iterative process ensures convergence to the optimal solution. This is a desirable solution when realism is important but might become intractable, in this case the simulation would need to update a closed-loop traffic control system to adapt to uncertainties in daily traffic. Data-driven approaches to calibration have gained more and more attention due to their capability to extract insights from large quantities of historical traffic data. Yet, despite their autonomy in extracting relevant features of traffic, they are using expensive computational mechanisms and are not guaranteed to converge to a plausible/realistic solution. To overcome these limitations and targeting real-world deployments, we propose a novel framework and system that combines fuzzy modeling, inference and rule-based systems with physical models of driver dynamics to perform calibration. The target of the proposed system is fast driver models behavior calibration with plausible parameter configurations, which simultaneously capture low-level driver peculiarities and the impact they have on the macro-scale traffic characteristics. This is achieved trough an orchestrated use of expert knowledge, inference, and physical modeling that searches for plausible solutions at both driver level dynamics (coupled longitudinal and lateral motion) and traffic level (lane and road segment flow and density). Our initial experiments on both synthetic and real traffic scenarios demonstrate that there is an alternative to the optimization-based driver behavior models calibration. Benefiting from expert knowledge, known physics models, and inference capabilities the calibration framework provides a very good trade-off between realistic reproduction of real traffic dynamics, plausible choice of driver parameters, and a very fast calibration procedure. Our next steps are oriented in improving the rule-base for lateral dynamics and including more complex merge-yield situations for robust urban calibrations. Our initial results give us confidence that this framework could capture realistic behaviors also in mixed urban-extraurban traffic.

References

Albeaik, S., Bayen, A., Chiri, M. T., Gong, X., Hayat, A., Kardous, N., Keimer, A., McQuade, S. T., Piccoli, B., & You, Y. (2021). Limitations and improvements of the intelligent driver model (idm). *arXiv preprint arXiv:2104.02583*, .

Alhariqi, A., Gu, Z., & Saberi, M. (2021). Calibration of the intelligent driver model (idm) with adaptive parameters for mixed autonomy traffic using experimental trajectory data. *Transportmetrica B: transport dynamics*, (pp. 1–20).

Axenie, C., Tudoran, R., Bortoli, S., Al Hajj Hassan, M., & Brasche, G. (2019). Narpca: Neural accumulate-retract pca for low-latency high-throughput processing on datastreams. In *International Conference on Artificial Neural Networks* (pp. 253–266). Springer.

Barceló, J., & Casas, J. (2005). Dynamic network simulation with aimsun. In *Simulation approaches in transportation analysis* (pp. 57–98). Springer.

Berghaus, M., Kallo, E., & Oeser, M. (2021). Car-following model calibration based on driving simulator data to study driver characteristics and to investigate model validity in extreme traffic situations. *Transportation research record*, *2675*, 1214–1232.

Bhattacharyya, R. P., Senanayake, R., Brown, K., & Kochenderfer, M. J. (2020). Online parameter estimation for human driver behavior prediction. In *2020 American Control Conference (ACC)* (pp. 301–306). IEEE.

Brockfeld, E., Kühne, R. D., & Wagner, P. (2004). Calibration and validation of microscopic traffic flow models. *Transportation Research Record*, *1876*, 62–70.

Cameron, G. D., & Duncan, G. I. (1996). Paramics—parallel microscopic simulation of road traffic. *The Journal of Supercomputing*, *10*, 25–53.

Cascan, E. T., Ivanchev, J., Eckhoff, D., Sangiovanni-Vincentelli, A., & Knoll, A. (2019). Multi-objective calibration of microscopic traffic simulation for highway traffic safety. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 4548–4555). IEEE.

Chandler, R. E., Herman, R., & Montroll, E. W. (1958). Traffic dynamics: studies in car following. *Operations research*, *6*, 165–184.

Chang, Y.-H. O., & Ayyub, B. M. (2001). Fuzzy regression methods–a comparative assessment. *Fuzzy sets and systems*, *119*, 187–203.

Ciuffo, B., Punzo, V., & Torrieri, V. (2008). Comparison of simulation-based and model-based calibrations of traffic-flow microsimulation models. *Transportation Research Record*, *2088*, 36–44.

Genser, A., Hautle, N., Makridis, M., & Kouvelas, A. (2021). An experimental urban case study with various data sources and a model for traffic estimation. *Sensors*, *22*, 144.

Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, *15*, 105–111.

Guo, J., Kurup, U., & Shah, M. (2020). Is it safe to drive? an overview of factors, metrics, and datasets for driveability assessment in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, *21*, 3135–3151. doi:10.1109/TITS.2019.2926042.

Guo, Q., Li, L., & Ban, X. J. (2019). Urban traffic signal control with connected and automated vehicles: A survey. *Transportation research part C: emerging technologies*, *101*, 313–334.

Hang, P., Lv, C., Xing, Y., Huang, C., & Hu, Z. (2021). Human-like decision making for autonomous driving: A noncooperative game theoretic approach. *IEEE Transactions on Intelligent Transportation Systems*, *22*, 2076–2087. doi:10.1109/TITS.2020.3036984.

Higgs, B., & Abbas, M. (2014). Segmentation and clustering of car-following behavior: Recognition of driving patterns. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 81–90.

Kao, C., & Chyu, C.-L. (2003). Least-squares estimates in fuzzy regression analysis. *European Journal of Operational Research*, *148*, 426–435.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*, 422–440.

Kesting, A., & Treiber, M. (2008). Calibrating car-following models by using trajectory data: Methodological study. *Transportation Research Record*, *2088*, 148–156.

Kesting, A., Treiber, M., & Helbing, D. (2007). General lane-changing model mobil for car-following models. *Transportation Research Record*, *1999*, 86–94.

Krajzewicz, D. (2010). Traffic simulation with sumo–simulation of urban mobility. In *Fundamentals of traffic simulation* (pp. 269–293). Springer.

Kučera, T., & Chocholáč, J. (2021). Design of the city logistics simulation model using ptv vissim software. *Transportation Research Procedia*, *53*, 258–265.

Kuefler, A., Morton, J., Wheeler, T., & Kochenderfer, M. (2017). Imitating driver behavior with generative adversarial networks. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 204–211). IEEE.

Kurtc, V., & Anufriev, I. (2015). Local stability conditions and calibrating procedure for new car-following models used in driving simulators. In *Traffic and Granular Flow'13* (pp. 453–461). Springer.

Li, N., Zhang, M., Yildiz, Y., Kolmanovsky, I., & Girard, A. (2019). Game theory-based traffic modeling for calibration of automated driving algorithms. In *Control Strategies for Advanced Driver Assistance Systems and Autonomous Driving Functions* (pp. 89–106). Springer.

Liu, H. X., Xin, W., Adam, Z., & Ban, J. (2007). A game theoretical approach for modelling merging and yielding behaviour at freeway on-ramp sections. *Transportation and traffic theory*, *3*, 197–211.

Lownes, N. E., & Machemehl, R. B. (2006). Vissim: a multi-parameter sensitivity analysis. In *Proceedings of the 2006 Winter Simulation Conference* (pp. 1406–1413). IEEE.

Ma, J., Dong, H., & Zhang, H. M. (2007). Calibration of microsimulation with heuristic optimization methods. *Transportation Research Record*, *1999*, 208–217.

Ma, X., & Andreasson, I. (2007). Behavior measurement, analysis, and regime classification in car following. *IEEE transactions on intelligent transportation systems*, *8*, 144–156.

Magdalena, L. (2015). Fuzzy rule-based systems. In *Springer Handbook of Computational Intelligence* (pp. 203–218). Springer.

Mahmassani, H., & Sheffi, Y. (1981). Using gap sequences to estimate gap acceptance functions. *Transportation Research Part B: Methodological*, *15*, 143–148.

Makridis, M., Mattas, K., Anesiadou, A., & Ciuffo, B. (2021). Openacc. an open database of car-following experiments to study the properties of commercial acc systems. *Transportation research part C: emerging technologies*, *125*, 103047.

Markkula, G., Benderius, O., Wolff, K., & Wahde, M. (2012). A review of near-collision driver behavior models. *Human factors*, *54*, 1117–1143.

Massera Filho, C., Terra, M. H., & Wolf, D. F. (2017). Safe optimization of highway traffic with robust model predictive control-based cooperative adaptive cruise control. *IEEE Transactions on Intelligent Transportation Systems*, *18*, 3193–3203. doi:10.1109/TITS.2017.2679098.

Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? In *Human behavior and traffic safety* (pp. 485–524). Springer.

Mirman, J. H. (2019). A dynamical systems perspective on driver behavior. *Transportation research part F: traffic psychology and behaviour*, *63*, 193–203.

Monteil, J., Billot, R., Sau, J., Buisson, C., & Faouzi, N.-E. E. (2014). Calibration, estimation, and sampling issues of car-following parameters. *Transportation research record*, *2422*, 131–140.

Najmi, A., Rashidi, T. H., Vaughan, J., & Miller, E. J. (2020). Calibration of large-scale transport planning models: a structured approach. *Transportation*, *47*, 1867–1905.

Ossen, S., & Hoogendoorn, S. P. (2008). Validity of trajectory-based calibration approach of car-following models in presence of measurement errors. *Transportation Research Record*, *2088*, 117–125.

Punzo, V., & Ciuffo, B. (2009). How parameters of microscopic traffic flow models relate to traffic dynamics in simulation: Implications for model calibration. *Transportation Research Record*, *2124*, 249–256.

Punzo, V., Zheng, Z., & Montanino, M. (2021). About calibration of car-following dynamics of automated and human-driven vehicles: Methodology, guidelines and codes. *Transportation Research Part C: Emerging Technologies*, *128*, 103165.

Shapiro, A. F. (2005). Fuzzy regression models. *Article of Penn State University*, *102*, 373–383.

Sharma, A., Zheng, Z., & Bhaskar, A. (2019a). Is more always better? the impact of vehicular trajectory completeness on car-following model calibration and validation. *Transportation research part B: methodological*, *120*, 49–75.

Sharma, A., Zheng, Z., Bhaskar, A., & Haque, M. M. (2019b). Modelling car-following behaviour of connected vehicles with a focus on driver compliance. *Transportation research part B: methodological*, *126*, 256–279.

Spolaor, S., Nobile, M. S., Mauri, G., Cazzaniga, P., & Besozzi, D. (2019). Coupling mechanistic approaches and fuzzy logic to model and simulate complex systems. *IEEE Transactions on Fuzzy Systems*, *28*, 1748–1759.

Sugiyama, Y. (1999). Optimal velocity model for traffic flow. *Computer Physics Communications*, *121*, 399–401.

Toledo, T., Koutsopoulos, H. N., & Ben-Akiva, M. (2007). Integrated driving behavior modeling. *Transportation Research Part C: Emerging Technologies*, *15*, 96–112.

Treiber, M., Hennecke, A., & Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Phys. Rev. E*, *62*, 1805–1824. URL: `https://link.aps.org/doi/10.1103/PhysRevE.62.1805`. doi:`10.1103/PhysRevE.62.1805`.

Treiber, M., & Kesting, A. (2012). Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns. *Transportation research part C: emerging technologies*, *21*, 31–41.

Treiber, M., & Kesting, A. (2013). Microscopic calibration and validation of car-following models–a systematic approach. *Procedia-Social and Behavioral Sciences*, *80*, 922–939.

U.S. Department of Transportation Federal Highway Administration (2016). Next generation simulation (ngsim) vehicle trajectories and supporting data. Provided by ITS DataHub through Data.transportation.gov. Accessed 2022-05-29 from `http://doi.org/10.21949/1504477`.

Wang, L.-X. (1995). Design and analysis of fuzzy identifiers of nonlinear dynamic systems. *IEEE transactions on automatic control*, *40*, 11–23.

Wiedemann, R. (1974). *Simulation des verkehrsflusses*. Ph.D. thesis IfV.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*, 495–518.

Xu, X., Wang, X., Wu, X., Hassanin, O., & Chai, C. (2021). Calibration and evaluation of the responsibility-sensitive safety model of autonomous car-following maneuvers using naturalistic driving study data. *Transportation research part C: emerging technologies*, *123*, 102988.

Yang, D., Jin, P., Pu, Y., & Ran, B. (2013). Safe distance car-following model including backward-looking and its stability analysis. *The European Physical Journal B*, *86*, 1–11.

Yang, Z., Masuda, S., & Tatsubori, M. (2021). Online adaptation of parameters using gru-based neural network with bo for accurate driving model. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems* (pp. 33–36).

Zaky, A. B., Gomaa, W., & Khamis, M. A. (2015). Car following markov regime classification and calibration. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 1013–1018). IEEE.

Zehe, D., Nair, S., Knoll, A., & Eckhoff, D. (2017). Towards citymos: a coupled city-scale mobility simulation framework. *5th GI/ITG KuVS Fachgespräch Inter-Vehicle Communication*, *2017*, 03.

Zheng, Z. (2014). Recent developments and research needs in modeling lane changing. *Transportation research part B: methodological*, *60*, 16–32.

Zhu, M., Wang, X., Tarko, A. et al. (2018). Modeling car-following behavior on urban expressways in shanghai: A naturalistic driving study. *Transportation research part C: emerging technologies*, *93*, 425–445.

Appendix A. Computation of M_1, the velocity-density characteristic.

1. Let $d_{t,n}$ be the density at time step t of lane n and $d_{t,n} = \frac{\|K_{t,n}\|}{L_n}$ where $K_{t,n}$ is the set of cars on lane n in time step t and L_n is the length of lane n in meters.
2. Let $v_{t,n}$ be the average speed at time step t on lane n and $v_{t,n} = \sum_{i \in K_{t_n}} \frac{v_{t,n,i}}{\|K_{t,n}\|}$ where $v_{t,n,i}$ is the speed of car i at lane n at time step t.
3. Let d_n be the overall density of lane n over time period T defined as $d_n = \sum_{t \in T} \frac{d_{t,n}}{\|T\|}$
4. Let v_n be the overall average speed of lane n over time period T defined as $v_n = \sum_{t \in T} \frac{v_{t,n}}{\|T\|}$
5. Let the vector l to specify the overall velocity and density of lane n in 2-dimensional space. $\boldsymbol{l_n} = \binom{v_n}{d_n}$
6. We defined the calibration metric M_1 as the average L2 distance between the measured $({l_n}^0)$ and simulated (l_n)overall densities and velocities: $M_1 = \frac{\sum_{n \in N} \left\| l_n - {l_n}^0 \right\|_2}{\|N\|}$ where N is the set of lanes.

Appendix B. Computation of M_2, the joint velocity-acceleration-headway distribution.

1. Let the area of interest be defined as $\mathcal{X}$
2. Let $v_{t,x,i}$ be the speed of car i at time step t at position x.
3. Let $a_{t,x,i}$ be the acceleration of car i at time step t at position x.
4. Let $h_{t,x,i}$ be the headway in meters of car i at time step t at position x.
5. Let S be the amount of data records inside the area of interest: $S = \|\{\forall v_{t,x,i} : x \in \mathcal{X}\}\|$
6. Let r_x be a the collection of velocities(r_v), accelerations(r_a), or headways(r_h) inside a range $r : \mathcal{R}^2$
7. Let p_{r_v,r_a,r_v} be the estimated probability that at any point during the data collection time a car from the real world data set will have speed within r_v, acceleration within r_a, and headway within r_h:
   $p_{r_v,r_a,r_v} = \dfrac{\|\{(t,x,i) : v_{t,x,i} \in r_v \wedge a_{t,x,i} \in r_a \wedge h_{t,x,i} \in r_h\}\|}{S}$
8. Let q_{r_v,r_a,r_v} be defined in the same way as p_{r_v,r_a,r_v} but estimating the probabilities form the data from the simulation run.
9. Let M_2 be the Kullback–Leibler divergence of the simulation probability distribution with respect to the real world data set distribution: $M_2 = \sum\limits_{\forall r_v,r_a,r_v} p_{r_v,r_a,r_v} \log \dfrac{p_{r_v,r_a,r_v}}{q_{r_v,r_a,r_v}}$

Appendix C. Computation of M_3, the number of lane changes.

1. Let C^0 and C be the total number of lance changes in the recorded data set and in the simulation respectively
2. Let M_3 be the relative error between the measured number of changes and the simulated one: $M_3 = \dfrac{|C^0 - C|}{C^0}$

Appendix D. Computation of the realism (goodness-of-fit) based on the Metrics M_i

1. Let q_i be a chosen maximum (maximum error) value for objective i
2. Assume that 0 is the minimum value for any objective (minimum error).
3. Let x_i be a value of objective function i for a given run of the simulation scenario/s
4. Let r_i be the realism score of objective i: $r_i = 100(1 - \dfrac{x_i}{q_i})$
5. Let R be the realism score of the scenario/s: $R = \dfrac{\sum\limits_{i \in M} r_i}{|M|}$