Challenges of Microsimulation Calibration with Traffic Waves using Aggregate Measurements

Sadman Ahmed Shanto
Department of Physics and Astronomy
Texas Tech University
2500 Broadway
Lubbock, TX 79409
sadman-ahmed.shanto@ttu.edu

George Gunter (corresponding author)
Department of Civil and Environmental Engineering
Institute for Software Integrated Systems
Vanderbilt University
1025 16th Avenue South
Nashville, TN 37212
george.l.gunter@vanderbilt.edu

Rabie Ramadan
Department of Mathematics
Temple University
1805 N. Broad Street
Philadelphia PA 19122, USA
rabie.ramadan@temple.edu

Benjamin Seibold
Department of Mathematics
Temple University
1805 N. Broad Street
Philadelphia PA 19122, USA
seibold@temple.edu

Daniel B. Work
Department of Civil and Environmental Engineering
Institute for Software Integrated Systems
Vanderbilt University
1025 16th Avenue South
Nashville, TN 37212

0 words + 2 table = 6063 words

ABSTRACT
This work explores the challenges associated with calibrating parameters of microscopic models with aggregate speed data, e.g., obtained from roadside sensors. Using the Intelligent Driver Model, we explore how reliably parameters that do not influence the equilibrium flow (i.e., the Fundamental Diagram), but do control the stability of those equilibria, can be determined from aggregate speed data. Using a carefully controlled computational setup, we show that standard loss functions used for calibrating microsimulation models can perform poorly when the true parameters result in an unstable traffic state. Precisely, it is found that all of the considered loss functions frequently return different and incorrect parameter sets that minimize the expected value of the loss function. These results highlight the need for improved loss functions, or even fundamental additions to the model calibration procedure.

INTRODUCTION
The calibration of models from measured data is a core problem to many transportation engineering applications, particularly when examining models that describe the flow of vehicular traffic. In general, calibration is concerned with finding numerical values for otherwise unknown parameters within a model, with the goal that the model becomes capable of reproducing recorded data to an acceptable degree. While vehicular traffic can be described on various scales, including large-scale macroscopic models [1, 15, 16, 19, 20, 25, 34], here we focus on microscopic models [2, 18, 21] that describe the interactions between nearby vehicles via systems of ordinary differential equations (ODEs).

Microscopic calibration commonly means that a certain model structure is postulated with a handful of free parameters that are to be fitted so that the model reproduces available measurement data suitably well. Specifically, the "best fit" parameters are determined as the solution of an optimization problem

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad L(Y_{\text{real}}, Y_{\text{sim}}(\theta, \lambda)) \tag{1}$$

where $\boldsymbol{\theta}$ are the free decision variables to be determined, $\boldsymbol{\lambda}$ are (non-free) hyper-parameters that are known a-priori, Y_{real} are the measurement data, Y_{sim} are the corresponding data generated by the simulated model under a given parameter choice, and L is a loss function that defines a suitable distance between data and model prediction. The data itself can be either macroscopic measurements (e.g., from roadside sensors such as inductive loops or radar units), microscopic data (e.g., collected from individual vehicles with GPS devices or on-board sensors); or a combination of microscopic and macroscopic data. In this work we restrict to car-following calibration, i.e., we do not consider perimeters associated with origin destination calibration or lane changing logic.

Real traffic flow is known to exhibit (in certain flow regimes) instabilities and nonlinear waves, and certain microscopic models reproduce this behavior [2, 12]. This paper studies the fundamental question of to what extent the calibration problem (1) can reliably and robustly identify the decision variables that govern instability and waves behavior for models that do have the capability of exhibiting such features. To that end, synthetic data are generated from a realizable model, the Intelligent Driver Model (IDM) [12, 31], and a number of commonly used loss functions are systematically investigated. We focus on calibration using macroscopic data, and assess the potential (or lack thereof) to calibrate microscopic models potentially containing instabilities.

Foundations and related work

The difficulty of the task of calibrating a traffic model depends critically on what scales and features of the real flow should be resolved. In real traffic, three scales must be distinguished: (i) the vehicle scale, on which heterogeneities across vehicles/drivers matter; (ii) the waves scale, on which non-equilibrium phenomena manifest, such as nonlinear traffic waves; and (iii) the fully macroscopic scale on which non-equilibrium effects average out. Given that commonly available data do not resolve vehicle-specific information, models usually aim to capture traffic on scale (ii) or on scale (iii). For the latter, simple first-order models or fully stable car-following models suffice, and the task reduces to calibrating the fundamental diagram (FD) of traffic flow, i.e., a function that relates the flow rate, q, vs. the vehicle density, ρ. In contrast, for the former, in addition to calibrating for an averaged FD, one also needs to calibrate for instabilities and waves. It is this situation that this paper explicitly considers.

Traffic waves are fundamental features of highway traffic flow. Traffic models that reproduce them do so via instabilities at uniform flow that grow into nonlinear traveling waves; both in car-following [2, 12] and macroscopic [6] models. The fact that traffic waves can arise via dynamic instabilities from uniform flow has been demonstrated experimentally [29]; and understanding these waves is of practical interest because of their adverse effects on flow efficiency, fuel economy, and emissions [27, 35], and the impact of vehicle automation [28].

In car-following models, such as the IDM, traffic waves arise when a uniform flow state fails to be "string stable", i.e., a vehicle's velocity perturbation generates a larger perturbation on the vehicle that follows. As shown below, two parameters in the IDM govern the model's stability properties.

Traffic data frequently come in aggregated form. Here we consider data from traditional stationary detectors, or cameras on highways that directly count passing vehicles and their speeds, and from those recover *average* flow rates and vehicle densities (density = flow/speed) over short time intervals (30 sec to 5 min) [5, 10]. If considered without temporal ordering, these data points can be used to form a FD cloud, and a suitably fitted function can be obtained as the FD curve $q = Q(\rho)$ [24]. The presence of instabilities and waves, however, will produce a spread in the FD data (cf. [26]), and one may attempt to employ the temporal information encoded in the aggregated data to obtain information about these non-equilibrium features, or here: the model parameters that shape them.

With regards to that last aspect, the aggregation time of the measurements is important: traveling waves are known to exist on scales as short as 230m [29], and an aggregation time of 5 min would completely average them out. In contrast, an aggregation time of 30 sec (which is not an uncommon practical choice) would conduct some averaging, but still retain a temporal signature of waves. Thus, we consider precisely that latter aggregation time in this study.

Calibration of traffic models to data can come in many facets. Here, we consider the problem of calibrating a second-order car-following model (all vehicles identical) to aggregate measurements. The calibration task of constructing a FD function $q = Q(\rho)$ from aggregated data is reasonably well established (see [4] and references therein), and proper choices of measurement locations, observables, aggregation times, and parameter choices have been established [7, 22]. Hence, for this study, we make the simplifying assumption that the model parameters determining the FD have already been calibrated and thus are known (without error). In contrast, the task of calibrating car-following models explicitly for non-equilibrium features, particularly via aggregated data, is far from well-established. Some specific effort has been done in [14, 33], with characteristics of waves and the regions they form employed to inform the model parameters. However, due to the complexity of instabilities and nonlinear waves, there is no general procedure; and this study aims to highlight some fundamental challenges incurred with the calibration problem in models and data with waves.

It is also worth stressing that the related, but different, task of micro-to-micro calibration, i.e., determining model parameters based on trajectory data, is significantly better understood [11, 13, 23, 30]. Specifically, the robust calibration of the IDM has been studied based on genetic algorithms [11] and based on parameter reduction [23].

Contributions and paper outline

The main contribution of this work is to highlight intrinsic challenges incurred with calibrating car-following model parameters from aggregated data, in the presence of instabilities. Focusing on the IDM this study specifically zooms in on the question how well the model parameters responsible for the strength of instability and waves can be determined. To facilitate a clean problem formulation, the IDM itself is used to generate a time series of aggregated speed measurements under a set of known "true" parameters. Given these measurements, we then explore the inverse problem to recover these "true" parameters. Using a Monte Carlo approach and on a fixed grid in parameter space, we demonstrate that commonly used loss functions for (1) are not reliable indicators (i.e., they are not minimal) of the true model parameters. We illustrate that the challenges appear even in the most simplified settings, i.e., on single lane roadways when the inflow and outflow conditions are known.

The remainder of the article is organized as follows. In the Section *Model Specifications* we review the IDM, its equilibrium properties, and string stability. In the Section *Methods*, the setup for the computational experiments is laid out. The Section *Results* contains the findings of the numerical experiments, highlighting the challenges of calibrating the IDM model using only macroscopic data. Finally, the *Conclusions* Section highlights potential next steps for investigation.

MODEL SPECIFICATIONS

The different components of the simulation environment are discussed in this section. We briefly review the IDM car-following model used in the numerical experiments presented later. We also isolate the parameters corresponding to equilibrium features (and correspondingly the fundamental diagram), and describe how to determine the regions in which the model is string stable.

Intelligent Driver Model

In order to describe the trajectories of individual vehicles, each vehicle is modeled via an ordinary differential equation that either describes the vehicle velocity (first order models), or the velocity and acceleration (second-order models). Second-order car-following models are of the form

$$\dot{v}(t) = f(\boldsymbol{\theta}, s(t), v(t), \Delta v(t)) \ , \tag{2}$$

where $f(\boldsymbol{\theta}, s(t), v(t), \Delta v(t))$ models the acceleration of the vehicle at time t. Here s represents the spacing between the ego vehicle and the vehicle ahead, v the speed of the vehicle, and Δv is velocity gap between the vehicle and the vehicle ahead (which is also the rate of change in the space gap, or the negative approach rate). The vector $\boldsymbol{\theta}$ contains the parameters that characterize the behavior of the model. The IDM [31] is a special case of (2), and reads

$$f(\boldsymbol{\theta}, s, v, \Delta v)_{\text{IDM}} = a\left[1 - \left(\frac{v}{v_0}\right)^{\delta} - \left(\frac{s^*(v, \Delta v)}{s}\right)^2\right], \tag{3}$$

where $s^*(v, \Delta v)$ is defined as

$$s^*(v, \Delta v) = s_0 + vT + \frac{\max\{0, v\Delta v\}}{2\sqrt{ab}} \ . \tag{4}$$

The IDM has six parameters $\boldsymbol{\theta} = [a, b, v_0, T, \delta, s_0]$. The parameter $v_0 > 0$ represents the desired empty road velocity, and $s_0 > 0$ represents the minimum desired spacing between vehicles. In addition, $T > 0$ is the desired time headway, which is the minimum possible time to reach the vehicle ahead, and δ is called the acceleration exponent, which is usually set to $\delta = 4$ [17]. The parameters a and b are both positive and measured in m/s^2, and they correspond to the maximum vehicle acceleration and minimum comfortable deceleration, respectively. In fact, the last four parameters in $\boldsymbol{\theta}$ (i.e., v_0, T, δ, s_0) determine the FD of the traffic flow. In contrast, the parameters a and b do no affect the FD, but they affect the traffic flow dynamics via stability and waves.

To allow instabilities to manifest, the model (2) is augmented by a noise term (Gaussian with zero mean and standard deviation σ), i.e., we actually integrate the stochastic differential equation (written in derivative form)

$$\dot{v}_{\text{IDM}}(t) = f_{\text{IDM}}(\boldsymbol{\theta}, s(t), v(t), \Delta v(t)) + \mu(0, \sigma) \ . \tag{5}$$

Determining parameters that influence the fundamental diagram

The fact that a and b do not determine the shape of the (equilibrium) FD [31] can be found by deriving an equation to define the equilibria for the IDM. This is done by solving $\dot{v}_{\text{IDM}} = f_{\text{IDM}}(\boldsymbol{\theta}, s, v, 0) = 0$ for s in terms of v. Put differently, this means finding the equilibrium spacing function $s_{eq}(v)$, that gives an equilibrium spacing value for a given speed. For $a > 0$, the equilibrium spacing function reads

$$s_{eq}(v) = \sqrt{\frac{s_0 + vT}{1 - \left(\frac{v}{v_0}\right)^{\delta}}} \ , \tag{6}$$

and its inverse is the equilibrium velocity function $v_{eq}(s)$. Realistic car-following behavior is generally assumed to require that $s_{eq}(v)$ and its inverse $v_{eq}(s)$ are strictly increasing functions. For the IDM, this is indeed the case when $v < v_0$, for any admissible parameter choices of $\boldsymbol{\theta}$. As a consequence, all equilibrium states can be parametrized by a single state variable. Moreover, with the vehicle density $\rho = 1/s + \ell$, where ℓ is the vehicle length, we also obtain the FD function $Q(\rho) = v_{eq}(1/\rho) * \rho$. As $v_{eq}(s)$ depends only on the parameters s_0, T, v_0, δ (see (6)), the FD depends only on the same parameters as well.

Stability of the IDM equilibria

To study the stability of a given equilibrium state, a linear stability analysis is usually employed [3]. First, consider vehicles on a single-lane road with positions $x_i(t)$, where vehicle i follows vehicle $i - 1$, and the motion of all vehicles is described by the car-following model (2). Second, equation (2) is linearized around an equilibrium state, by choosing the position of vehicle i to be $x_i(t) = (s_{eq} + \ell)i + v_{eq}t + y_i$, where y_i is an infinitesimal perturbation. Then, substituting these vehicle positions into (2), Taylor-expanding around the equilibrium state, and keeping only the linear terms, the perturbation equation is obtained

$$\ddot{y}_i(t) = \alpha_1(y_{i-1} - y_i) - \alpha_2\dot{y}_i + \alpha_3\dot{y}_{i-1} \ , \tag{7}$$

$$\text{where} \qquad \alpha_1 = \frac{\partial f}{\partial s} \ , \qquad \alpha_2 = \frac{\partial f}{\partial(\Delta v)} - \frac{\partial f}{\partial v} \ , \qquad \alpha_3 = \frac{\partial f}{\partial(\Delta v)} \ , \tag{8}$$

and all the partial derivatives are evaluated at the equilibrium state in consideration. Then, the growth/decay behavior of solutions to (7) is characterized by performing a Laplace transform ansatz $y_i(t) = c_i e^{\omega t}$, where $c_i, \omega \in \mathbb{C}$. This yields to the interpretation of (7) as an input/output (I/O) system, $c_i = F(\omega)c_{i-1}$, with the transfer function

$$F(\omega) = \frac{\alpha_1 + \alpha_3\omega}{\alpha_1 + \alpha_2\omega + \omega^2} \ . \tag{9}$$

$\mathrm{Re}(\omega)$ and $\mathrm{Im}(\omega)$ in equation (9) represent, respectively, the temporal growth/decay and the frequency of oscillation of the lead vehicle's velocity profile. The transfer function's modulus $|F|$ is the growth/decay of the perturbation amplitude from one vehicle to the next.

With this setup, string stability means that $|F(\omega)| \leq 1 \ \forall \omega \in i\mathbb{R}$.

This stability criterion can be written as a condition on the partial derivatives of f, as follows:

$$\alpha_2^2 - \alpha_3^2 - 2\alpha_1 \geq 0 \ . \tag{10}$$

Hence, an equilibrium state's stability is determined by the partial derivatives of $f(\boldsymbol{\theta}, s, v, \Delta v)$ with respect to the state variables at that equilibrium state.

As discussed above, the parameters a and b of the IDM are not involved in determining the shape of the FD, but they do play a critical role in determining whether an equilibrium state is stable or unstable. In Figure 1, two FDs are generated under the same set of parameters (v_0, T, δ, s_0), but different a and b. The equilibrium points in which the model is string unstable are marked in red, while stable regions are marked in blue. The figure shows that the choice of $a = 0.7$ m/s^2 and $b = 1.5$ m/s^2 leads to instabilities at higher flow-rates than $a = 1.4$ m/s^2 and $b = 1.0$ m/s^2.

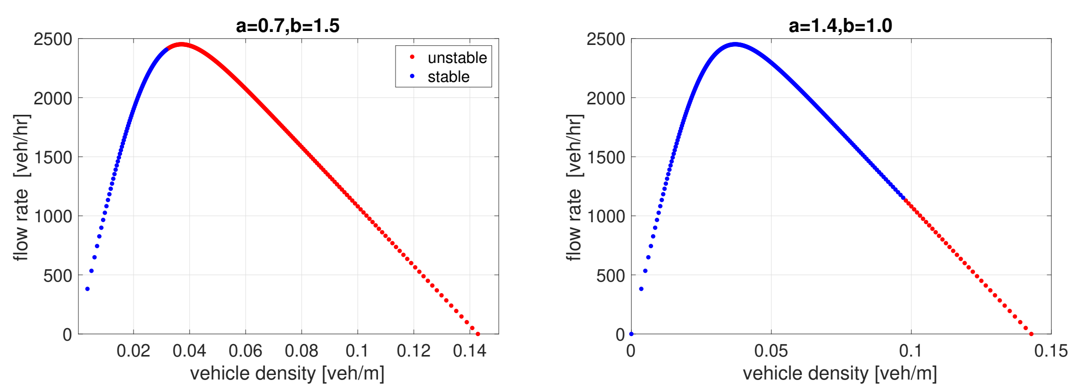

FIGURE 1 : Comparison between the string stability regions for two different choices of a and b in the IDM.

FIGURE 2 : A graphic representing the road geometry combined with a single radar sensor generating aggregate speed measurements.

EXPERIMENTAL SETUP

To illustrate the challenges of calibration using macroscopic data, we set up a test environment under carefully controlled settings. Recalling that all parameters in the IDM except a and b control the shape of the fundamental diagram, we assume those parameters are known (estimated) from historical aggregate data.

Instead, we focus on understanding the choices of a and b, the two parameters that cannot be estimated from equilibrium data. Note that even though a and b do not affect equilibrium states (and thus the FD shape), they *do* change the non-equilibrium behavior and the stability of the equilibria. As a result it is in principle possible to see changes in macroscopic/aggregate measurements that are a result of changes in these microscopic parameters.

In the experiments below, we fix the following equilibrium parameters according to nominal values reported in [12]:

$$\begin{aligned} v_0 &= 30 \text{ m/s} , \\ s_0 &= 2 \text{ m} , \\ T &= 1 \text{ s} , \\ \delta &= 4 . \end{aligned}$$

The magnitude of the additive noise (see (5)) is fixed at $\sigma = 0.1$ m/s^2. In order to numerically solve the simulations, a ballistic integration method, as described in [32], is employed at a step size of 0.4 seconds, with a simple Euler-Maruyama treatment of the noise term.

Network geometry

To isolate the effects of a and b, we consider a highly structured setup. A single lane road segment is considered to remove the potential confounding effects of lane changing or routing logic as specified through Origin/Destination demand data. An overview of the setup is shown in Figure 2. The road is the domain $x \in [0, 2100\text{m}]$, and a single sensor is located at $x = 500$m. The sensor reports the average speed of vehicles passing the sensor, aggregated over 30 second increments. These values were chosen to allow for suitable space for waves to develop and be sensed.

Traffic flow conditions

The traffic is loaded onto the network with a free-flowing inflow rate of 2250 vehicles per hour. The outflow rate is restricted to generate congestion, resulting in an outflow rate of 1600 vehicles per hour. The simulation is run for 1800 seconds (30 min), but only the final 750 seconds are used for the study to avoid the influence of the simulation warm up or the propagation of the traffic

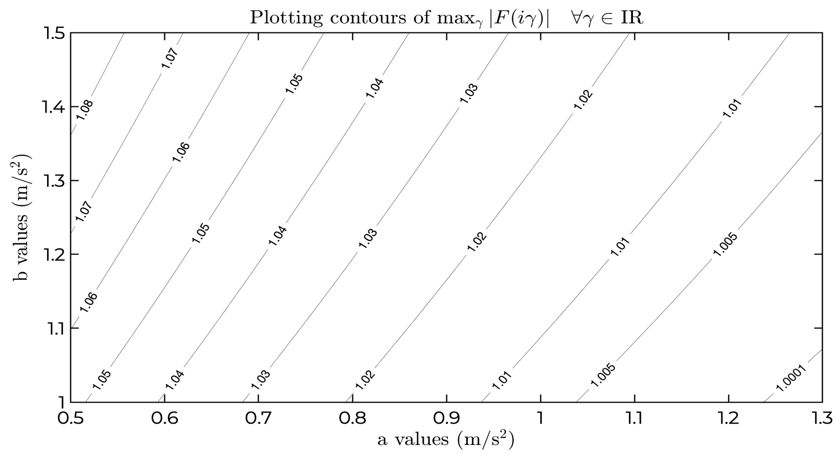

FIGURE 3 : Contour plot showing the growth rate of waves with respect to parameters a and b under the experimental setting. A growth rate above 1 indicates string instability.

congestion conditions to the sensor location. As a result, the generated sensor data are contained entirely in a congested traffic state. Under these traffic conditions, the stability of traffic can be computed following the analysis presented in the Section *Model Specifications*. The resulting wave growth rate is shown for a range of parameter values for a and b in Figure 3.

METHODS

Optimization approach

Given the simplified experimental setting described above, the resulting calibration problem stated in (1) amounts to $\boldsymbol{\theta}$ containing only a and b, which can then be compared to the true a and b used to generate the data. In order to isolate the consequences of the loss function from the consequences of the optimization solver, here we adopt a brute force parameter sweep solution approach (which is costly but it eliminates any error in the optimization procedure itself). Namely we consider solving the optimization problem on a fixed grid in the (a, b) parameter space, by varying the parameters in increments of 0.1 within ranges of $[0.5, 1.3]$ and $[1.0, 1.5]$ respectively (with units m/s^2, for simplicity omitted here and below). As we will illustrate in the Results section, the loss function hinders the ability to correctly calibrate the model, even when solved via a brute force approach.

Loss functions

Numerous loss functions have been proposed in the literature to determine a parameter set that best matches (in the measurement space) the observed data. An excellent review of the use of these loss functions can be found in [8]. The definitions of the considered functions are given in Table **??**. Because some loss functions operate on the data point-wise, we summarize the notation used in the loss functions. Let $Y_{real} \in \mathbb{R}^N$ denote the vector of the true (or real) sensor data, and $Y_{sim} \in \mathbb{R}^N$ is the simulated sensor data under a given parameter set. The notation $Y(i)$ represents the i-th element (or measurement) from the data set of length N.

A loss function is needed to compare two sets of measurements, in this case to define the distance between the true (real) data set, and one that is simulated under a candidate set of parameters. Given that the model in question is stochastic in nature (and may exhibit instabilities that amplify perturbations), it is not the case that two different simulation runs that use the same

Loss Function Name	Abbreviation	Function Definition
Mean error	L_{ME}	$\frac{1}{N}\sum_{i=1}^{N}(Y_{sim}(i) - Y_{real}(i))$
Mean normalized error	L_{MNE}	$\frac{1}{N}\sum_{i=1}^{N}\frac{Y_{sim}(i)-Y_{real}(i)}{Y_{real}(i)}$
Root mean normalized squared error	L_{RMSNE}	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_{sim}(i)-Y_{real}(i)}{Y_{real}(i)}\right)^2}$
Mean absolute normalized error	L_{MANE}	$\frac{1}{N}\sum_{i=1}^{N}\frac{\lvert Y_{sim}(i)-Y_{real}(i)\rvert}{Y_{real}(i)}$
Squared sum error	L_{SSE}	$\sum_{i=1}^{N}(Y_{sim}(i) - Y_{real}(i))^2$
Root mean squared error	L_{RMSE}	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{sim}(i) - Y_{real}(i))^2}$
Mean absolute error	L_{MAE}	$\frac{1}{N}\sum_{i=1}^{N}\lvert Y_{sim}(i) - Y_{real}(i)\rvert$
Thiel's inequality coefficient	L_U	$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{real}(i)-Y_{sim}(i))^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{real}(i))^2}+\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{sim}(i))^2}}$

TABLE 1 : A summary of the loss functions considered in this study.

parameter set will return the exact same time series of measurements, even under the highly structured simulation setting considered in this work. This means that for a given parameter, one might record a non-zero loss value between two measurement time series corresponding to the exact same parameter set.

In order to account for this, ensemble averages of multiple simulations must be considered. Specifically, it is generally recommended that a number of simulations are repeated on a given parameter set and that the collection of simulations are used, are used rather than just single simulation [9]. This involves evaluating the loss function to compare each simulation under the same parameters Y_{sim} to the single Y_{real} time series, and then minimizing the sample average of the loss function evaluations.

In particular, to evaluate the loss function under a given parameter set $\boldsymbol{\theta}$, suppose a total of M simulations are conducted. Then, the effective loss function between the simulated data for this $\boldsymbol{\theta}$ and real data is given by

$$\hat{L}\left(Y_{real}, Y_{\boldsymbol{\theta}}\right) = \frac{1}{M}\sum_{j=1}^{M} L(Y_{real}, Y_{\boldsymbol{\theta}}^{j}) \ , \tag{11}$$

where $L(Y_{real}, Y_{\boldsymbol{\theta}}^{j})$ is one of the loss functions summarized in Table 1 above, evaluated between the real data and the j-th set of simulated data (of the total M simulations) at the parameter choice $\boldsymbol{\theta}$. From now onward, the hats will be omitted for notational simplicity, and L will refer to the average of M loss functions evaluated using M simulated data sets that share the same $\boldsymbol{\theta}$.

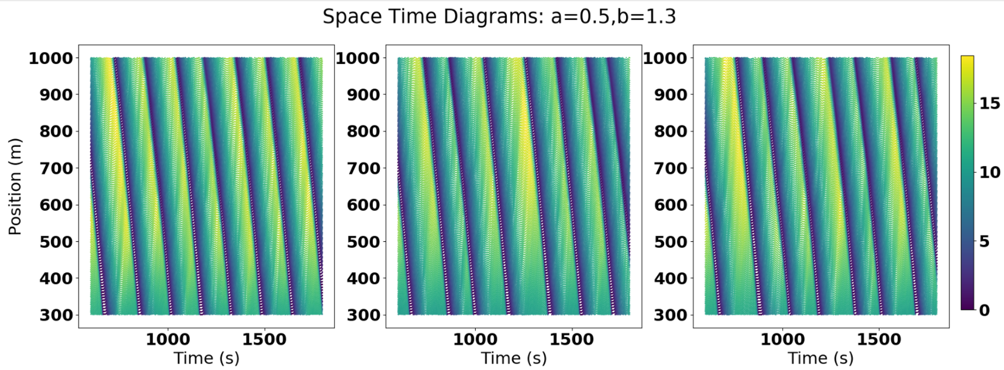

FIGURE 4 : Three time space diagrams colored by speed in (m/s) produced from identical simulations except for the random seed, with $(a, b) = (0.5, 1.3)$. Waves are present and small variations occur in the phase and amplitude of the waves.

RESULTS

This section details the computational experiments conducted on the IDM under the experimental settings defined in the *Experimental Setup* Section. The parameters of the IDM that can be determined from the Fundamental Diagram of a roadway are fixed, while a parameter sweep on a and b is conducted that performs multiple simulations at each parameter set in order to account for stochasticity in the simulation. Several loss functions are then evaluated on this data set by using a hold-out simulation for each parameter set considered as a “true” measurement and then comparing the loss for each parameter set to these hold-outs. It is found that all loss functions considered have limitations in performance in terms of returning lowest expected losses at the true parameter values used to make a hold-out measurement set, suggesting that currently utilized objective functions in micro-model calibration incur fundamental challenges when applied to calibrating traffic in which instabilities are present.

Investigation of the influence of stochastic forcing and model instability

Given the stochastic and linearly unstable nature of the microsimulation model in question, simulation runs for the exact same set of parameters may return different measurement values when run multiple times (which is representative of unstable systems in reality). A consequence of this is that no single simulation run is completely representative of a choice of parameters. Since the goal of calibration is to choose an optimal set of parameter values, one must be able to compare the performance of a candidate parameter choice both back to a set of true measurements and to other parameters.

A demonstration of possible variation across simulations for the same parameter set can be seen in Figure 4. Here three space-time plots are shown, all generated using the same set of parameters, $(a, b) = (0.5, 1.3)$. All three simulations are conducted under the exact same set of model parameters but with different random seeds. The waves that are present have similar shape, but no two of the simulations are exactly the same, nor are the resulting measurement values. As a result, a loss function that calculates a difference between a set of recorded measurements (what is being calibrated for) and a set of simulated measurements has the potential to return different evaluations depending on the simulation run.

In comparison to Figure 4, Figure 5 shows the same setup: three different space-time

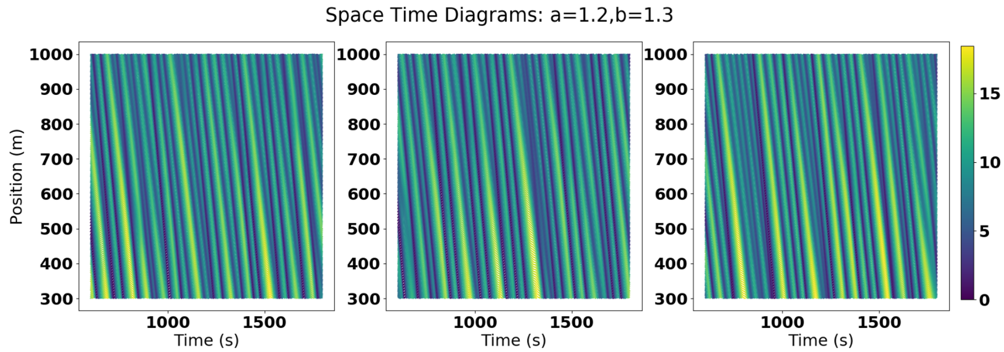

FIGURE 5 : Three time space diagrams colored by speed in (m/s) produced from identical simulations except for the random seed, with $(a, b) = (1.2, 1.3)$. Waves are present and small variations occur in the phase and amplitude of the waves.

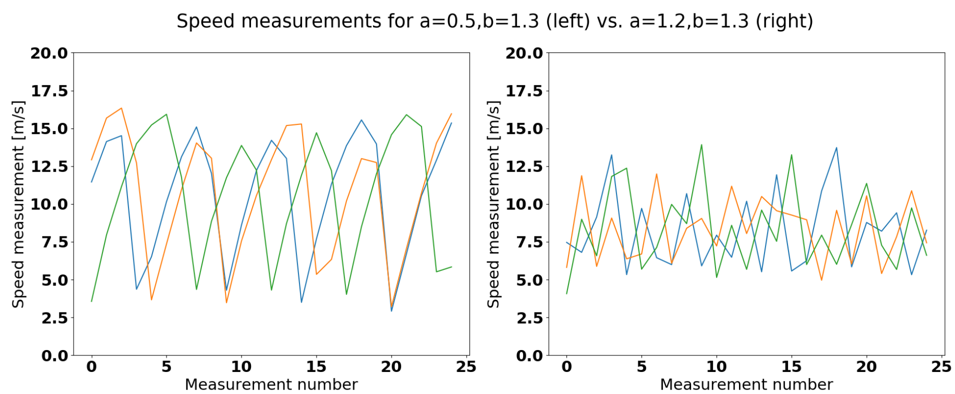

FIGURE 6 : Illustration of the time series measurements recorded for the three simulations under $(a, b) = (0.5, 1.3)$ (left) and $(a, b) = (1.2, 1.3)$ (right).

diagrams from three simulations, except now with $(a, b) = (1.2, 1.3)$. Again the same intra-parameter variation in the wave formation is observable. It is worth noting however, that the waves that form from this parameter set are distinct in behavior from those in Figure 4, occurring more frequently and with lower magnitude. This result suggests that while certain variations happens across simulations for the same parameter choice, there are more fundamental differences in the simulation results across different parameters.

The core of the calibration problem is to exploit the differences between the parameters, as manifested through the measurement data, so that the correct parameters can be found. Figure 6 shows how the variations across simulations with the same parameter sets result in variations in the measurement data time series. Under the two parameter choices, notable differences are present again both within each parameter set across multiple simulations, and also between different parameter sets.

Investigation of the RMSE loss function to recover true parameters

We next consider the distribution of the loss function when evaluated under the same and distinct parameter settings. The setup is as follows. We first generate the true measurement data by a single run of the IDM under the true parameters $(a, b) = (0.5, 1.3)$. Next we conduct 50

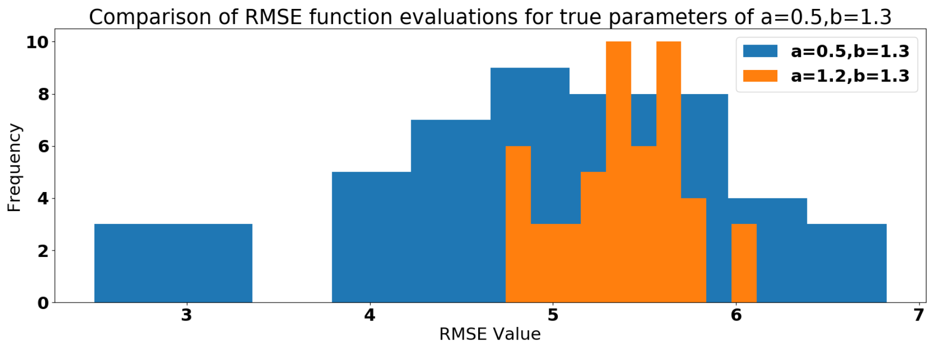

FIGURE 7 : Histogram of RMSE loss function evaluations comparing a holdout Y_{sim} generated under the true parameters $(a, b) = (0.5, 1.3)$. 50 simulations are conducted to generate 50 sets Y_{sim} timeseries each under $(a, b) = (0.5, 1.3)$ (blue), and $(a, b) = (1.2, 1.3)$ (orange). The sample average is lower under the true parameters than the incorrect ones (sample average RMSE 4.91 compared to sample average RMSE 5.38).

simulations under the same IDM parameters (i.e., $(a, b) = (0.5, 1.3)$). For each run, we evaluate the RMSE loss function to quantify the consistency of the measurement data with the true data, and the resulting histogram is shown in Figure 6 (blue). Next, we evaluate the RMSE loss function with 50 additional simulations under incorrect parameters (i.e., $(a, b) = (1.2, 1.3)$), with the resulting RMSE histogram shown in orange in Figure 6. The sample average RMSE under the true parameters (4.91) is lower than the sample average RMSE under the incorrect parameters (5.38). When used as a loss function for calibration, we could correctly rule out the incorrect parameter set from the sample average RMSE. However, the fact that there are realizations of the loss function for the incorrect parameter set that are lower (better) than the loss function realizations for the correct parameter set illustrates the importance of using multiple runs to evaluate the loss function. Moreover, the fact that the loss function under the true parameters has a wide variance indicates that a large number of samples might be necessary to obtain an accurate estimate.

The analysis so far has considered the distribution of the loss function when evaluated on a single incorrect parameter set, compared to an evaluation on the correct parameter set. We now repeat the analysis but for all 54 parameter pairs in the discrete search space. Again, given a true set of measurements Y_{real} generated from a single run of the IDM under $(a, b) = (0.5, 1.3)$ we evaluate the sample average of the loss function for all parameter pairs in the search space. This requires running 50 simulations for each of the 54 parameter pairs, extracting Y_{sim} for each run, and computing the loss function.

Here we find a discouraging result, illustrated in Figure 8. The objective function evaluation, which is the expected loss function value, is shown in blue, while the minimum and maximum loss function evaluations are shown in yellow and green, respectively. Those metrics are shown in red specifically for the set corresponding to $\boldsymbol{\theta_{true}}$. All values are presented in sorted order according to their expected loss function evaluation. Of the 54 parameter pairs, 42 pairs resulted in a lower sample RMSE than the sample RMSE when evaluated at the true parameters. In other words, in the setting where one minimizes the sample expected loss in order to determine the values of (a, b), the optimization solver can easily find an incorrect but smaller (in sample average loss) parameter pair. Figure 8 also shows the minimum and maximum of the loss function evaluations for the 54

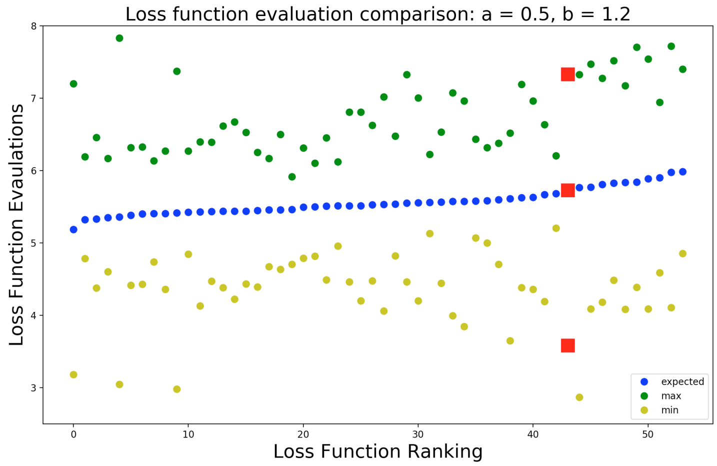

FIGURE 8 : Loss function evaluations using L_{RMSE} and a true parameter set of (a, b)=(0.5, 1.2) are shown for every parameter set, sorted by order of expected loss.

parameter pairs, highlighting that other statistical measures beyond the sample average (e.g., the sample minimum), will not circumvent the problem.

Investigation of the RMSE loss function to recover true parameters: sensitivity to the true parameters

To further expand on the issue raised in the previous section, we next consider if the challenge is localized for the single choice of the true parameters $(a, b) = (0.5, 1.3)$, or if the problem appears for other choices of the true parameters. We remind the reader (see Figure 3) that the search space of parameters corresponds to string unstable traffic in the controlled numerical settings considered here, so all true parameter sets will generate instabilities. For each of the 54 points in the parameter search space, we first generate a true measurement data set Y_{real} under those true parameters. Then we repeat the analysis above. Namely, we evaluate, under each of the 54 parameter pairs and using 50 simulations for each pair, the loss function distribution and its sample average. We then check to see if the lowest sample average loss function corresponds to the true parameters.

We quantify the rate of failure through the *Point-wise Percentage Failure* (PPF). The PPF for a given $\boldsymbol{\theta_{true}}$ is how many other parameter sets return lower expected loss values normalized by the number of parameter sets considered in total. For example, when the sample average loss function is lowest on the true parameter, the PPF is 0%, while if the sample average loss is highest on the true parameter set, the PPF is 100%. The PPF when the true parameters are $(a, b) = (0.5, 1.2)$ (Figure 8) is 42/54 = 78%.

Figure 9 plots the PPF for each true parameter pair in the parameter space. Some parameter pairs have low PPF, while others have PPF in excess of 80%. The large variance in the performance of L_{RMSE} suggests that for certain zones of parameters it may perform better or worse in terms of convergence to $\boldsymbol{\theta_{true}}$.

To better understand how much error occurs when minimizing the sample average loss function when the incorrect parameters are recovered, two other metrics are calculated. The *Point-wise Divergence in a* (PD_a) and the *Point-wise Divergence in b* (PD_b) determine the average difference on the parameter values. Figure 10 and Figure 11 plot these measures as functions of the

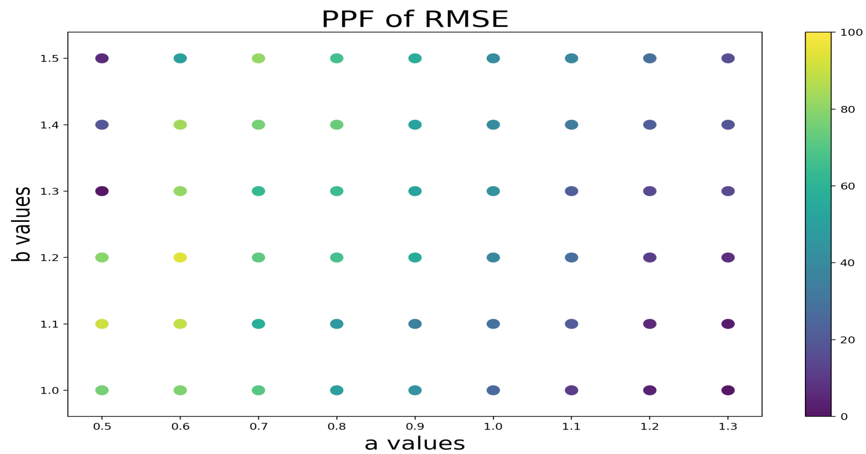

FIGURE 9 : The RMSE loss function is evaluated by calculating the percentage of points with a lower objective function score than the evaluation on the true parameter set.

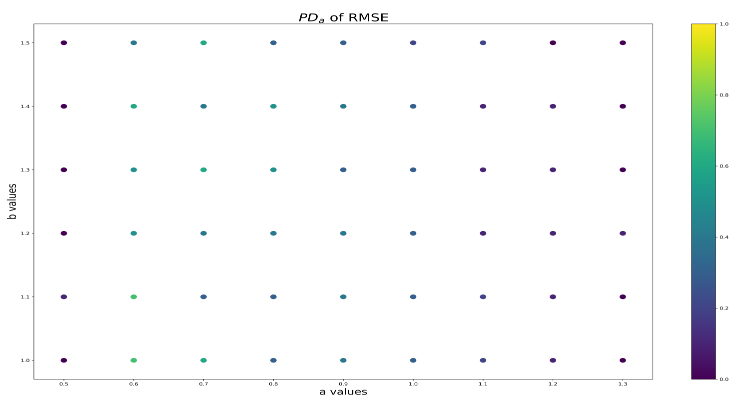

FIGURE 10 : Average Divergence in the parameter a using L_{RMSE} as the loss function.

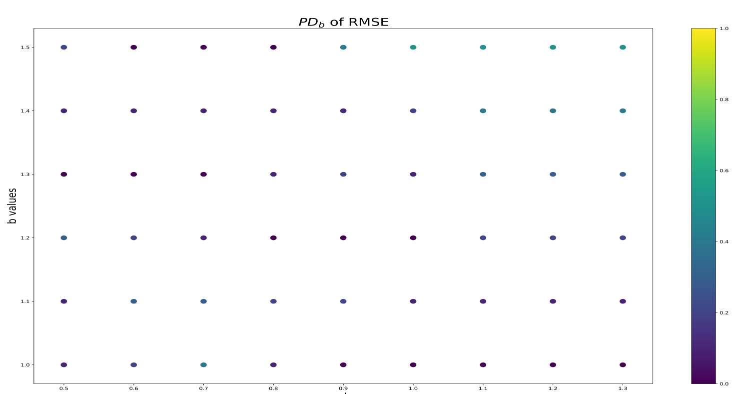

FIGURE 11 : Average divergence in the parameter b using L_{RMSE} as the loss function.

Loss Function	Average % Failure	Average Divergence in a	Average Divergence in b
ME:	49.1	0.40	0.25
MNE:	49.0	0.40	0.25
RMSNE:	47.5	0.39	0.24
MANE:	47.1	0.37	0.24
SSE:	44.4	0.28	0.20
RMSE:	43.5	0.26	0.17
MAE:	42.1	0.24	0.19
U:	31.4	0.19	0.18

TABLE 2 : Reporting of three different error metrics on each candidate loss function. All loss functions are found to have similar and high degrees of error in their performance.

true parameters. From the figures it is clear that the largest PD_a are realized for small true a values. Generally the PD_b values are lower than the corresponding differences in a. The consequences of these differences depends on the goal of the simulation, but these results indicate that the difference on a is likely to be higher than on b.

Another important observation is that the curves of comparable PPF in Figure 9 tend to exhibit some interesting qualitative agreements with the growth rate contours shown in Figure 3 (except for a-values towards the lower boundary). This observation appears to indicate that there is some fundamental connection between the (lack of) robustness of the calibration problem and the strength of waves growth.

Comparison of Loss Functions

So far the analysis has been restricted to a single loss function, namely the RMSE. Table **??** shows the aggregate performance scores for each loss function across all parameter sets. Each of PPF, PD_a, and PD_b are averaged across all a, b pairs and reported as a total score.

From the summary statistics it is apparent that the challenges observed with the RMSE loss function are also present in the majority of other loss functions. The average failure rates across all parameters and for all loss functions are in the range of 31%–49%. Similarly the average divergence values range from 0.18-0.40 on a, and 0.17-0.25 on b. While small performance improvements are observed depending on the loss function used, no loss function has overall excellent performance. The best performing loss function is *Thiel's inequality coefficient* (U) scores the lowest of all considered loss functions in both average PPF and average PD_a, while only scoring behind L_{RMSE} in average PD_b, suggesting it may be the best of all loss functions considered for the calibration task proposed.

CONCLUSIONS

Motivated by a desire to calibrate car-following model parameters from aggregated traffic measurements that cannot be found through standard techniques, a comprehensive review of the suitability of several commonly used loss functions is performed using a simulation environment. Each loss function is assessed on its ability to recover known parameter values that were used to create synthetic measurements, where well-performing loss functions would return a globally lowest value at the parameter set used to create those measurements. Despite the advantageous test configuration (with data generated from the model, and equilibrium parameters known), it is nevertheless found that out of all loss functions considered, none have satisfactory performance, meaning: for many parameter sets used to create measurements a different parameter set returns a lower loss value. In three different metrics of performance for a given candidate loss function, *Thiel's inequality coefficient* performs the best in two, and second best in a third, suggesting it may be the best choice out of those considered. In general however, the poor performance of these standard loss functions suggests that better loss function formulations past what is currently employed are needed in order to reliably calibrate microscopic models from aggregate data; or even more: loss functions that explicitly distill fundamental characteristics of traffic waves that arise in the unstable regime of traffic.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant CNS-1837652, and work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Vehicle Technologies Office award number CID DE-EE0008872. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[1] A. Aw and M. Rascle. Resurrection of second order models of traffic flow. *SIAM J. Appl. Math.*, 60:916–944, 2000.

[2] M. Bando, Hesebem K., A. Nakayama, A. Shibata, and Y. Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Phys. Rev. E*, 51(2):1035–1042, 1995.

[3] S. Cui, B. Seibold, R. Stern, and D. B. Work. Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1336–1341. IEEE, 2017.

[4] W. Daamen, C. Buisson, and S.P. Hoogendoorn. *Traffic Simulation and Data: Validation Methods and Applications*. CRC Press, 2014.

[5] C. C. De Wit, F. Morbidi, L. L. Ojeda, A. Y. Kibangou, I. Bellicot, and P. Bellemain. Grenoble traffic lab: An experimental platform for advanced traffic monitoring and forecasting [applications of control]. *IEEE Control Systems Magazine*, 35(3):23–39, 2015.

[6] M. R. Flynn, A. R. Kasimov, J.-C. Nave, R. R. Rosales, and B. Seibold. Self-sustained nonlinear waves in traffic flow. *Phys. Rev. E*, 79(5):056113, 2009.

[7] G. Gomes, A. May, and R. Horowitz. A microsimulation model of a congested freeway using VISSIM. *TRB 2004 Annual Meeting*, 11 2003.

[8] Y. Hollander and R. Liu. The principles of calibrating traffic microsimulation models. *Transportation*, 35(3):347–362, 2008.

[9] Yaron Hollander and Ronghui Liu. The principles of calibrating traffic microsimulation models. *Transportation*, 35(3):347–362, May 2008.

[10] Z. Jia, C. Chen, B. Coifman, and P. Varaiya. The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors. In *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings*, pages 536–541. IEEE, 2001.

[11] A. Kesting and M. Treiber. Calibrating car-following models by using trajectory data: Methodological study. *Transportation Research Record*, 2088(1):148–156, 2008.

[12] A. Kesting, M. Treiber, and D. Helbing. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4585–4605, 2010.

[13] J. Kim and H. S. Mahmassani. Correlated parameters in driving behavior models: Car-following example and implications for traffic microsimulation. *Transportation Research Record*, 2249(1):62–77, 2011.

[14] V. Kurtc and M. Treiber. Calibrating the local and platoon dynamics of car-following models on the reconstructed NGSIM data. In *Traffic and Granular Flow '15*, Cham, 2016. Springer.

[15] J.-P. Lebacque. Les modeles macroscopiques du traffic. *Annales des Ponts.*, 67:24–45, 1993.

[16] M. J. Lighthill and G. B. Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proc. Roy. Soc. A*, 229(1178):317–345, 1955.

[17] R. Malinauskas. The intelligent driver model: Analysis and application to adaptive cruise control. Master's thesis, Clemson University, 2014.

[18] G. F. Newell. Nonlinear effects in the dynamics of car following. *Operations Research*, 9:209–229, 1961.

[19] H. J. Payne. Models of freeway traffic and control. *Proc. Simulation Council*, 1:51–61, 1971.

[20] H. J. Payne. FREEFLO: A macroscopic simulation model of freeway traffic. *Transp. Res. Rec.*, 722:68–77, 1979.

[21] L. A. Pipes. An operational analysis of traffic dynamics. *Journal of Applied Physics*, 24:274–281, 1953.

[22] V. Punzo and B. Ciuffo. How parameters of microscopic traffic flow models relate to traffic dynamics in simulation: Implications for model calibration. *Transportation Research Record*, 2124(1):249–256, 2009.

[23] V. Punzo, M. Montanino, and B. Ciuffo. Do we really need to calibrate all the parameters? Variance-based sensitivity analysis to simplify microscopic traffic flow models. *Intelligent Transportation Systems, IEEE Transactions on*, 16:184–193, 02 2015.

[24] X. Qu, S. Wang, and J. Zhang. On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. *Transportation Research Part B: Methodological*, 73:91–102, 2015.

[25] P. I. Richards. Shock waves on the highway. *Operations Research*, 4:42–51, 1956.

[26] B. Seibold, M. R. Flynn, A. R. Kasimov, and R. R. Rosales. Constructing set-valued fundamental diagrams from jamiton solutions in second order traffic models. *Netw. Heterog. Media*, 8(3):745–772, 2013.

[27] R. E. Stern, Y. Chen, M. Churchill, F. Wu, M. L. Delle Monache, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work. Quantifying air quality benefits resulting from few autonomous vehicles stabilizing traffic. *Transportation Research Part D: Transport and Environment*, 67:351–365, 2019.

[28] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmann, F. Wu, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work. Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. *Transportation Research Part C: Emerging Technologies*, 89:205–221, 2018.

[29] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-I. Tadaki, and S. Yukawa. Traffic jams without bottlenecks – Experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics*, 10(3):033001, 2008.

[30] E. Talbot, R. Chamberlin, B. A. Holmen, and K. M. Sentoff. Calibrating a traffic microsimulation model to real-world operating mode distributions. *TRB 93rd Annual Meeting Compendium of Papers*, 2014.

[31] M. Treiber, A. Hennecke, and D. Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical review E*, 62(2):1805, 2000.

[32] M. Treiber and V. Kanagaraj. Comparing numerical integration schemes for time-continuous car-following models. *Physica A: Statistical Mechanics and its Applications*, 419:183–195, 2015.

[33] M. Treiber and A. Kesting. Microscopic calibration and validation of car-following models — A systematic approach. *Procedia Social and Behavioral Sciences*, 80:922–939, 2013.

[34] R. Underwood. Speed, volume, and density relationships: Quality and theory of traffic flow. Technical report, Yale Bureau of Highway Traffic, 1961.

[35] F. Wu, R. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmann, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work. Tracking vehicle trajectories and fuel rates in oscillatory traffic. *Transportation Research Part C: Emerging Technologies*, 88:82–109, 2018.