

# Calibrating Car-Following Models Using SUMO-in-the-Loop and Vehicle Trajectories From Roadside Radar

## Calibrating CF model Parameters

Maxwell Schrader<sup>1</sup> , Arya Karnik<sup>1</sup> , Alexander Hainen<sup>1</sup> , and Joshua Bittle<sup>1</sup> 

<sup>1</sup>University of Alabama, Tuscaloosa, Alabama, USA

**Abstract:** This paper presents an innovative calibration method for car-following (CF) models in the Simulation of Urban MObility (SUMO) using real-world trajectory data from a 1.5 km signalized urban corridor, captured by roadside radars. By applying a sophisticated track-level association and fusion methodology, the study extends trajectory analysis beyond individual radar fields of view. The enhanced data is then utilized to refine the Krauss, IDM, and W99 CF models within SUMO, addressing the literature gap by integrating SUMO into the calibration loop, thereby accommodating the simulator's integration scheme and any model adaptations. The research identifies that default SUMO models tend to exhibit shorter time headways compared to real-world data, with calibration effectively reducing this discrepancy. Moreover, the W99 model, despite its unrealistic acceleration profiles when calibrated without considering acceleration, most accurately captures the higher-end energy consumption distribution. Conversely, the IDM model, with its default parameters, provides the closest approximation to observed acceleration behaviors, highlighting the nuanced performance of CF models in traffic simulation and their implications for energy consumption estimation. Detailed results of optimized parameters for each CF model are provided in appendix in addition to distribution information that may be useful for other modelers to use directly or other datasets to be compared in the future (including expansion of the work to include vehicle classification).

**Keywords:** Traffic Micro-Simulation, Car-Following Models, Car-Following Calibration, Intelligent Driver Model, Roadside Radar Data

## 1 Introduction

Traffic micro-simulation can play a critical role in evaluating vehicle emissions and energy consumption of any size traffic network, with the accuracy of these simulations being heavily dependent on the underlying models that simulate individual driver behavior and energy consumption. Among these, the car-following (CF) model is particularly important, as it governs the longitudinal interactions between vehicles. As traffic simulations become increasingly utilized for a variety of applications, from environmental impact studies to traffic management, the precision of CF models in mirroring

actual driving patterns is paramount. It is this precision, combined with robust energy consumption models, that lends confidence to the simulation results, making the calibration of CF models a fundamental step in the development of reliable and robust traffic simulation systems.

The well-documented need for calibration in traffic simulation is highlighted by the diversity of approaches present in the literature, commonly categorized into capacity calibration, route/demand calibration, and individual trajectory-based calibration [1], [2]. While capacity and route/demand calibrations often rely on aggregate measures like loop detector counts, travel times, and saturation flow rates, they may not capture the intricacies of individual vehicle behavior accurately [2], [3] which would lead to poor accuracy in energy consumption estimates. For applications where the precision of each vehicle's trajectory is necessary, such as in the estimation of energy or emissions, it is crucial to ensure that simulated driver behavior is closely aligned with real-world data in the region of interest. This alignment is essential because variables such as acceleration, deceleration, jerk, and speed greatly influence vehicle-level emissions and fuel consumption [4]–[6]. Calibrating a model to both traffic volume and localized CF model behavior is thus vital. However, the CF model calibration process is computationally demanding and typically depends on a database of real vehicle trajectories, which are challenging to obtain in the field. Researchers often resort to using the NGSIM database, which was collected nearly two decades ago and primarily includes free-way driving data that is distinctly different from the signalized corridors modeled in this study [7]. To address this, previous research investigated the "physicality" of CF models — whether real-world measurable parameters, such as acceleration distributions, could be directly applied to CF models, thereby circumventing the need for complex optimization algorithms. The findings indicated that this approach was insufficient and that trajectory-based CF model calibration was necessary [8].

Despite the extensive body of research on CF model calibration, there appears to be a gap concerning calibration that incorporates traffic simulation software directly within the loop. This inclusion is critical, as an examination of the SUMO source code reveals modifications to the CF models<sup>1</sup>. Additionally, the discrete integration scheme employed in numerical simulations and the impact of simulation step size warrant consideration. Previous research has explored these aspects, with Treiber et al. recommending the ballistic integration scheme for small step sizes [9], while Ciuffo et al. suggested that the influence of integration step size is not easily discernible [10].

Building on this foundation, the current work presents the calibration of Simulation of Urban MObility (SUMO) [11] CF models using trajectories from a 1.5 km long urban corridor. These trajectories were captured by roadside radars, which were deployed as part of an intelligent transportation system (ITS) being used for traffic signal control optimization development. A sophisticated track-level association and fusion methodology is introduced, enhancing the trajectories beyond the field of view of individual radars and enabling the identification of leader-follower pairs with intricate interactions. For the first time, to the authors' knowledge, SUMO is utilized in the loop for CF model calibration, moving beyond the traditional use of mathematical representations. The calibration results are detailed and benchmarked against the default vehicle models in SUMO, with recommendations offered to guide future modelers in their simulation endeavors.

---

<sup>1</sup>[Reference to IDM Modifications](#)

## 2 Background

### 2.1 Car Following Models

Since the 1950s, various CF models have been developed, ranging from early concepts by Reuschel [12] and Pipes [13] to contemporary mathematical, data-driven, and hybrid models [14]. Among these, mathematical models like the Intelligent Driver Model (IDM) [15] remain widely used in traffic simulation software, including SUMO. This section briefly introduces the Krauss model, the default CF model in SUMO, the IDM, and the Wiedemann 99 (W99) model, each with its unique parameters and approaches to modeling vehicle dynamics. Below follows a brief description of each of the considered models.

**Krauss Model:** The Krauss CF model, the default for SUMO, ensures collision-free travel by calculating a safe speed,  $v_{\text{safe}}$ , for each following vehicle at every simulation step [16], [17]:

$$v_{\text{safe}}(t) = v_l + \frac{g(t) - v_l(t) \cdot \tau}{\frac{v_f}{b \cdot v_f} + \tau} \quad (1)$$

Here,  $g(t)$  represents the gap to the leading vehicle,  $\tau$  the reaction time, and  $b$  the comfortable braking deceleration. The desired speed,  $v_{\text{des}}(t)$ , factors in the vehicle's maximum acceleration and the driver's behavior:

$$v_{\text{des}}(t) = \min [v_{\text{safe}}(t), v_f(t) + a, v_0] \quad (2)$$

with  $a$  characterizing the driver's preferred maximum acceleration and  $b$  the maximum deceleration. The model allows tuning of  $a$ ,  $b$ , and  $\tau$  to match driver behavior [18].

**Intelligent Driver Model (IDM):** The IDM CF model, conceptualized by Treiber et al. in 2000, defines a vehicle's acceleration  $\dot{v}_f$  as a function of its velocity, the gap to the leading vehicle, and the velocity difference between the vehicles [15]:

$$\dot{v}_f(v_f, s, \Delta v) = a \left[ 1 - \left( \frac{v_f}{v_0} \right)^\delta - \left( \frac{s^*(v_f, \Delta v)}{s} \right)^2 \right] \quad (3)$$

The desired minimum gap  $s^*$  depends on the current velocity and the relative velocity, with the free-flow speed  $v_0$  modeled as a speed factor  $\text{SF}_v$  times the speed limit:

$$s^*(v_f, \Delta v) = s_0 + \tau v_f + \frac{v_f \Delta v}{2\sqrt{ab}} \quad (4)$$

where  $s_0$  represents the minimum acceptable gap between vehicles during standstill.

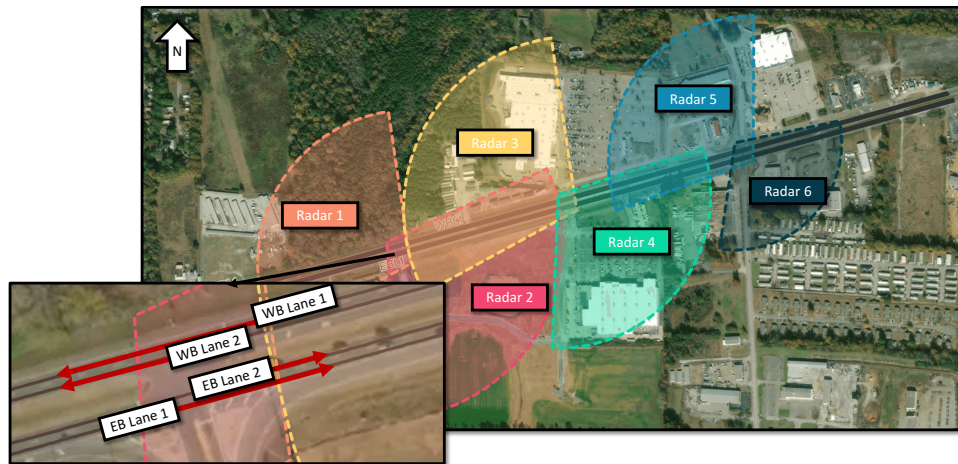
**W99 Model:** The Wiedemann 99 (W99) CF model, is an advancement of Wiedemann's 1974 model, which has been widely adopted in traffic simulation software such as VISSIM. It simulates vehicular behavior by integrating both physical and psychological factors of driving, including stationary and dynamic following distances, perception thresholds, and acceleration behavior [19]. The W99 model is comprised of several equations, which have been skipped for brevity, but in general, it is parameterized by 10 parameters  $\{cc0, cc1, \dots, cc9\}$ .

## 3 Methods

### 3.1 Radar Processing

The focus of this work is a four-lane divided state highway in Tuscaloosa, Alabama, which is a primary conduit for east-west traffic flow, especially during peak commuting

hours. This arterial route connects residential areas in the west with the urban center of Tuscaloosa to the east. The study zeroes in on the highway's main lanes, as depicted in black in Figure 1, to derive detailed vehicle trajectories across the network's signalized intersections.



**Figure 1.** The network of interest shown with the FoV of all 6 radars. The colored region is drawn to the 95th percentile of range. The zoomed intersection shows the lane centerlines that are used as the Frenet frames in both the eastbound (EB) and westbound (WB) directions. The centroid of the network is located at (33.235, -87.614).

To capture traffic dynamics, six iSYS-5220 radar units from InnoSent GmbH<sup>2</sup> incorporated into Econolite's EVO system are employed. These radars, which perform onboard data processing and output vehicle tracks via an API, have undisclosed proprietary algorithms. Consequently, this research approaches the radar system as a black-box. Data collection is executed at 75ms intervals, with synchronization on a central computer and subsequent resampling to a 100ms interval for analysis. The radar units are strategically positioned at three adjacent, signalized intersections along the studied highway, as illustrated in Figure 1, which presents an overview of the network centered at coordinates (33.235, -87.614). The shaded areas in the figure represent the radars' Field of View (FoV), determined by the 95<sup>th</sup> percentile of vehicle detection range. The average radar track from an individual radar extends 146 meters.

### 3.1.1 Fusion and Association

Work by Sharma et. al has shown that car-following calibration is highly sensitive to the length and completeness of vehicle trajectories [20]. The vehicle tracklets from the individual radars are in-complete, that is they may only capture one acceleration or deceleration event. To form longer, more complete trajectories, sensor fusion of the six radars is utilized. Below follows a brief description of the sensor fusion process. Readers more interested in car-following calibration can skip, those looking for more information are directed to a to-be-published work [21]

The first step in the fusion process involves mapping measurements into a unified coordinate system. This study employs a two-step transformation: initially from the radar's local system to Universal Transverse Mercator (UTM) coordinates, and subsequently from UTM to the Frenet coordinate system. The radar-to-UTM transformation is anchored on the known UTM position of the radar and the calibrated rotation angle,  $\theta$ . After transforming radar coordinates to Universal Transverse Mercator (UTM) coor-

<sup>2</sup><https://www.innosent.de/>

ordinates, the data is further converted to the Frenet or curvi-linear coordinate system. This curvilinear coordinate system is parameterized with the path distance ( $s$ ) and the perpendicular distance from the centerline,  $d$ . It is particularly advantageous for modeling the constrained motion of vehicles as they navigate within lanes, adhere to traffic patterns, and execute maneuvers such as lane changes [22], [23].

After the radar data has been transformed in the Frenet frame, the data is filtered using an Interacting Multiple Model (IMM) filter [24]. This filter runs multiple Kalman filters in parallel, each assuming a different motion model, thus enhancing the tracking of maneuvering targets. The IMM maintains a set of model probabilities  $\mu_t$  that indicate the likelihood of each motion model describing the current vehicle behavior, aiding in the categorization of different driving patterns such as lane changes and constant or variable speed driving [22]. This study incorporates three motion models into the IMM filter: constant velocity lane-keeping (CVLK), constant acceleration lane-keeping (CALK), and constant acceleration lane-changing (CALC), as informed by [25]. For short-term trajectory predictions, especially when radar data is missing at the edges of the FoV, the IMM filter uses these models to extrapolate vehicle positions, setting the probability of lane-change and acceleration to zero to reduce lateral and longitudinal error. Vehicle lane occupancy is then inferred by comparing the  $d$  dimension of the filtered trajectory to the known lane widths, thus determining whether a vehicle is in Lane 1 or Lane 2 as they are shown in Figure 1.

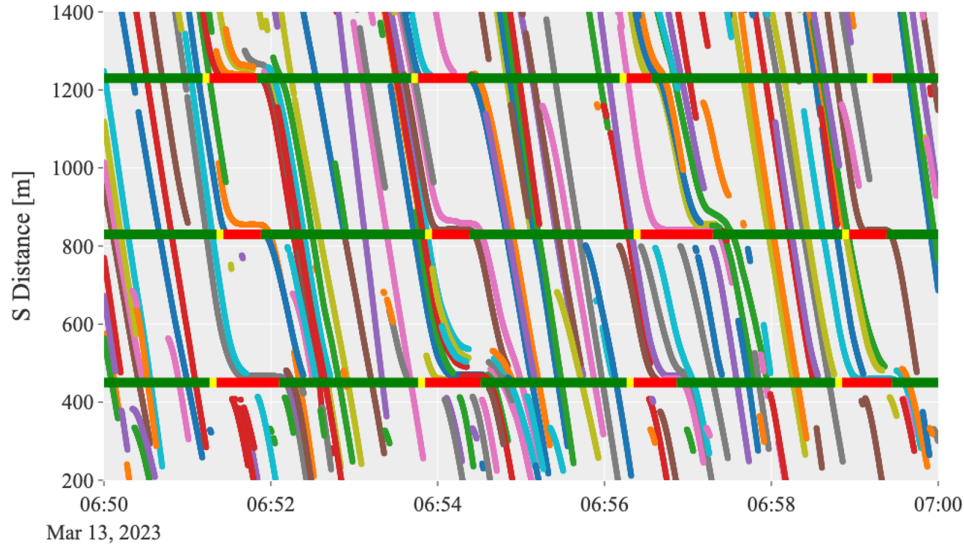
The IMM filter is applied to trajectories from the 6 radars independently. After the filter is applied, the trajectories are associated, using a probabilistic gating method based on the association likelihood distance [26]. This study utilizes a simplified vehicle track association method that relies on the geometric configuration of the roadways. By assuming a single leader for each follower vehicle and that the path distance of the leader exceeds that of the follower, the association process is streamlined. For each identified leader-follower pair, the association likelihood distance is calculated, incorporating the position of multiple points (front, rear, and centroid) on the vehicle to mitigate biases introduced by radar hand-offs.

After associating vehicle tracks, Covariance Intersection (CI) is utilized to fuse the associated radar tracks. CI is ideal when the correlation between tracks is unknown, allowing for the combination of state estimates and covariances without explicit correlation data [27]. Specifically, an adaption of CI, namely ImprovedFastCI, is utilized to streamline the process [28], [29]. After fusion, the trajectories that have been successfully associated and fused have an average length of 705 meters, as compared to the average length of 146 meters recorded by any one radar.

### 3.2 Trajectory Identification & Processing

The calibration of car-following models requires precise identification of leader-follower vehicle pairs. To facilitate this process, data was collected using radar throughout a 24-hour period from the afternoon of March 12, 2023, to the evening of March 13, 2023. The substantial dataset, consisting of 73,084 vehicles, served as a foundation for a highly selective identification process of leader-follower pairs, which is pivotal for the calibration of the models. A snippet of this dataset is shown in Figure 2, which highlights the abundance of high density traffic flow, as well as the impact that traffic signals have on the flow.

As the calibration has been shown to be sensitive to the completeness of vehicle trajectories, and considering that incomplete data can significantly skew validation results,



**Figure 2.** Time-space diagram showing fused trajectories from the EB Lane 1 in Figure 1. Plotted underneath are the light states of the three traffic signals in the network at their corresponding  $s$  position. Trajectories that are the same color represent the same vehicle, with colors being recycled every 10 vehicles.

it is recommended to utilize comprehensive trajectories. These trajectories should encapsulate a variety of driving behaviors, such as free acceleration, cruising, acceleration, deceleration, following, and standstill [20]. The research aimed to select trajectories that included most of these driving conditions to enhance the accuracy of the calibration process while also managing the size of the calibration dataset.

The algorithmic identification process sorted vehicles by their longitudinal position within the lane to differentiate leaders from followers. To ensure that the calibration was done on vehicles that had periods of both acceleration and deceleration, followers were filtered to only include those exhibiting deceleration less than  $-0.5m/s^2$  and acceleration greater than  $0.5m/s^2$ . Furthermore, only leader-follower pairs that maintained their relationship for more than 10s were retained. The time headway between vehicles was either required to be less than 5s at some point during the encounter, based on the assertion that headway times longer than 5s fall outside of the car-following regime [5], [30], or, the distance between pairs needed to reduce to between 1 to 10m, and the following vehicle's velocity had to dip below  $5m/s$  and exceed  $15m/s$  at various points.

Applying these selection parameters resulted in a significant reduction of the initial dataset to 2,397 trajectory pairs. The study did not include the identification of free acceleration and deceleration phases due to the complexity of the traffic network, which features traffic signals and side streets that impact driver behavior and decision-making. The complex interactions within the network present challenges that go beyond the scope of this study, focusing instead on the scenarios most relevant to traditional car-following dynamics.

### 3.2.1 Trajectory Smoothing

The data association and fusion process, as detailed in Section 3.1.1, undergoes further refinement to isolate smooth trajectories suitable for calibration. The literature presents various methods for smoothing data, including low-pass filters, Kalman filtering, and wavelet transforms, with a comprehensive review in [31]. A primary concern

in this process is ensuring "data consistency," which entails maintaining both spacing and velocity in a manner that accurately reflects the vehicle's movement.

To achieve this, the study employs a low-pass Butterworth filter with a cutoff frequency of 0.25Hz. The Interactive Multiple Model (IMM) filter, described in Section 3.1.1, is used on data that has already undergone preliminary filtering. Due to the unknown true process and noise covariance matrices, the study assumes high process noise with  $\sigma_s = 8m/s^2$ . This assumption is necessary to account for the positional "jumps" that can occur during radar handoffs, which introduce noise into the filter's output. To more accurately capture the vehicle's true dynamics, a sixth-order Butterworth filter is implemented. The low cutoff frequency of 0.25Hz for the Butterworth filter effectively attenuates high-frequency noise, preserving the fundamental dynamics of the vehicle's motion. Following the work of [32], the filter is applied to the velocity measurements, independent of the positional data.

### 3.3 Calibration

The calibration of car-following models in SUMO framework necessitates the careful selection of both an optimization method and an error metric. The academic discourse includes extensive discussions on these topics; notably, Punzo et al. have provided a comprehensive summary of different error metrics and offered recommendations on the most suitable measure of performance (MoP) and optimization algorithm for such tasks [33]. The overarching optimization problem inherent in calibrating a car-following model can be formulated as:

$$\begin{aligned} \min_{\beta} \quad & f(MoP_{obs}, MoP_{sim}) \\ \text{s.t.} \quad & LB_{\beta} \leq \beta \leq UB_{\beta} \end{aligned}$$

In this context,  $f(MoP_{obs}, MoP_{sim})$  denotes the car-following behavior to be optimized, serving as the objective function, while  $\beta$  represents the set of model parameters subject to calibration [20]. The evaluation of the objective function is conducted using SUMO-in-the-loop within this work. Due to the inherent non-linear behavior of car-following models and the use of a discrete simulator, the resulting reward structure is characteristically noisy. This noise can cause conventional gradient-based optimization methods to become trapped in local minima, rendering them ineffective [34]. Consequently, the literature generally favors genetic algorithms or other gradient-free methods, which are better suited to handling non-smooth objective functions and have demonstrated superior performance in this domain.

In this research, the Nevergrad optimization library is employed, which features the NGOpt meta-algorithm. NGOpt is designed to intelligently select the most appropriate gradient-free method for a given problem, taking into account the problem's dimensionality and the computational resources available [35].

#### 3.3.1 SUMO Simulation Setup

The SUMO network used for calibration is a comprised of a single edge with a length of 1.8 km, ensuring that it was able to accommodate the longest trajectories in the dataset. The speed limit within the simulation was set at 22.35 meters per second to match the real-world speed limit of 50 miles per hour. Each leader-follower pair was introduced into the simulation environment using SUMO's Traffic Control Interface (TraCI). TraCI was used to control the leader vehicle's position and velocity throughout each trial, also storing both the leader and follower's positions and velocities. The follower vehicle's

parameters were set using SUMO's `vType` tag in an additional file, as not all parameters are accessible via the TraCI API.

### 3.3.2 Measures of Performance

Previous publications have recommended using Normalized Root Mean Square Error of spacing ( $s$ ), velocity ( $v$ ), and acceleration ( $a$ ) over other metrics in terms of calibration precision [33], as defined as

$$\text{NRMSE}(s, v, a) = \text{NRMSE}(s) + \text{NRMSE}(v) + \text{NRMSE}(a) \quad (5)$$

where the Normalized Root Mean Squared Error for every observed variable,  $X$ , is written as

$$\text{NRMSE}(X) = \frac{\text{RMSE}(X, \hat{X})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i)^2}} \quad (6)$$

and  $\hat{X}$  represents the simulated variables.

As shown in Punzo et al., it is preferable to use Eq. (5) as the error metric, namely due to it being Pareto-efficient [33]. Pareto-efficiency is key as optimizing to only one objective (e.g.,  $s$ ) may worsen error on other objectives (e.g.,  $v$  or  $a$ ). Using a Pareto-efficient metric will optimize error on all concerned objectives. Because the dataset included smoothed acceleration data, as outlined in Section 3.2.1, Eq. (5) includes  $a$ . According to the recommendation presented by Punzo et al., utilizing  $a$  is preferable if acceleration data is clean.

## 4 Results

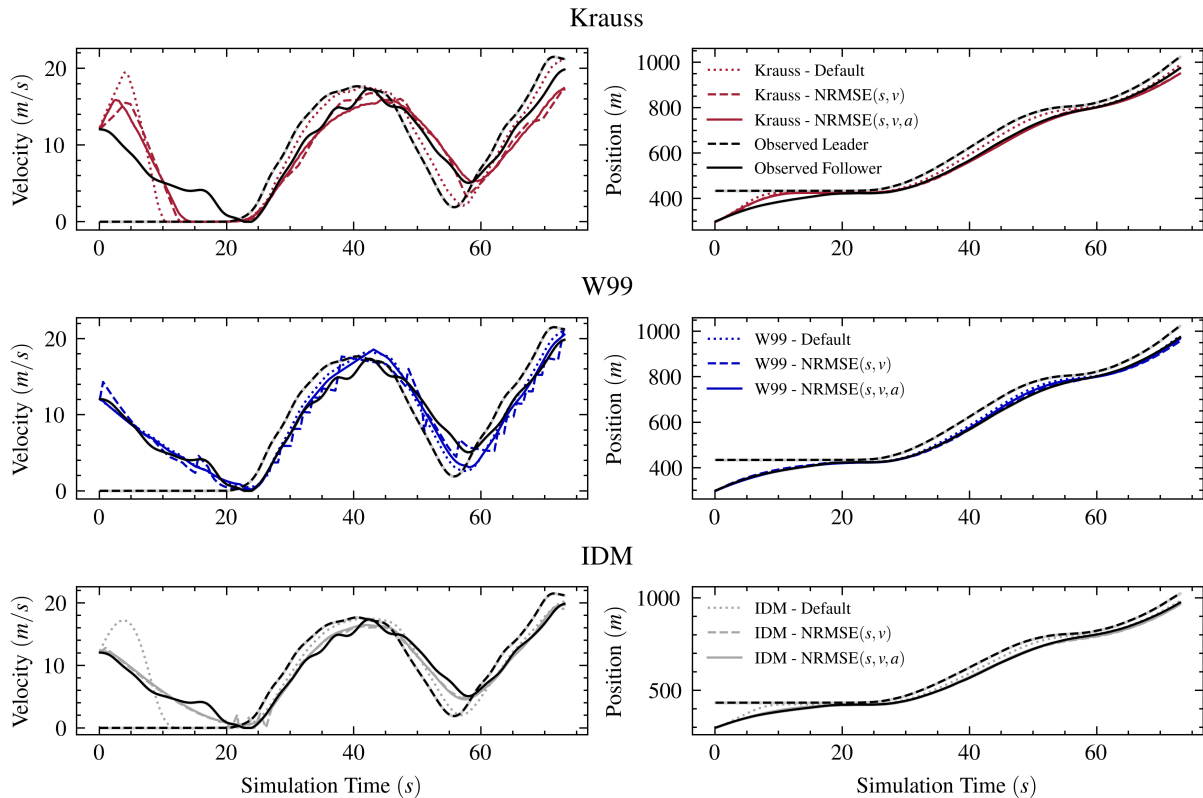
The described calibration methodology was systematically applied to the identified trajectories. Prior to the initiation of calibration, measures of performance (MoPs) were evaluated for each leader-follower pair using the default parameters provided by SUMO. When the departure speed was higher than default speed factor, SUMO updated the speed factor of the inserted vehicle to match the departure speed. After simulating with the default parameters, the optimization of each car-following model was conducted utilizing the NGOpt algorithm, with a computational budget set to 2000 rounds and an early termination criterion after 100 rounds without improvement. Throughout the optimization process, a recurring issue arose where TraCI reported a collision between leader-follower pairs. To address the impact of these collisions on the outcomes, a substantial penalty score was assigned to simulation runs reporting a collision to discourage unsafe value assignment during optimization.

This optimization process was executed on a server equipped with 128 cores, enabling 128 parallel optimization instances. Each simulation iteration was configured with a step size of 0.1 seconds, and the ballistic integration scheme was activated within SUMO, version v1.19.0. The mean optimization time for one trajectory pair was  $44.82s$  ( $\sigma = 40.64s$ ,  $P_{50\%} = 30.76s$ ). The associated code has been made publicly accessible on [GitHub](#).

Before presenting the results of the calibration across the entire set of trajectories, a detailed examination of the calibration outcomes for a selected vehicle is illustrated in Figure 3. The vehicle in question was chosen based on its ranking, which aligns closely with the 75th percentile in terms of RMSE for both spacing and velocity. The figure elucidates several overarching trends that will be discussed later. In this sample case, the leader is initially stationary approximately 150 m ahead of the follower as it



enters the network. Notably, it is observed that the default parameter settings for both the IDM and Krauss models tend to favor shorter time headways. This is evident in the positional plots most easily around 40 seconds as the dotted line is always closest to the leader, as well as in the acceleration profiles when approaching a stationary vehicle at the beginning of the simulation interval with all default models initially accelerating. Additionally, a distinctive "sawtooth" velocity profile emerges when the W99 model is calibrated using  $\text{NRMSE}(s, v)$ .

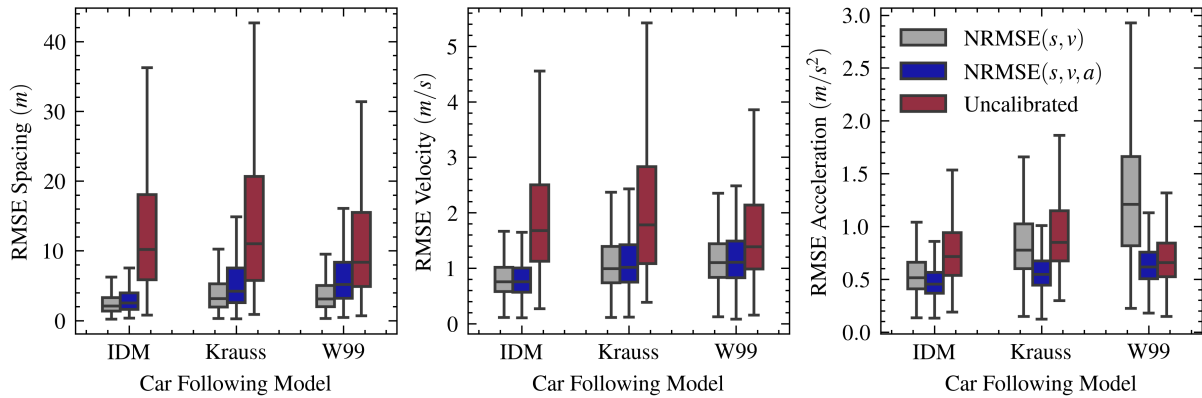


**Figure 3.** Comparison of calibrated vs default trajectories in a car-following scenario. The three rows are the three models considered, with velocity in the first column and position in the second. The plotted trajectory was selected by finding the vehicle nearest the 75th percentile of loss in acceleration, velocity, and spacing when simulating with the default parameters.

The figure also highlights the ability of the calibrated IDM model to best approximate the smooth velocity of the real vehicle, and subsequently results in smooth accelerations and jerk, a topic which has been found in other literature as well [36].

#### 4.1 Error Comparison

The calibration outcomes for the three car-following models detailed in Section 2.1 are depicted in Figure 4, with further summary provided in Table 1. As previously mentioned a subset of simulations led to collisions, with the default parameters of the Krauss model resulting in 762 collisions between following and leading vehicles. This is contrasted with 81 collisions for the W99 model and only 3 for the IDM when utilizing default parameters. For the purpose of analysis in Figure 4, leader-follower pairs that ended in a collision have been excluded. Despite this, the benefit of calibration on the car-following models' performance is evident, with notable improvements observed in the RMSE of both spacing and velocity.



**Figure 4.** RMSE of position, velocity, and acceleration after calibration of the Krauss, IDM, and W99 models. The red color denotes the MoPs evaluated with SUMO default parameters, while grey and blue represent calibration with the  $NRMSE(s, v)$  and  $NRMSE(s, v, a)$  MoPs, respectively.

The calibration of traffic simulation models markedly improves their predictive accuracy, as evidenced by the performance metrics summarized in Table 1. Focusing on the default parameters, the IDM model appears less accurate with a spacing RMSE ( $P_{50\%}$ ) of 10.21 and a velocity RMSE ( $P_{50\%}$ ) of 1.61 compared to its calibrated states. However, upon calibration targeting both spacing and velocity ( $NRMSE(s, v)$ ), the IDM model achieves the best spacing RMSE ( $P_{50\%}$ ) of 2.14 and a notably good velocity RMSE ( $P_{50\%}$ ) of 0.79. This demonstrates a significant improvement over the default settings, indicating the IDM model’s potential when properly tuned. When considering the inclusion of acceleration in the calibration target ( $NRMSE(s, v, a)$ ), the IDM model again stands out, achieving the best acceleration RMSE ( $P_{50\%}$ ) of 0.47 and the second-best spacing RMSE ( $P_{50\%}$ ) of 2.6. This calibration also results in the highest number of best fit vehicles at 973, tying for the fewest crashes at 1, which underscores the model’s robustness in traffic flow representation and safety.

**Table 1.** Statistically Summary of the calibration process. The best value in each column is bolded, with the second best being underlined. The best fit vehicle is selected by identifying number of leader-follower pairs where that model minimizes the ( $NRMSE(s, v, a)$ ).

Model	Model Label	RMSE(a)		RMSE(s)		RMSE(v)		Vehicle #	
		$P_{50\%}$	$P_{95\%}$	$P_{50\%}$	$P_{95\%}$	$P_{50\%}$	$P_{95\%}$	Best Fit	Crash
IDM	$NRMSE(s, v)$	<u>0.53</u>	1.21	<b>2.14</b>	<b>7.94</b>	<u>0.79</u>	<b>1.93</b>	<u>346</u>	<b>1</b>
	$NRMSE(s, v, a)$	<b>0.47</b>	<b>0.99</b>	<u>2.6</u>	<u>8.74</u>	<b>0.78</b>	<u>1.97</u>	<b>973</b>	<b>1</b>
	Default	0.69	1.35	10.21	42.0	1.61	4.56	11	3
Krauss	$NRMSE(s, v)$	0.79	1.54	2.97	10.74	0.97	2.32	49	25
	$NRMSE(s, v, a)$	0.55	<u>1.03</u>	3.98	15.95	0.99	2.39	141	25
	Default	0.85	1.65	11.04	44.95	1.78	5.03	0	762
W99	$NRMSE(s, v)$	1.29	2.4	3.07	11.11	1.13	2.4	17	<b>1</b>
	$NRMSE(s, v, a)$	0.63	1.11	5.23	16.5	1.13	2.68	77	<b>1</b>
	Default	0.66	1.25	8.38	39.5	1.36	4.06	18	81

In simulations using default parameters, the W99 model stands out with the lowest RMSE for spacing ( $\mu_s = 13.11, \sigma = 15.33$ ), statistically outperforming the IDM ( $\mu_s = 14.74, \sigma = 15.34$ ) and Krauss ( $\mu_s = 16.03, \sigma = 16.3$ ) models as confirmed by a post-hoc Durbin-Conover test. Despite this strong baseline, the W99 model does not show the same level of improvement upon calibration; its best calibrated spacing RMSE is 3.07, which is higher than the best results of the IDM. Similarly, the Krauss model’s calibrated spacing RMSE peaks at 2.97, unable to surpass the IDM’s calibrated ver-

sions. Furthermore, including acceleration in the calibration target generally detracts from the spacing RMSE performance for all models.

In contrast, the IDM model's calibrated performance is superior, irrespective of the calibration target. Whether the calibration target is  $\text{NRMSE}(s, v, a)$  ( $\mu_s = 3.73, \sigma = 6.45$ ) or  $\text{NRMSE}(s, v)$  ( $\mu_s = 3.19, \sigma = 6.16$ ), the IDM achieves the lowest RMSE for spacing, with the latter target outperforming the former. This pattern is also reflected in the RMSE of velocity, where the IDM excels when calibrated. Additionally, the IDM maintains the highest vehicle fit count at 973 and ties with the W99 for the fewest crashes at 1, highlighting its calibration potential and robustness in traffic flow prediction and safety.

#### 4.1.1 Headway

Further analysis of the impact is presented through empirical cumulative distribution functions (eCDFs) in Figure 5. The four subfigures represent the eCDFs of velocity, acceleration, headway and energy consumption. Plotted in solid black is the eCDF of real-world observations, with the CF-models plotted on top. The velocity, acceleration, and headways are instantaneous values. The energy eCDF is generated using the total energy consumption of the follower vehicle, with further discussion in Section 4.2.

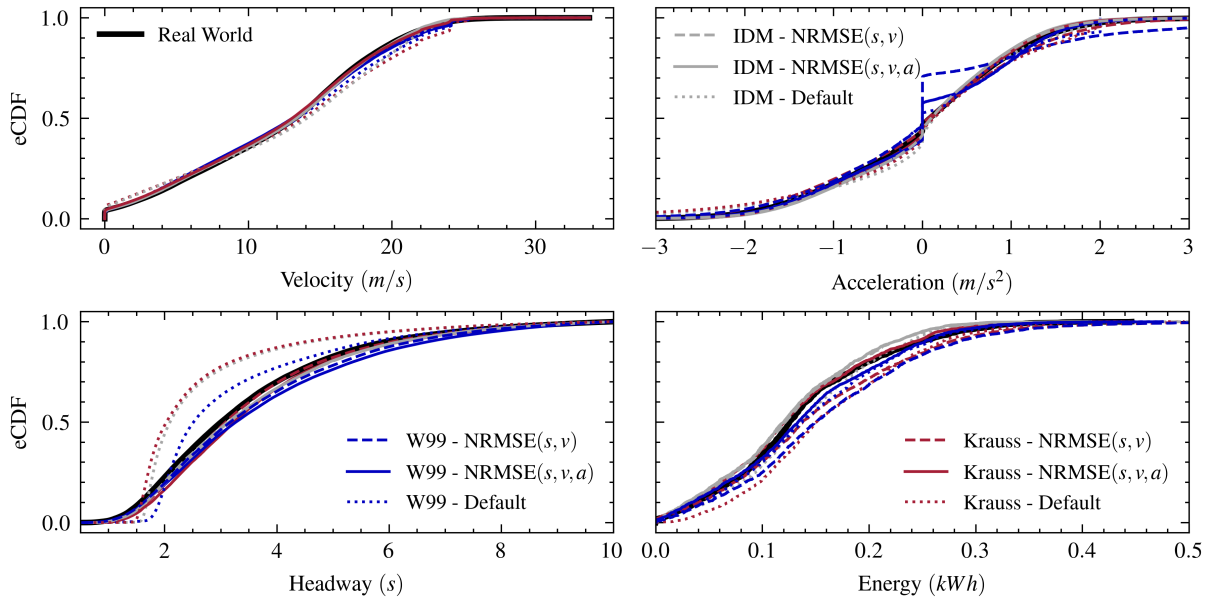
Both spacing and velocity factor into the the time headway ( $\tau$ ) distribution depicted in the bottom left of the figure. Time headway is defined as the ratio of the spacing between the leading ( $s_l$ ) and following ( $s_f$ ) vehicles to the velocity of the following vehicle ( $v_f$ ), as expressed below

$$\tau = \frac{s_l - s_f}{v_f} \quad (7)$$

The empirical cumulative distribution functions (eCDFs) indicate that simulations employing default SUMO parameters generally result in a compressed range of headways. To enhance the relevance of the headway measurements, instances where the velocity is less than  $1m/s$  are excluded to mitigate the influence of low speeds on headway inflation. Furthermore, the headway data is confined to a range of  $0 - 10s$ . The median real-world headway is observed to be  $3.22s$ , with an interquartile range (IQR) spanning from  $2.16s$  to  $4.89s$ . By comparison, the default configurations of the IDM ( $P_{50\%} = 2.15s, IQR = 1.81 - 3.26s$ ) and Krauss ( $P_{50\%} = 2.09s, IQR = 1.72 - 3.14s$ ) models result in notably narrower headways. Even though the W99 model with its default parameters offers a better approximation, it still tends to underestimate driver headways. The implications of these findings are significant for localized simulation, given that time headway is a fundamental element of traffic flow analysis, with the inverse of time headway correlating to traffic flow rate [37].

The disparity between simulated and empirical headways might appear to contradict earlier studies, which asserted that default IDM and Krauss parameters closely mirrored observed headways [8]. However, it is crucial to recognize that the previous research focused on headway measurements at a static location, while the data presented in Figure 5 includes headways calculated at every step of the simulation, thus encompassing a broad spectrum of network locations and driving scenarios. Consequently, although default parameters might accurately replicate headways at intersections, as observed in the previous study, they tend to undervalue headways when considering an entire corridor in this study.

Nevertheless, the headway eCDFs in Figure 5 demonstrate the effectiveness of calibration in aligning with the observed headway distribution. Post-calibration, all



**Figure 5.** Cumulative distribution functions of velocity, acceleration, headway, and energy consumption moving from top-left to bottom right with each CF-model shown for each parameter. Plotted in black are the observed values from the real world. The three colors represent the 3 considered CF-models, with the line style representing the model calibration scheme - including simulating with the SUMO default parameters.

three models closely match the empirical distribution. Notably, the IDM calibrated with  $\text{NRMSE}(s, v)$  achieves the lowest RMSE at  $0.68s$  and exhibits a strong positive correlation ( $r(475597) = 0.93, p \ll 0.01$ ) with the real-world headways. The IDM is followed by Krauss and W99 in their performance rankings, with calibrations targeting  $\text{NRMSE}(s, v)$  showing statistically significant improvements for all models, as confirmed by the Mann-Whitney U test.

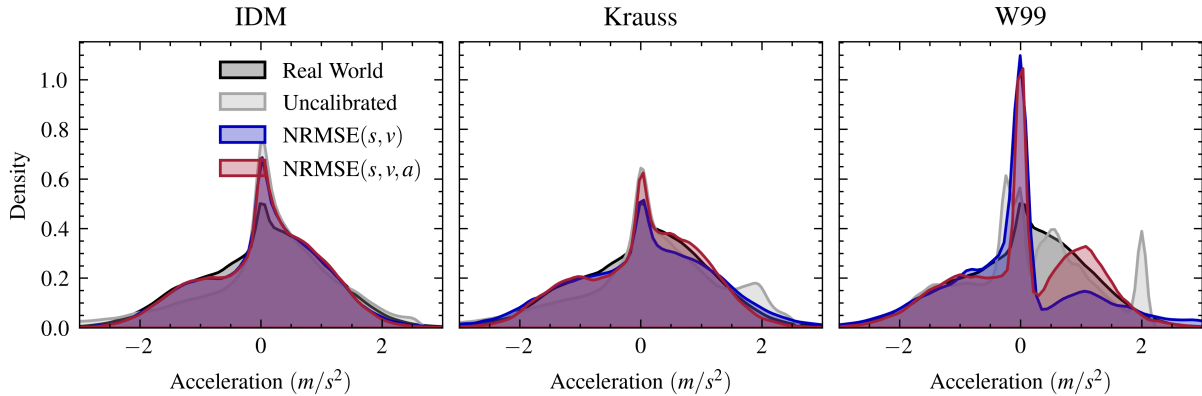
#### 4.1.2 Acceleration

The results presented in Figure 5 for acceleration reveal that both the IDM and W99 models are capable of replicating realistic acceleration behaviors when using default parameter values. Both models exhibit a mean RMSE of less than  $0.8m/s^2$ , with the default parameters of both models surpassing the performance of the Krauss ( $\mu_{RMSE} = 0.85m/s^2$ ) and W99 ( $\mu_{RMSE} = 1.29m/s^2$ ) when calibrated using  $\text{NRMSE}(s, v)$  as the optimization objective. Prior work characterizing similar eCDF results based on road side radar did not find this agreement and rather noted a strong disagreement [8]. After detailed analysis this was a result of the approach to extract behaviors using a piecewise linear fit rather than the smoothing approach used in this work as well as its use of all trajectories rather than only followers.

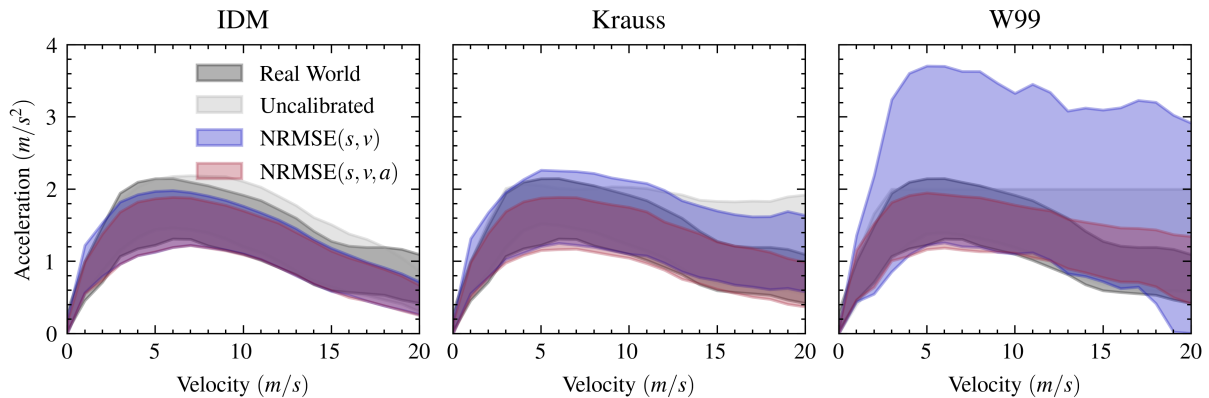
Incorporating acceleration into the measure of performance (MoP) reduces the mean RMSE for all models, with the IDM model achieving the best performance ( $\mu_{RMSE} = 0.85m/s^2$ ). However, according to the principle of Pareto efficiency, optimizing for acceleration in the objective function can result in diminished performance in other dimensions, such as spacing [33], [38]. This trade-off is observed in this study, as employing  $\text{NRMSE}(s, v, a)$  leads to increased spacing errors for all models. Notably, the calibration of the W99 model using  $\text{NRMSE}(s, v)$  results in a higher RMSE for acceleration compared to simulations with the default parameters, as indicated by Figure 4. This phenomenon is also reflected in the characteristic "sawtooth" pattern observed

in the velocity profile of the W99 model, prevalent across trajectories calibrated with  $\text{NRMSE}(s, v)$ .

Figure 6a further explores the distribution of accelerations through kernel density estimates (KDEs) of instantaneous acceleration for the three models, overlaid with the observed acceleration distribution. The KDEs underscore default IDM parameters proficiency in approximating the shape of the real-world acceleration distribution, albeit with a marginally higher likelihood of sharp decelerations and aggressive accelerations.



(a) Comparison of instantaneous acceleration distributions between real-world observations and simulations with default and calibrated parameters.



(b) Filled-area plots showing the range from the 50th to the 90th percentile of positive accelerations versus the velocity at which the acceleration occurred.

**Figure 6.** Exploration of acceleration behavior for the three car-following models, showcasing simulations with both default and optimized parameters.

Figure 6b further investigates acceleration by incorporating the velocity dimension. Here, positive acceleration values are categorized in  $1\text{m/s}$  bins, and the range from the 50th to the 90th percentile is plotted. The curve of real-world values clearly demonstrates the decreasing tendency and capability for acceleration as vehicle speed increases. The IDM model's propensity for high acceleration with default parameters is more pronounced, where simulations consistently overestimate the 90th percentile of acceleration across various speeds. In contrast, the calibrated IDM model tends to underestimate the 90th percentile of acceleration as a function of velocity. This is likely a result of the coupled optimization when calibrating to  $\text{NRMSE}(s, v, a)$ .

The Krauss model performs very well with default parameters, though it is slightly worse than IDM. When calibrating to  $\text{NRMSE}(s, v, a)$  it seems to be the best at capturing the high speed accelerations. Otherwise Krauss performs similarly to IDM when calibrated.

The acceleration distribution of the W99 model differs markedly from both the real-world data and the IDM, as evidenced by Figures 6a and 6b. The distributions exhibit multimodality, whether calibrated or not. In its uncalibrated state, the W99 model overestimates the prevalence of high accelerations, as denoted by the prominent gray peak in the KDE at  $2m/s^2$ . Moreover, the 90th percentile of positive acceleration remains constant at  $2m/s^2$  across all velocities. This model artifact, which leads to unrealistic maximum accelerations, has been previously acknowledged in the literature [4]. While an adaptation to the model was proposed to mitigate this issue, our findings suggest that through calibration, the W99 model can more accurately represent the acceleration-speed relationship, closely mirroring the empirical data as shown in Figure 6b calibrating to  $\text{NRMSE}(s, v, a)$ .

However, Figure 6b also highlights the acceleration error introduced when calibrating to  $\text{NRMSE}(s, v)$ , as it drastically over-estimates the  $P_{90\%}$  at all speeds. This overestimation is corroborated by the fat-tail of high accelerations observed in both Figures 6a and 5, suggesting a systematic bias in the calibration process when acceleration is not explicitly included in the MoP. Additionally, the W99 model still exhibits an "acceleration cliff," a phenomenon identified by Lu et al. [39] as stemming from transitions between model regimes. This results in a lack of acceleration instances in the  $(0m/s^2, 0.5m/s^2]$  range, leading to abrupt transitions from no acceleration to relatively high acceleration and potentially contributing to perceived jerkiness in vehicle motion.

## 4.2 Energy Consumption

When the goal of traffic simulation is emissions or fuel consumption estimation, the accurate modelling of the vehicle trajectory becomes increasingly important, particularly the power requirements under varying traffic conditions. To analyze the impact of calibration on energy consumption estimates and the implications of acceleration differences presented in Section 4.1.2, the instantaneous power  $P$  exerted by the vehicle's wheels, as a function of velocity  $v$  and acceleration  $a$ , was computed using the equation<sup>3</sup>

$$P(v, a) = \frac{v}{1000} \left( g(m + L) (Fr_0 + Fr_1v + Fr_4v^4) + \frac{AC_w\rho}{2}v^2 + a(mr_f + R + L) \right) \quad (8)$$

where  $m$  represents the mass of the vehicle,  $L$  the loading,  $g$  the gravitational acceleration,  $Fr_{0-4}$  are rolling resistance coefficients,  $A$  the frontal area,  $C_w$  the drag coefficient,  $\rho$  the air density,  $r_f$  the rotational mass factor, and  $R$  the reduced mass of the wheel. The constants were taken from the PHEMLightV5 PC\_EU4\_G model in SUMO<sup>4</sup>.

The integration of this power over the course of a driving cycle yields the total energy consumption  $E$ , given by

$$E = \int \max(0, P(v(t), a(t))) dt, \quad (9)$$

which accounts for velocity and acceleration changes over time, with assumption of zero energy required during deceleration, and is commonly written in units of  $kWh$ .

To analyze the difference between car-following models in predicting the fuel consumption of a trajectory, we consider each followers  $E$ . Thus, the car-following model becomes a parametric function to predict energy consumption:

$$\hat{E} = f_{\text{CF model}}(\beta_{l,f}, s_l, v_l) \quad (10)$$

<sup>3</sup>Taken from SUMO's wheel power calculation

<sup>4</sup>PHEMLightV5 PC\_EU4\_G constants

**Table 2.** Model evaluation metrics for the predicted energy consumption of the calibrated car following models. The best performing model is bolded, with the second best being underlined. All units are in kWh.

Model	Opt Metric	Model Evaluation		Summary Statistics (kWh)			
		$R^2$	RMSE	$P_{50\%}$	$P_{90\%}$	$P_{99\%}$	Total
IDM	$\text{NRMSE}(s, v)$	0.853	0.030	<b>0.121</b>	0.224	0.306	204
	$\text{NRMSE}(s, v, a)$	<u>0.873</u>	<u>0.028</u>	0.119	0.223	0.297	201
	Default	0.785	0.036	0.134	0.239	0.323	<u>230</u>
Krauss	$\text{NRMSE}(s, v)$	0.584	0.051	0.137	0.275	0.451	250
	$\text{NRMSE}(s, v, a)$	<b>0.886</b>	<b>0.027</b>	<u>0.120</u>	<u>0.241</u>	0.317	211
	Default	0.590	0.050	0.150	0.268	0.374	267
W99	$\text{NRMSE}(s, v)$	0.486	0.056	0.151	0.284	0.453	265
	$\text{NRMSE}(s, v, a)$	0.851	0.030	0.133	<b>0.252</b>	<b>0.342</b>	<b>227</b>
	Default	0.775	0.037	0.136	0.257	<u>0.356</u>	236
Real World	Observed	-	-	0.125	0.247	0.342	221

where  $\beta_{l,f}$  denotes the set of car-following model parameters describing the follower,  $s_l$  is the leader's position,  $v_l$  is the leader's velocity, and  $\hat{E}$  is the predicted energy consumption. The accuracy of each model is assessed using loss metrics, such as the  $R^2$  score, which quantifies the proportion of variance in the observed energy consumption that is predictable from the model.

The primary objective of utilizing traffic microsimulation is not to predict the fuel consumption of known trajectories; in those instances, the trajectories would be directly passed through a fuel consumption model. Instead, the aim is to model the fuel consumption within a traffic network where it is impractical for the modeler to observe all trajectories. Under these circumstances, the overall statistics highlighted in Table 2 are less pertinent. This is due to interactions between subsequent vehicles, where the error in predicting one vehicle's behavior is not independent of the errors in recreating the trajectories of other vehicles, especially that of the leading vehicle. Therefore, while individual model performance metrics such as  $R^2$  scores, RMSE, and descriptive statistics are presented in 2, they must be contextualized within the broader objective of modeling network-level fuel consumption where the interdependencies of vehicle behaviors play a significant role.

The evaluation metrics for predicted energy consumption across various calibrated car-following models reveal that the Krauss model calibrated with  $\text{NRMSE}(s, v, a)$  outperforms the others, achieving the highest  $R^2$  value of 0.886 and the lowest RMSE of 0.027 kWh. This indicates a strong alignment with real-world energy consumption patterns. The IDM model follows closely when calibrated with  $\text{NRMSE}(s, v, a)$ , with a  $R^2$  value of 0.873 and an RMSE of 0.028 kWh though calibrating to  $\text{NRMSE}(s, v)$  is only slightly worse for IDM. For Krauss however, calibrating to  $\text{NRMSE}(s, v)$  actually has a worse  $R^2$  than default, which underscores the value of including acceleration in the optimization metric for enhancing model accuracy. In terms of median energy consumption ( $P_{50\%}$ ), the IDM model prediction of 0.121 kWh when calibrated with  $\text{NRMSE}(s, v)$  most closely matches the real-world observation of 0.125 kWh (shown in bottom row of Table 2). However, the IDM and Krauss calibrated models underestimate the high energy consumers, as indicated by their underprediction of the  $P_{99th}$  and the total energy consumption. This could be partly explained by Figure 6b, where neither capture the upper-bound of accelerations.

On the other hand, the W99 model, despite its improvements when calibrated with acceleration, shows a lower performance compared to the Krauss and IDM models,

with an  $R^2$  value of 0.851 and an RMSE of 0.030 kWh. However, it is noteworthy that the W99 model calibrated with  $\text{NRMSE}(s, v, a)$  predicts the highest  $P_{90\%}$  and  $P_{99\%}$  values, suggesting a better capture of higher-end energy consumption scenarios. The W99 model, although not the best performer in other metrics, still manages to predict the total energy consumption most accurately at 227 kWh when calibrated with  $\text{NRMSE}(s, v, a)$  compared to 221 kWh predicted from the real-world follower trajectories.

In reviewing the default configurations of the car-following models, it is evident that all of them overestimate the total energy consumption compared to the observed real-world data. This can be seen in both Figure 5 with the eCDF of energy and in Table 2 in the total kWh results. Among the default models, the IDM exhibits the most accurate prediction, with a total energy consumption of 230 kWh. It has the best fit with an  $R^2$  value of 0.785, indicating that while the model can capture a substantial portion of the variance in energy consumption, there is still a significant amount that it does not account for. The Krauss model's default configuration is the least predictive, with an  $R^2$  value of 0.590, highlighting a considerable gap between the model's predictions and actual energy usage. The calibration process enhances the  $R^2$  values across all models, with the Krauss model showing the most significant improvement when acceleration is included in the optimization, bringing its  $R^2$  up to 0.886.

When considering the calibration of the Krauss and W99 models with  $\text{NRMSE}(s, v)$ , the performance is suboptimal. The Krauss model, in particular, has an  $R^2$  value of just 0.584 under this calibration, and the W99 model's performance is similarly poor with an  $R^2$  of 0.486. This aligns with expectations set by the acceleration distributions discussed in Section 4.1.2, where it was noted that both models struggle with accurately capturing realistic acceleration behaviors without the inclusion of acceleration in the calibration metric.

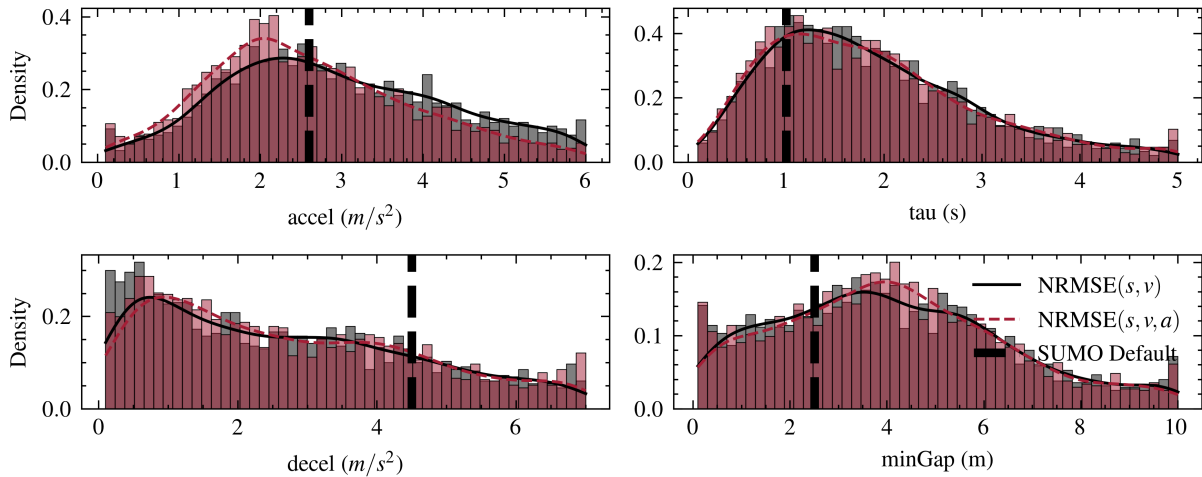
### 4.3 Calibrated Parameter Distribution Discussion

The primary aim of this research was to identify the car-following model parameters that best fit the network depicted in Figure 1. As demonstrated in Section 4.1, calibration significantly enhances the performance of the default car-following models within the SUMO environment. Given that calibration was applied to 2,397 trajectory pairs, the "optimal" parameters derived do not represent singular values but rather span a range of values, each assuming different distributional forms. A comprehensive summary of the calibrated parameters for all three car-following models is provided in the appendix. It is crucial to acknowledge that the dataset encompasses a diverse fleet, including passenger cars, SUVs, and various heavy trucks. Although SUMO permits the modeling of such mixed fleets, this comparison has been simplified to focus solely on personal cars for clarity. Future work may update this result to include classification based on radar or camera identification. For ease of implementation, the model parameters are presented below with their SUMO naming convention<sup>5</sup>. As an example,  $\tau$  from Section 2.1 is tau,  $b$  is decel,  $s_0$  is minGap, and  $a$  is accel.

A subset of model parameters, particularly those commonly subject to calibration, are illustrated in the histograms (Figures 7, 8, 9), with the SUMO default values indicated by solid vertical lines. In line with the discussions from Section 4.1.2, the IDM parameters depicted in both Figure 7 and Table 3 in the appendix exhibit remarkable similarity, regardless of the calibration metric used. Comparing the calibrated IDM model median

<sup>5</sup>[SUMO Vehicle Type Definitions](#)



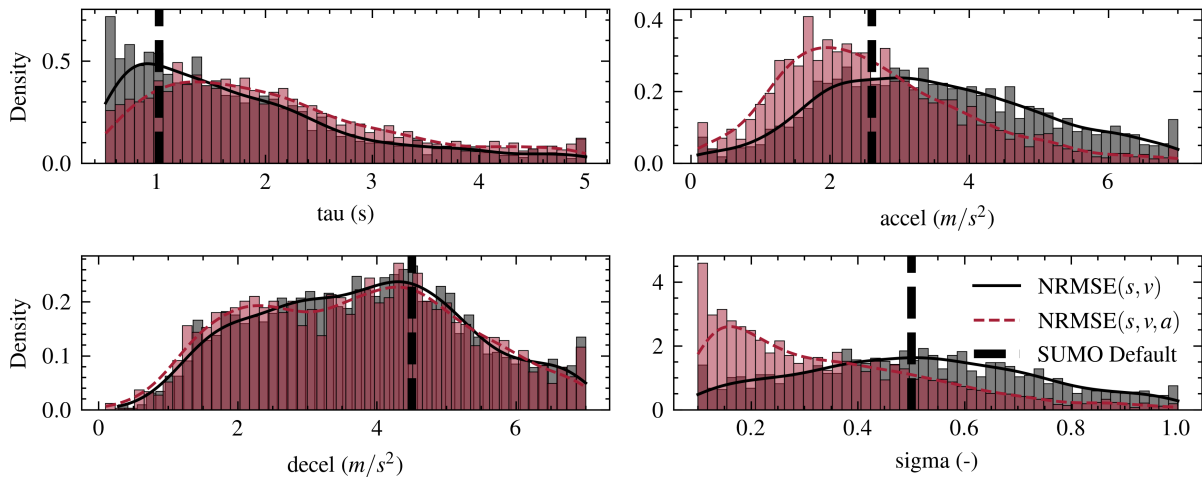


**Figure 7.** Histogram of select IDM parameters calibrated using  $NRMSE(s, v)$  vs.  $NRMSE(s, v, a)$ . The vertical line represents the SUMO default value for the parameter.

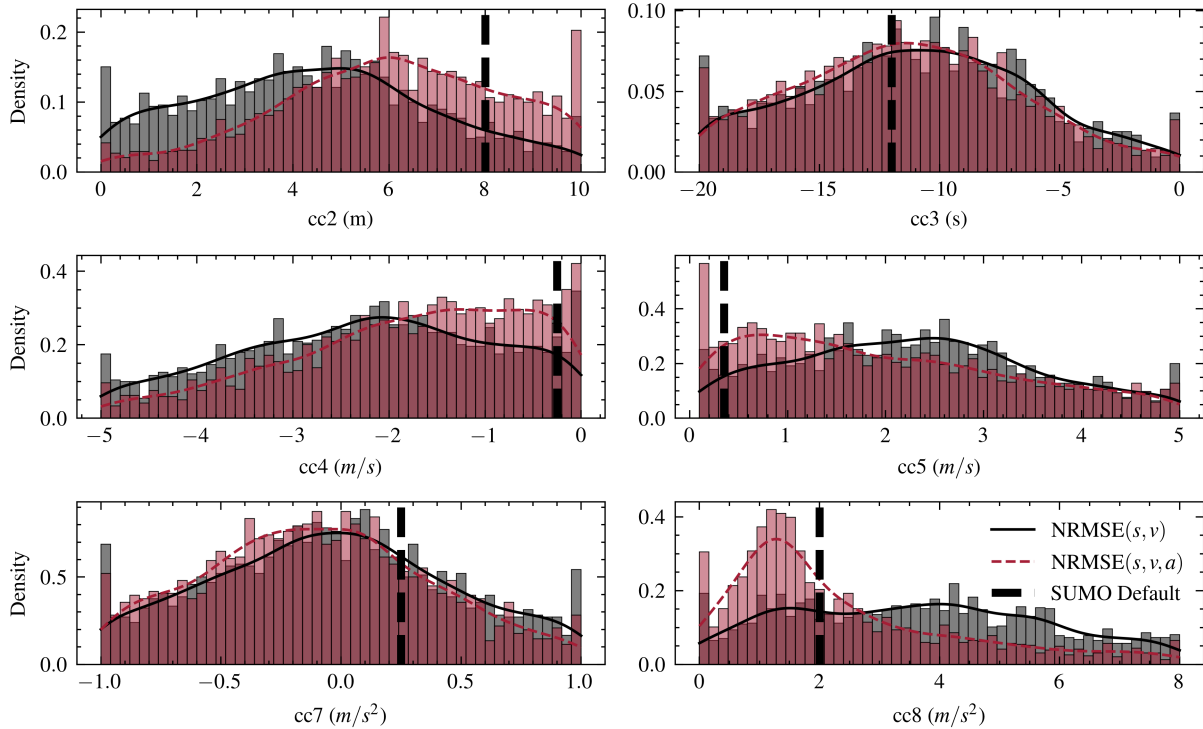
values to the default parameters, notable differences emerge in the decel, minGap, tau, and actionStepLength parameters with generally good agreement for accel.

The deceleration distribution of the calibrated IDM models are right skewed with a mean ( $2.73m/s^2$ ) and median ( $2.41m/s^2$ ) much lower than the default value in SUMO. Because the parameter is shared between multiple car-following models, its likely that the value was originally chosen according to the Krauss model (the default model in SUMO) where the calibrated mean and the median ( $\mu = 3.70, P_{50\%} = 3.75m/s^2$ ) match more closely. The calibrated findings of IDM revealing a deceleration parameter  $\ll 4.5$  matches results from prior literature [36], [40].

The calibrated minGap values for the IDM ( $P_{50\%} = 3.97m$ ) and W99 ( $P_{50\%} = 5.53m$ ) models are substantially larger than the default model's values in SUMO. This discrepancy may be due to measurement inaccuracies in the radar data, particularly as the radar's vehicle length output, which is directly relevant to minGap, has been found to contain significant errors. The calibration may be compensating for these inaccuracies in the radar length estimates. When simulating with a step size of less than 1 second (0.1 seconds in this study), the SUMO documentation recommends using an



**Figure 8.** Histogram of select Krauss parameters calibrated using  $NRMSE(s, v)$  vs.  $NRMSE(s, v, a)$ . The vertical line represents the SUMO default value for the parameter.



**Figure 9.** Histogram of select W99 parameters calibrated using  $NRMSE(s, v)$  vs.  $NRMSE(s, v, a)$ . The vertical line represents the SUMO default value for the parameter.

actionStepLength larger than the simulation step size. Despite this recommendation, the default remains as the step size. The calibration results from this study suggest an optimal actionStepLength range of 0.2 - 0.4 for all three models.

It is also noteworthy that the Krauss and IDM parameter distributions are not Gaussian and exhibit a right skew. For the Krauss model, the optimal sigma distributions for  $NRMSE(s, v)$  versus  $NRMSE(s, v, a)$  differ markedly, suggesting that sigma is a critical parameter for approximating smooth and realistic accelerations, alongside acceleration itself. However, a more detailed sensitivity analysis is required for a definitive conclusion. Similar to the IDM, the calibrated Krauss decel parameter is lower than SUMO's default values, albeit to a lesser extent. Additionally, the mean of the calibrated headway parameter (tau for Krauss and IDM, and cc1 for W99) is higher than the default values, which is consistent with the headway results presented in Section 4.1.1.

## 5 Summary, Conclusions & Future Work

This research presents a comprehensive methodology for calibrating SUMO's CF models, utilizing detailed vehicle trajectory data obtained from a 1.5 km urban corridor equipped with ITS radar sensors. The track-level association and fusion approach used in this study facilitates the identification of complex leader-follower dynamics. Notably, this work represents the first known instance of employing SUMO in a closed-loop calibration process, marking a departure from the conventional practice of utilizing abstract mathematical models. The calibration outcomes have been thoroughly presented and assessed against SUMO's default vehicle models, yielding recommendations that are expected to significantly benefit future modeling efforts within traffic simulations.

The research findings highlight that different models excel in different respects. For instance, the W99 model, with its default parameters, minimizes errors in spacing, ac-

celeration, or velocity. However, once calibrated, the IDM model outperforms others across various metrics. To optimize the W99 model, it was imperative to factor in acceleration error in the minimization metric to avoid introducing velocity inconsistencies and subsequent unrealistic acceleration behaviors. Examining the best  $NRMSE(s, v, a)$  among the three calibrated models, the IDM adjusted using  $NRMSE(s, v, a)$  emerges as superior for 60% of the vehicles, with IDM tailored to  $NRMSE(s, v)$  ranking second at 22%.

Moreover, when the objective is to replicate realistic accelerations and generate follower trajectories that align closely with real-world energy consumption, the calibrated Krauss model stands out in predictive capability overall, though the calibrated W99 model best represents high energy-consuming vehicles. The calibration process also disclosed that all default models are inclined to overestimate total energy consumption, with the IDM being the most precise and the Krauss model exhibiting the most substantial overestimation among the default models.

The study also reveals notable discrepancies between the optimal parameters and SUMO's default parameters. It suggests that an `actionStepLength` of 0.2-0.4 seconds is ideal when simulating with a step size of 0.1 seconds. Furthermore, the calibrated parameters for time headway demonstrate a mean higher than the default settings in SUMO by at least 0.5 seconds. It is unclear how generalizable these conclusions are to other traffic scenarios; however, detailed results of optimized parameters for each CF model are provided in appendix in addition to distribution information that may be useful for other modelers to use directly or other datasets to be compared in the future.

The study acknowledges limitations, including the lack of a hold-out or test set, which means the models were assessed on the same dataset used for training, potentially impacting the broader applicability of the results. Additionally, the ultimate purpose of car-following models in SUMO extends beyond generating precise follower trajectories; it also encompasses the accurate reproduction of overall traffic flow. Future work will aim to overcome these limitations by incorporating a validation phase and investigating the effects of correlated versus uncorrelated sampling at the trajectory level. It will also assess the influence of calibrated versus uncalibrated models on macroscopic traffic flows. A sensitivity analysis is proposed as a valuable future endeavor to pinpoint critical parameters, which could streamline the calibration process by narrowing the calibration space.

Subsequent research will delve into the resilience of the models to variations in simulation step size and the impact of diverse driving behaviors, including lane-changing, approaching traffic signals, starting from a stop, and merging maneuvers. The study also highlights the importance of accounting for different vehicle types in simulations and suggests that the calibration of the W99 model should consider velocity profile smoothness. Introducing penalties for erratic acceleration during calibration could refine the accuracy and realism of the simulations. These future research avenues will build on the current findings to create traffic simulations that more faithfully reflect real-world driving patterns, thereby advancing the efficacy of intelligent transportation systems.

## CRediT authorship contribution statement

**Maxwell Schrader:** Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing - original draft, Writing - review & editing. **Arya Karnik:** Methodology, Software, Writing - original draft **Alexander Hainen:** Resource, Project administra-

tion, Funding acquisition. **Joshua Bittle:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgements and Declarations

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the the Vehicle Technologies Office (VTO) and through Energy Efficient Mobility Systems (EEMS) program award number DE-EE0009210. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This work was also supported by The University of Alabama's Center for Advanced Vehicle Technology (CAVT).

The authors have no conflicts of interests to declare that are relevant to the content of this article.

Currently data provided by request to corresponding author, though future plans include making this data and more from the dataset publicly available online.

## References

- [1] J. Barceló *et al.*, *Fundamentals of traffic simulation*. Springer, 2010, vol. 145. DOI: [10.1007/978-1-4419-6142-6](https://doi.org/10.1007/978-1-4419-6142-6).
- [2] T. V. da Rocha, L. Leclercq, M. Montanino, *et al.*, "Does traffic-related calibration of car-following models provide accurate estimations of vehicle emissions?" *Transportation research part D: Transport and Environment*, vol. 34, pp. 267–280, 2015. DOI: [10.1016/j.trd.2014.11.006](https://doi.org/10.1016/j.trd.2014.11.006).
- [3] V. Punzo and F. Simonelli, "Analysis and comparison of microscopic traffic flow models with real traffic microscopic data," *Transportation Research Record*, vol. 1934, no. 1, pp. 53–63, 2005. DOI: [10.3141/1934-06](https://doi.org/10.3141/1934-06).
- [4] G. Song, L. Yu, and Z. Geng, "Optimization of wiedemann and fritzsche car-following models for emission estimation," *Transportation Research Part D: Transport and Environment*, vol. 34, pp. 318–329, 2015. DOI: [10.1016/j.trd.2014.11.023](https://doi.org/10.1016/j.trd.2014.11.023).
- [5] L. Jie, H. Van Zuylen, Y. Chen, F. Viti, and I. Wilmink, "Calibration of a microscopic simulation model for emission calculation," *Transportation Research Part C: Emerging Technologies*, vol. 31, pp. 172–184, 2013. DOI: [10.1016/j.trc.2012.04.008](https://doi.org/10.1016/j.trc.2012.04.008).
- [6] B. Notter, M. Keller, and B. Cox, "Handbook emission factors for road transport 4.2," *INFRAS, Bern*, 2022. [Online]. Available: [https://www.hbefa.net/e/help/HBEFA42\\_help\\_en.pdf](https://www.hbefa.net/e/help/HBEFA42_help_en.pdf).
- [7] "Next Generation Simulation (NGSIM) Program I-80 Videos.," U.S. Department of Transportation Federal Highway Administration, Tech. Rep., 2016. DOI: [10.21949/1504477](https://doi.org/10.21949/1504477).
- [8] M. Schrader, M. Al Abdraboh, and J. Bittle, "Comparing measured driver behavior distributions to results from car-following models using sumo and real-world vehicle trajectories from radar: Sumo default vs. radar-measured cf model parameters," in *SUMO Conference Proceedings*, vol. 4, 2023, pp. 41–54. DOI: [10.52825/scp.v4i.214](https://doi.org/10.52825/scp.v4i.214).
- [9] M. Treiber and V. Kanagaraj, "Comparing numerical integration schemes for time-continuous car-following models," *Physica A: Statistical Mechanics and its Applications*, vol. 419, pp. 183–195, 2015. DOI: [10.1016/j.physa.2014.09.061](https://doi.org/10.1016/j.physa.2014.09.061).
- [10] B. Ciuffo, V. Punzo, and M. Montanino, "Thirty years of gipps' car-following model: Applications, developments, and new features," *Transportation research record*, vol. 2315, no. 1, pp. 89–99, 2012. DOI: [10.3141/2315-10](https://doi.org/10.3141/2315-10).

- [11] P. A. Lopez, M. Behrisch, L. Bieker-Walz, *et al.*, "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*, IEEE, 2018. [Online]. Available: <https://elib.dlr.de/124092/>.
- [12] A. Reuschel, "Fahrzeugbewegungen in der kolonne," *Osterreichisches Ingenieur Archiv*, vol. 4, pp. 193–215, 1950.
- [13] L. A. Pipes, "An operational analysis of traffic dynamics," *Journal of applied physics*, vol. 24, no. 3, pp. 274–281, 1953. DOI: [10.1063/1.1721265](https://doi.org/10.1063/1.1721265).
- [14] Y. Zhang, X. Chen, J. Wang, Z. Zheng, and K. Wu, "A generative car-following model conditioned on driving styles," *Transportation research part C: emerging technologies*, vol. 145, p. 103926, 2022. DOI: [10.1016/j.trc.2022.103926](https://doi.org/10.1016/j.trc.2022.103926).
- [15] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000. DOI: [10.1103/PhysRevE.62.1805](https://doi.org/10.1103/PhysRevE.62.1805).
- [16] S. Krauß, "Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics," 1998. [Online]. Available: <https://www.osti.gov/etdeweb/biblio/627062>.
- [17] S. Krauß, P. Wagner, and C. Gawron, "Metastable states in a microscopic model of traffic flow," *Physical Review E*, vol. 55, no. 5, p. 5597, 1997. DOI: [10.1103/PhysRevE.55.5597](https://doi.org/10.1103/PhysRevE.55.5597).
- [18] L. Bieker-Walz, M. Behrisch, M. Junghans, and K. Gimm, "Evaluation of car-following-models at controlled intersections," Tech. Rep., 2017. [Online]. Available: <https://elib.dlr.de/115720>.
- [19] R. Wiedemann and U. Reiter, "Microscopic traffic simulation: The simulation system mission, background and actual state," *Project ICARUS (V1052) Final Report*, vol. 2, pp. 1–53, 1992.
- [20] A. Sharma, Z. Zheng, and A. Bhaskar, "Is more always better? The impact of vehicular trajectory completeness on car-following model calibration and validation," *Transportation research part B: methodological*, vol. 120, pp. 49–75, 2019. DOI: [10.1016/j.trb.2018.12.016](https://doi.org/10.1016/j.trb.2018.12.016).
- [21] M. Schrader and J. Bittle, "Extracting vehicle trajectories from partially overlapping roadside radar," Manuscript in Preparation.
- [22] K. Jo, M. Lee, and M. Sunwoo, "Track fusion and behavioral reasoning for moving vehicles based on curvilinear coordinates of roadway geometries," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 9, pp. 3068–3075, 2017.
- [23] J. Huang, Z. He, Y. Arakawa, and B. Dawton, "Trajectory planning in frenet frame via multi-objective optimization," *IEEE Access*, 2023. DOI: [10.1109/ACCESS.2023.3294713](https://doi.org/10.1109/ACCESS.2023.3294713).
- [24] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with markovian switching coefficients," *IEEE transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988. DOI: [10.1109/9.1299](https://doi.org/10.1109/9.1299).
- [25] K. Jo, M. Lee, J. Kim, and M. Sunwoo, "Tracking and behavior reasoning of moving vehicles based on roadway geometry constraints," *IEEE transactions on intelligent transportation systems*, vol. 18, no. 2, pp. 460–476, 2016. DOI: [10.1109/TITS.2016.2605163](https://doi.org/10.1109/TITS.2016.2605163).
- [26] R. Altendorfer and S. Wirkert, "Why the association log-likelihood distance should be used for measurement-to-track association," in *2016 IEEE intelligent vehicles symposium (IV)*, IEEE, 2016, pp. 258–265. DOI: [10.1109/IVS.2016.7535395](https://doi.org/10.1109/IVS.2016.7535395).
- [27] S. Julier and J. K. Uhlmann, "General decentralized data fusion with covariance intersection," in *Handbook of multisensor data fusion*, CRC Press, 2017, pp. 339–364.

- [28] W. Nielsen, "Information fusion based on fast covariance intersection filtering," in *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002.(IEEE Cat. No. 02EX5997)*, IEEE, vol. 2, 2002, pp. 901–904. DOI: [10.1109/ICIF.2002.1020907](https://doi.org/10.1109/ICIF.2002.1020907).
- [29] D. Franken and A. Hupper, "Improved fast covariance intersection for distributed data fusion," in *2005 7th International Conference on Information Fusion*, IEEE, vol. 1, 2005, 7–pp. DOI: [10.1109/ICIF.2005.1591849](https://doi.org/10.1109/ICIF.2005.1591849).
- [30] D. R. King, A. Gold, K. B. Phillips, and D. A. Krauss, "Headway times on urban, multiple lane freeways," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, vol. 66, 2022, pp. 1225–1229. DOI: [10.1177/1071181322661413](https://doi.org/10.1177/1071181322661413).
- [31] S. P. Venthuruthiyil and M. Chunchu, "Vehicle path reconstruction using recursively ensembled low-pass filter (relp) and adaptive tri-cubic kernel smoother," *Transportation Research Part C: Emerging Technologies*, vol. 120, p. 102 847, 2020. DOI: [10.1016/j.trc.2020.102847](https://doi.org/10.1016/j.trc.2020.102847).
- [32] X. Hu, Z. Zheng, D. Chen, X. Zhang, and J. Sun, "Processing, assessing, and enhancing the waymo autonomous vehicle open dataset for driving behavior research," *Transportation Research Part C: Emerging Technologies*, vol. 134, p. 103 490, 2022. DOI: [10.1016/j.trc.2021.103490](https://doi.org/10.1016/j.trc.2021.103490).
- [33] V. Punzo, Z. Zheng, and M. Montanino, "About calibration of car-following dynamics of automated and human-driven vehicles: Methodology, guidelines and codes," *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103 165, 2021. DOI: [10.1016/j.trc.2021.103165](https://doi.org/10.1016/j.trc.2021.103165).
- [34] V. Punzo, B. Ciuffo, and M. Montanino, "Can results of car-following model calibration based on trajectory data be trusted?" *Transportation research record*, vol. 2315, no. 1, pp. 11–24, 2012. DOI: [10.1017/S1041610212001457](https://doi.org/10.1017/S1041610212001457).
- [35] J. Rapin and O. Teytaud, *Nevergrad - A gradient-free optimization platform*, <https://github.com/facebookresearch/nevergrad>, 2018.
- [36] D. Salles, S. Kaufmann, and H.-C. Reuss, "Extending the intelligent driver model in sumo and verifying the drive off trajectories with aerial measurements," *SUMO Conference Proceedings*, vol. 1, pp. 1–25, Jul. 2022. DOI: [10.52825/scp.v1i.95](https://doi.org/10.52825/scp.v1i.95).
- [37] R. Salter and R. Salter, "Headway distributions in highway traffic flow," *Highway Traffic Analysis and Design*, pp. 107–124, 1974.
- [38] D. A. Iancu and N. Trichakis, "Pareto efficiency in robust optimization," *Management Science*, vol. 60, no. 1, pp. 130–147, 2014. DOI: [10.1287/mnsc.2013.1753](https://doi.org/10.1287/mnsc.2013.1753).
- [39] H. Lu, G. Song, and L. Yu, "The "acceleration cliff": An investigation of the possible error source of the vsp distributions generated by wiedemann car-following model," *Transportation Research Part D: Transport and Environment*, vol. 65, pp. 161–177, 2018. DOI: [10.1016/j.trd.2018.08.004](https://doi.org/10.1016/j.trd.2018.08.004).
- [40] M. Zhu, X. Wang, A. Tarko, et al., "Modeling car-following behavior on urban expressways in shanghai: A naturalistic driving study," *Transportation research part C: emerging technologies*, vol. 93, pp. 425–445, 2018. DOI: [10.1016/j.trc.2018.06.009](https://doi.org/10.1016/j.trc.2018.06.009).

## Appendix: Calibrated CF-Model Parameters

The IDM, Krauss and W99 were calibrated with two objectives:  $\text{NRMSE}(s, v)$  and  $\text{NRMSE}(s, v, a)$ . The below tables, Table 3, 4, & 5 represent the results from calibration, with the mean ( $\mu$ ), standard deviation ( $\sigma$ ), and the 10th, 50th, and 90th ( $P_{10-90th}$ ) percentiles being presented. Also shown are the SUMO default values for the passenger

car class, with the value being given as the  $\mu$ . The lower and upper bounds of the optimization constraints are also shown. Certain parameters were further constrained to be multiples of the simulation step size, such as `actionStepLength`.

**Table 3.** Calibrated results of the IDM CF-model. LB stands for the lower-bound on optimization and UB the upper bound. All parameters are in the naming convention and units presented in the [SUMO documentation](#)

Parameter	Calibration		Cal. Target	Calibrated Parameters				
	LB	UB		$\mu$	$\sigma$	$P_{10\%}$	$P_{50\%}$	$P_{95\%}$
accel	0.1	6.0	Default	2.60	-	-	-	-
			$\text{NRMSE}(s, v, a)$	2.63	1.27	1.13	2.45	5.04
			$\text{NRMSE}(s, v)$	2.95	1.38	1.28	2.78	5.45
actionStepLength <sup>a</sup>	0.1	1.0	Default	0.10	-	-	-	-
			$\text{NRMSE}(s, v, a)$	0.33	0.25	0.10	0.20	0.90
			$\text{NRMSE}(s, v)$	0.43	0.27	0.10	0.40	0.90
decel	0.1	7.0	Default	4.50	-	-	-	-
			$\text{NRMSE}(s, v, a)$	2.73	1.86	0.56	2.41	6.35
			$\text{NRMSE}(s, v)$	2.66	1.87	0.44	2.39	6.19
delta	1.0	10.0	Default	4.00	-	-	-	-
			$\text{NRMSE}(s, v, a)$	4.75	2.54	1.44	4.57	9.21
			$\text{NRMSE}(s, v)$	4.99	2.59	1.48	4.92	9.43
minGap	0.1	10.0	Default	2.50	-	-	-	-
			$\text{NRMSE}(s, v, a)$	4.12	2.34	1.05	3.97	8.60
			$\text{NRMSE}(s, v)$	4.10	2.41	0.98	3.87	8.66
speedFactor	0.8	1.8	Default	1.00	-	-	-	-
			$\text{NRMSE}(s, v, a)$	1.27	0.27	0.89	1.26	1.71
			$\text{NRMSE}(s, v)$	1.27	0.27	0.89	1.26	1.73
stepping <sup>a</sup>	0.1	1.0	Default	0.25	-	-	-	-
			$\text{NRMSE}(s, v, a)$	0.55	0.23	0.23	0.55	0.94
			$\text{NRMSE}(s, v)$	0.54	0.23	0.22	0.54	0.92
tau	0.1	5.0	Default	1.00	-	-	-	-
			$\text{NRMSE}(s, v, a)$	1.87	1.07	0.67	1.69	4.01
			$\text{NRMSE}(s, v)$	1.89	1.05	0.69	1.68	4.00

<sup>a</sup>Constrained in optimization to a multiple of the simulation step size

**Table 4.** Calibrated results of the Krauss CF-model. LB stands for the lower-bound on optimization and UB the upper bound. All parameters are in the naming convention and units presented in the [SUMO documentation](#)

Parameter	Calibration		Cal. Target	Calibrated Parameters				
	LB	UB		$\mu$	$\sigma$	$P_{10\%}$	$P_{50\%}$	$P_{95\%}$
accel	0.1	7.0	Default	2.60	-	-	-	-
			$\text{NRMSE}(s, v, a)$	2.63	1.36	1.10	2.39	5.20
			$\text{NRMSE}(s, v)$	3.49	1.56	1.61	3.35	6.30
actionStepLength <sup>a</sup>	0.1	1.0	Default	0.10	-	-	-	-
			$\text{NRMSE}(s, v, a)$	0.40	0.28	0.10	0.30	0.90
			$\text{NRMSE}(s, v)$	0.42	0.27	0.10	0.40	0.90
decel	0.1	7.0	Default	4.50	-	-	-	-
			$\text{NRMSE}(s, v, a)$	3.70	1.56	1.63	3.75	6.36
			$\text{NRMSE}(s, v)$	3.81	1.52	1.77	3.82	6.48
sigma	0.1	1.0	Default	0.50	-	-	-	-
			$\text{NRMSE}(s, v, a)$	0.34	0.20	0.12	0.30	0.75
			$\text{NRMSE}(s, v)$	0.51	0.22	0.21	0.51	0.91
sigmaStep	0.1	1.0	Default	0.10	-	-	-	-
			$\text{NRMSE}(s, v, a)$	0.49	0.23	0.18	0.49	0.89
			$\text{NRMSE}(s, v)$	0.50	0.22	0.19	0.49	0.89
speedFactor	0.8	1.8	Default	1.00	-	-	-	-
			$\text{NRMSE}(s, v, a)$	1.30	0.26	0.94	1.30	1.73
			$\text{NRMSE}(s, v)$	1.30	0.26	0.94	1.31	1.73
tau	0.5	5.0	Default	1.00	-	-	-	-
			$\text{NRMSE}(s, v, a)$	2.10	1.10	0.85	1.87	4.41
			$\text{NRMSE}(s, v)$	1.77	1.06	0.66	1.51	4.01

<sup>a</sup>Constrained in optimization to a multiple of the simulation step size



**Table 5.** Summary statistics for the calibrated W99 model. LB stands for the lower-bound on optimization and UB the upper bound. All parameters are in the naming convention and units presented in the [SUMO documentation](#)

Parameter	Calibration		Cal. Target	Calibrated Parameters				
	LB	UB		$\mu$	$\sigma$	$P_{10\%}$	$P_{50\%}$	$P_{95\%}$
actionStepLength <sup>a</sup>	0.1	1.0	Default	0.10	-	-	-	-
			NRMSE( $s, v, a$ )	0.41	0.27	0.10	0.40	0.90
			NRMSE( $s, v$ )	0.45	0.28	0.10	0.40	1.00
cc1	0.0	5.0	Default	1.30	-	-	-	-
			NRMSE( $s, v, a$ )	1.69	0.98	0.54	1.56	3.57
			NRMSE( $s, v$ )	2.01	1.19	0.66	1.79	4.59
cc2	0.0	10.0	Default	8.00	-	-	-	-
			NRMSE( $s, v, a$ )	5.99	2.35	2.78	6.05	9.65
			NRMSE( $s, v$ )	4.45	2.49	1.04	4.41	8.82
cc3	-20.0	0.0	Default	-12.00	-	-	-	-
			NRMSE( $s, v, a$ )	-11.26	4.69	-17.59	-11.40	-3.01
			NRMSE( $s, v$ )	-10.88	4.85	-17.59	-10.86	-2.48
cc4	-5.0	0.0	Default	-0.25	-	-	-	-
			NRMSE( $s, v, a$ )	-1.81	1.25	-3.65	-1.62	-0.12
			NRMSE( $s, v$ )	-2.24	1.32	-4.12	-2.16	-0.19
cc5	0.1	5.0	Default	0.35	-	-	-	-
			NRMSE( $s, v, a$ )	1.97	1.33	0.36	1.72	4.52
			NRMSE( $s, v$ )	2.31	1.25	0.63	2.29	4.55
cc6	0.1	20.0	Default	6.00	-	-	-	-
			NRMSE( $s, v, a$ )	7.03	4.94	1.14	6.28	16.76
			NRMSE( $s, v$ )	8.24	5.25	1.39	7.94	17.96
cc7	-1.0	1.0	Default	0.25	-	-	-	-
			NRMSE( $s, v, a$ )	-0.10	0.47	-0.74	-0.11	0.73
			NRMSE( $s, v$ )	-0.04	0.50	-0.73	-0.04	0.84
cc8	0.0	8.0	Default	2.00	-	-	-	-
			NRMSE( $s, v, a$ )	2.38	1.89	0.52	1.72	6.62
			NRMSE( $s, v$ )	3.65	2.10	0.95	3.60	7.36
cc9	0.0	8.0	Default	1.50	-	-	-	-
			NRMSE( $s, v, a$ )	3.98	2.05	1.14	3.94	7.50
			NRMSE( $s, v$ )	4.02	2.03	1.21	4.05	7.44
minGap	0.0	20.0	Default	2.50	-	-	-	-
			NRMSE( $s, v, a$ )	6.18	4.30	1.00	5.53	14.38
			NRMSE( $s, v$ )	5.30	4.14	0.70	4.39	13.28
speedFactor	0.8	1.5	Default	1.00	-	-	-	-
			NRMSE( $s, v, a$ )	1.12	0.20	0.84	1.12	1.44
			NRMSE( $s, v$ )	1.12	0.19	0.85	1.13	1.44

<sup>a</sup>Constrained in optimization to a multiple of the simulation step size