IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 August 2022, accepted 17 October 2022, date of publication 3 November 2022, date of current version 11 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3219095

RESEARCH ARTICLE

Unveiling the Capacity Drop Phenomenon Through Traffic Microsimulation: Modeling, Calibration, and Sensitivity Analysis

FELIPE DE SOUZA[1] AND MARJAN MOSSLEMI[2]
[1]Energy Systems, Argonne National Laboratory, Lemont, IL 60439, USA
[2]Civil Engineering, University of California at Irvine, Irvine, CA 92697, USA

Corresponding author: Felipe de Souza (fdesouza@anl.gov)

This work was supported by the U.S. Department of Energy (DOE) Vehicle Technologies Office (VTO) under the Systems and Modeling for Accelerated Research in Transportation Mobility Laboratory (SMART) Consortium, an Initiative of the Energy Efficient Mobility Systems (EEMS) Program.

ABSTRACT The Capacity Drop (CD) is a sustained drop in a freeway bottleneck outflow when its upstream section becomes congested. Understanding how to replicate CD with microscopic models can help us to better understand its properties and mechanism. However, there is no published study replicating an observed occurrence of CD using microscopic models. In this study, we combine a lane-changing model with two widely known car-following models and calibrate the combined model parameters against field data, to replicate the CD in a merge bottleneck. We confirm that both car-following models can reproduce the CD. Moreover, we show that including the cooperative lane-changing behavior in the model is critical for the replication success. Further, we perform a sensitivity analysis with respect to the ramp demand and maximum acceleration parameters, showing: (i) slower acceleration rate leads to a lower bottleneck throughput, and (ii) higher ramp demand leads to decreased bottleneck throughput.

INDEX TERMS Capacity drop, microsimulation, IDM, gipps, calibration.

I. INTRODUCTION

Bottlenecks are critical elements of freeways as total delay and travel time are strongly dependent on the throughput of bottlenecks [1]. A key aspect of bottleneck throughput is capacity drop (CD) phenomenon - a drop in the bottleneck outflow when queues are formed at the upstream section of the bottleneck [2]. Understanding how to replicate CD with microscopic models can lead to a better understanding of this phenomenon, considering the level of fidelity these models.

Despite its importance, this phenomenon is often overlooked in common practice. In fact, the most common traffic flow models, such as the Lighthill-Witham-Richards (LWR) model [3], [4] and the discrete time models derived through different discretization techniques such as the cell transmission model [5]) are not able to reproduce the CD in their standard versions. However, there is a rich literature proposing model extensions (for a comparison among different models see [6]) which enable first order models to reproduce the CD. Additionally, second-order macroscopic models [7], [8], [9] can arguably reproduce this phenomenon inherently. Such potential of second-order models is often used as the key argument for simulating the traffic in situations the CD occurs (e.g., see [10]). Therefore, in one way or another, there is a rich literature on how to reproduce CD using macroscopic models.

However, the studies on CD from the microscopic modeling perspective is more scarce. Although in many cases applying macroscopic models is advantageous, it is also of interest to reproduce CD using microscopic models for two key reasons. First, modeling individual to individual interactions, which is hardly possible in macroscopic models, is essential for some applications such as adaptive cruise control), inter-vehicle communication (e.g., cooperative adaptive cruise control [11]). Second, an accurate microsimulation model can provide key insights into the

The associate editor coordinating the review of this manuscript and approving it for publication was Razi Iqbal.

underlying microscopic mechanism of CD. Specifically, two individual driving behavior aspects have been mentioned as contributing factors of CD: lane-changing [12], [13], [14] and driver heterogeneity [15], [16], [17]. Therefore, having accurate microsimulation models can help us to better understand CD in all its nuances.

Although scarce, there is a literature on microscopic models that directly or indirectly address CD. This includes the study by [18] analyzing the effect of drivers' relaxation on bottleneck capacity, and [19] replicating CD with the INTEGRATION software as a result of acceleration, lane-changing behavior, and fleet composition. Also, [20] replicated CD in Aimsun [21] by adjusting the model parameters, especially the acceleration and deceleration parameters. [22] applied a microscopic approach to investigate the impacts of two different driving behaviors, acceleration spread and reaction time, on the queue discharge rate when CD is observed. Also, [23] showed a reduction in the outflow at the onset of congestion consistent with the CD in their proposed sub-microscopic framework for modeling lateral movements.

Nevertheless, to the best of the authors' knowledge,[1] there is no published studies in the microscopic modeling literature that consistently reproduce CD of an actual case and informed by observed time-series data. In this context, the main contributions of this paper are as following:

- We show how to replicate CD at merge bottlenecks according to observed time-series data using widely known car-following [25], [26] and lane-changing models [27].
- We confirm the consistency of the results by replicating CD over different days including a validation data-set.
- We estimate the relative impacts of bounded acceleration and ramp demand on CD by performing a sensitivity analysis on the calibrated model.

The remainder of this paper is organized as follows. In Section II, we detail the simulation model used in this study. In Section III, we introduce the data, calibration method, and calibration results. In section IV, we present the results of the sensitivity analysis with respect to two key aspects of the model. Finally, in Section V, we state our conclusions.

II. SIMULATION MODEL

A schematic representation of mathematical notations applied in this paper is depicted in Figure 1. The position and speed of subject vehicle (S), respectively, is denoted as $x_S(t)$ and $v_S(t)$. Similar variables for the leading vehicle of vehicle (S), in its origin lane, are denoted as $x_L(t)$ and $v_L(t)$. Subject vehicle (S), aiming to make a lane-changing maneuver, possesses a potential following vehicle and leading vehicle in the target lane whose positions and speeds are denoted as $(x^\star_L(t), v^\star_L(t))$ and $(x^\star_F(t), v^\star_F(t))$ respectively. In addition, the right most lane is always referred as lane 1 and increasing for the left lanes. Finally, time t is discretized into fixed time steps of length Δt such that $t = i\Delta t$.

The model is a composition of car-following and lane-changing models. For car-following decisions, we apply two widely used car-following models in the literature, the Gipps [26] and the Intelligent Driver Model (IDM) [25]. With respect to lane-changing decisions, we use the lane-changing framework with explicit interactions by [27] which is a Gipps-like type of model [28], [29]. In the next sub-sections, we present the car-following and lane-changing models applied in this study.

A. CAR-FOLLOWING MODELS

We apply two distinct car-following models in this work: Gipps [26] and IDM [25]. While IDM is a continuous-time model, the Gipps model is formulated as an *iterative coupled map* [30] assuming the simulation time-step Δt and the reaction time τ to be the same. Since we intend to analyze the results assuming the same time steps (and $\Delta t < \tau$), we apply an interpolation method [30] to be able to compute the values of speed and position at any continuous time t. First, we assume Euler discretization:

$$x(t+\Delta t) = x(t) + \frac{1}{2}(v(t) + v(t+\Delta t))\Delta t \tag{1}$$

where speed $v(t)$ is defined for each specific car-following model. The position and speed for vehicle z within the time window $t - t_1$ is computed as:

$$z(t-t_1) = \beta z(t-(n+1)\Delta t) + (1-\beta) z(t-n\Delta t) \tag{2}$$

where

$$n = \lfloor \frac{t-t_1}{\Delta t} \rfloor, \qquad \text{and}$$
$$\beta = \frac{t-t_1}{\Delta t} - n \tag{3}$$

Therefore, whenever we need to compute the spacing or speed variable at a time between $t_1 = (i-1)\Delta t$ and $t_2 = i\Delta t$, we estimate based on the variables at time steps $i-1$ and i.

1) GIPPS CAR-FOLLOWING MODEL

Gipps car-following model [26] is one of the first behavioral car-following models. Gipps defines two speed regimes, one for deceleration regime $v^d_S(t)$ and another for acceleration regime $v^a_S(t)$, where vehicle speed v_S is updated as:

$$\begin{aligned}
v_S(t) &= \min\{v^d_S(t), v^a_S(t)\},\\
v^d_S(t) &= \max\{v^{d\star}_S(t), v_S(t-\Delta t) - B\Delta t\}\\
v^{d\star}_S(t) &= \Bigg(B^2(\frac{\tau}{2}+\theta)^2 + B\{2(x_L(t-\tau) - x_S(t-\tau)\\
&\quad -S_L) - \tau v_S(t-\tau) + \frac{v_L(t-\tau)^2}{\widehat{B}_L}\}\Bigg)^{1/2}\\
&\quad -B(\frac{\tau}{2}+\theta),\\
v^u_S(t) &= v_S(t-\tau) + 2.5A\tau\left(1 - \frac{v_S(t-\tau)}{V}\right)
\end{aligned}$$

[1] disregarding the authors' preliminary results [24]).

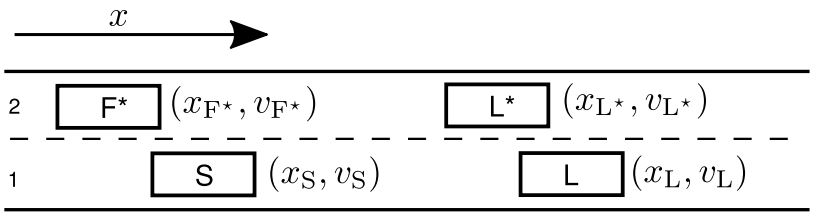


FIGURE 1. A schematic of variables defined at the neighborhood of the subject vehicle.

$$\times\sqrt{0.025+\frac{v_S(t-\tau)}{V}}, \tag{4}$$

where B is the maximum deceleration rate,A is the maximum acceleration rate, S_L is effective length of the leader, τ is the reaction time, θ is the safe margin, and $\widehat{B}_L$ is the maximum deceleration rate of the leading vehicle as estimated by the subject vehicle. These are all parameters to be estimated. The reasoning behind the deceleration speed $v_S^d(t)$ is that the subject vehicle keeps its spacing and speed so as to guarantee that will not hit the leading vehicle in the worst case scenario - braking at the maximum rate $\widehat{B}_L$, assuming the subject vehicle reacts after reaction time τ and brake at its maximum decceleration rate B. The second term of the max for determining $v_S^d(t)$ in Equation (4) operator is not present in the original Gipps model [26]. Unlike the IDM model which deceleration higher than the comfortable deceleration can be attained [31], the Gipps model is derived from the assumption B is the maximum deceleration and therefore it is necessary to explicitly enforce it. This term would not be necessary if the vehicle had been following the same leading vehicle at all times. However, after a tight lane-changing maneuver (we will provide details in Section II-B), the subject vehicle might become too close to the leading vehicle which would lead to $v_S^d(t)$ significantly lower than $v_S(t-\Delta t)-B\Delta t$. Hence, we added the second term to ensure the key behavior of the Gipps model which is to never decelerate at rates higher than the maximum deceleration rate, B.

2) INTELLIGENT DRIVER MODEL

The IDM [25] model is another widely used car-following model. Unlike the Gipps model, in the IDM there is no distinct equations for acceleration and deceleration regimes. On the other hand, likewise Gipps, IDM also has parameters A and B associated to acceleration and deceleration rates, although the decceleration parameter B has a slightly different meaning. In the IDM it is defined as the comfortable deceleration rate (unlike Gipps, there can be cases with deceleration greater than B). The IDM model computes acceleration as:

$$a=\dot{v}_S(t)=A\left[1-\left(\frac{v_S(t)}{V}\right)^{\delta}-\left(\frac{s^\star(t)}{s(t)}\right)^2\right],$$
$$s^\star(t)=\eta+v_S(t)t_h+\frac{v_S(t)(v_S(t)-v_L(t))}{2\sqrt{AB}}, \tag{5}$$

where t_h is desired time gap, η is jam spacing, and δ is free acceleration exponent. In IDM model, the equations are defined based on the rear to front bumper spacing $s(t)$ which, here, we assume as $s(t)=x_L(t)-x_S(t)-S_L$ where S_L is vehicle length. We note that unlike Gipps model, the IDM model allows deceleration rates higher than the comfortable deceleration [31] and therefore there is no need to enforce a maximum deceleration rate in the dynamic equations; however, this usually occurs at extreme cases. We also note that the maximum deceleration rate could be enforced on the calibration process to penalize a set of parameter that may lead to unrealistic deceleration.

B. LANE-CHANGING MODEL

In this study, we apply a framework for lane-changing model that conceptually follows [32] which follows the structure of lane-changing decisions by [29]. In this framework, lane-changing decisions are outcome of a rule-based framework that accounts for increasing the subject vehicle's utility (i.e., higher speed, avoid queue) while keeping its intended route. We first present the key elements of our applied lane-changing model; then, introduce the model of discretionary, mandatory and cooperative change of lanes.

For any lane-changing maneuver, the following gap acceptance rule must hold:

$$B^L=\frac{(v_2-v_1)^2}{x_1-x_2+\tau(v_2-v_1)-S_1}<B^{max},$$
$$X_1-X_2>g, \tag{6}$$

where g is the minimum gap to perform a change of lane, and B^L is the deceleration rate required for the lane-changing maneuver assuming the follower vehicle start decelerating after the reaction time τ is elapsed. The estimated deceleration rate B^L must be smaller than the maximum deceleration rate for the maneuver, B^{max}. The subscripts 1 and 2 on equation (6) refers to the leading and following vehicles, respectively, for any specific leader-follower combination. The first combination is the subject vehicle acting as a follower with respect to its potential leader at the target lane and for this case $x_2=x_L^\star(t)$, $v_2=v_L^\star(t)$, $x_1=x_S(t)$, and $v_1=x_S(t)$. The second combination is when the subject vehicle acts as a leader with respect to the follower at the target lane and for this case $x_2=x_S(t)$, and $v_2=x_S(t)$ and $x_1=x_F^\star(t)$, $v_2=v_F^\star(t)$. In summary, the subject vehicle is cognizant of the deceleration that both the subject vehicle and the potential follower need to undertake after the maneuver is performed and a change of lane can be performed if the deceleration required does not exceed B^{max} for both vehicles.

After a change of lane, we also model a relaxation time in which drivers avoid deceleration following a change of lane by accepting shorter spacing for a short of period of time and therefore undertaking speeds higher than the car-following model would prescribe [18], [33]. Denoting as $v_S^{CF}(t)$ and $v_L^{CF}(t)$ as the car-following speed of the subject and the leading vehicle respectively. During the relaxation time the car-following speed is overridden as:

$$v_S(t) = \max\{v_S^{CF}(t), v_L^{CF}(t), v_S(t - \Delta t) - B\Delta t\}, \quad (7)$$

which applies for both the vehicle performing the maneuver as well as to the potential (and henceforth effective) follower. The key aspect of Equation (7) is that both vehicles is aware of the the traffic ahead (i.e., the speed of the leader of leading vehicle) and therefore would be aware of the speed of their leader at current time $v_L^{CF}(t)$ and will not brake harder then that speed in the after that lane-changing. Notice that it also guarantees there will be no collision since the vehicle follow a speed no greater than its leader. Effectively, the relaxation avoids an unrealistic braking below the prevailing speed just downstream of the point in which the lane-changing is performed. In that case, the vehicle waits for the vehicle in front to increase its cruising speed to recover a comfortable spacing.

We specifically add the relaxation behavior since previous study has pointed out this specific behavior as an important aspects on merge bottlenecks [18] suggesting that the bottleneck may discharge at higher flow rate for a p than it would be predicted by the fundamental diagram capacity at that specific location during its activation.

We now present the models for discretionary, mandatory and, cooperative lane-changing.

1) DISCRETIONARY LANE-CHANGING

The discretionary lane-changing is performed for obtaining a speed advantage at the adjacent lanes. A vehicle considers a lane-changing maneuver if its current speed is lower than free-flow speed similar to [32]. However, the rule that governs a lane-changing decision uses lane by lane thresholds similar to [34].

The prevailing speed at the target lane l is denoted as v^1. We define thresholds $\Delta v_{1,2}$, $\Delta v_{2,3}$, ..., $\Delta v_{n-1,n}$ for left lane-changing maneuver and thresholds $\Delta v_{n,n-1}, \ldots, \Delta v_{2,1}$. A change of lane between lanes l_1, and l_2 occurs when:

1) the maneuver does not violate the gap acceptance constraint (6);
2) the prevailing speed at the target lane exceeds a lane-specific threshold $\Delta v_{l_1,l_2}$;
3) no other lane-changing was performed in the previous period of time, ω, and space, Δx.

The last item is introduced to discourage a sequence of lane-changing at the same position into the same lane. Although other vehicles might have been considering a change of lane, they may hold the maneuver until other lane-changing maneuver is completed.

2) COOPERATIVE LANE-CHANGING

Cooperative lane-changing refers to lane-changing maneuver with explicit interaction between the subject vehicle and its potential follower [32]. This approach is also referred to as courtesy merging [35], [36]. Here we present a lane-changing model implementation that is based on the model by [27] which was specifically developed for merging sections. This type of lane-changing is specifically necessary on congested traffic in which a gap respecting constraints (6) barely occurs either because the target lane is congested and there is no gap to perform the maneuver. This is the reason for an explicit interaction or a courtesy.

The formulation which we apply in this study has a few distinctions compared to [32]. First, the equation and rules are adapted to the car-following models assumed. Second, we assume homogeneous drivers and therefore we do not model the agressivity parameter. Finally, we also consider cooperative lane-changing for non mandatory lane-changing. We first present the mandatory lane-changing and then the extension for non mandatory lane-changing with explict interaction.

A given lane-changing is deemed mandatory when the subject the maneuver is strictly necessary to keep the intended route and the subject vehicle is less than 8 seconds away from that. In the case of merge bottlenecks, the focus of this study, mandatory lane-changing are required for vehicles coming from the on-ramp at the end of the acceleration lane.

When a mandatory lane-changing is required, the subject vehicle checks for gaps on the target lane in the same manner as a discretionary lane-changing and will perform the maneuver if there is an existent gap. Unlike the discretionary lane-changing, however, the maneuver does not require a threshold for speed advantage and neither would cancel the maneuver if previous changes of lane occurred in the intervals ω and Δx.

The explicit interaction occurs when the subject vehicle is not able to find an acceptable gap until a distance $L^\star$ from the limit in which the lane-changing must be performed. At this point the lane-changing maneuver occurs with direct communication between the subject vehicle and the potential follower. The subject vehicles and the potential follower undertake actions to complete the maneuver. Basically, both vehicles adjust their speeds so as to perform the change of lane safely.

Both vehicles consider a horizon of time in which they with to complete the maneuver, referred here as T. The subject vehicle, attempts to reach a reference position, $x_S^\star$ at time $t+T$ that is just behind the potential leader assuming the potential leader and computes a speed, $x_S^\star$ so as to reach that position:

$$\begin{aligned} x_S^\star(t) &= x_L^\star(t) + Tv_L^\star(t) - g(1+s), \\ v_S^\star(t) &= \frac{x_S^\star(t) - x_S}{T} \end{aligned} \quad (8)$$

where s is a relative safe margin for the maneuver (e.g., if $s = 1$ the vehicle will leave twice the minimum gap). Based on the reference position, the subject vehicle

adjust its speed as follows:

$$v_S(t) = \begin{cases} v_S^- - B^{lc}\Delta t & v_S^\star(t) < v_S^- - B^{lc}\Delta t, \\ v_S^- + A^{lc}\Delta t & v_S^\star(t) > v_S^- + A^{lc}\Delta t, \\ v_S^\star(t) & otherwise, \\ v_S^- = v_S(t - \Delta t) \end{cases} \quad (9)$$

In short, the subject vehicle keeps speed $v_S^\star(t)$ as long it does not violate maximum acceleration, A^{lc}, or maximum deceleration, B^{lc}, to complete the maneuver.

Likewise, the follower also adjusts the speed so as to leave enough gap for the subject vehicle to merge. However, the difference is that potentially the vehicle may be providing enough space for multiple vehicles (although typically it is only one). The potential follower sets reference position and speed as follows:

$$\begin{aligned} x_F^\star(t) &= \min\{x_{L^\star}(t) + Tv_{L^\star}(t) - g(n+s), x_M(t) \\ &\quad + Tv_M(t) - g(1+s)\}, \\ v_F^\star(t) &= \frac{x_F^\star(t) - x_S(t)}{T}, \end{aligned} \quad (10)$$

where n is the number of vehicles in which the follower at the target lane is providing a gap and the subscript M refers to the merging vehicle (if $n > 1$ refers to the most upstream vehicle). Therefore, the follower at the target lane attempts to open enough gap with respect to its current leader. Notice however it is also mindful of the current position and speed of the merging vehicle. Likewise the subject vehicle, the speed of the follower at target lane is updated as Eq. (9) in order to not violate maximum acceleration and deceleration rates.

At every time step, the merging vehicle checks if there is enough gap to perform the lane-changing maneuver according to (6). In congested traffic, in general the three vehicles - the subject vehicle, the potential follower and potential leader travels at similar speeds and therefore the deceleration required will be close to zero and typically the vehicle merges as soon as it has enough spacing.

Therefore, this lane-changing with explicit interaction ensures that all vehicles will be able to perform a lane-changing maneuver so as to keep their route as long as the distance $L^\star$ is long enough for the maximum acceleration and deceleration rates A^{lc} and B^{lc} during the maneuver.

The non-mandatory cooperative lane-changing occurs in similar situation as the mandatory lane-changing: drivers scans for gaps in the target lane, but it is not able to find one gap suitable for a change of lane because of the speed difference across different lanes. In the case of merge bottlenecks, the target lane for vehicles coming from the acceleration lane ends up to be much slower because it needs to accommodate all the ramp flow. Since they are traveling at speeds lower than free flow speed they attempt to perform discretionary lane-changing for speed advantage.

However, the speed at their current lane might be much slower than the speed at the left lane and therefore it would require a high deceleration rate for the potential follower a in the target (and faster) lane which can far exceed B^{max}. This situation would never dissolve due to the on-ramp traffic. The cooperative lane-changing is performed in this case so that vehicles at slower lanes are able to safely perform lane-changing.

The process for the maneuver execution follows the same equations as the mandatory case. The difference occurs on the condition that triggers the cooperative lane-changing. A subject vehicle S will perform a non-mandatory cooperative lane-changing when:

1) the ratio r between on target lane $v^l(t)$ and the subject vehicle $r = \frac{v^l(t)}{v_S}(t)$ remained higher than a specified ratio r^1 for a period of time higher than t_c;
2) no other cooperative lane-changing has been started in the in the previous interval ω^c and Δx^c in the same manner as the discretionary lane-changing.

If both conditions hold for that vehicle, an explicit interaction between the subject vehicle and its follower in the target lane will start following the same set of equations as the mandatory lane-changing. The only distinction on the execution between the non-mandatory and mandatory lane-changing is that a cooperative lane-changing may be cancelled after elapsed Δt^c while there is no maneuver cancellation in the mandatory lane-changing.

III. CALIBRATION RESULTS

Based on the modeling presented, we aim to set all the model parameters associated to car-following and lane-changing models so as to reproduce the dynamics of traffic flow at a merge bottleneck in all its aspects, especially the capacity drop phenomenon. To that end, we use data from a merge bottleneck in which occurrence of CD is consistently observed.

We use data from California Performance Measurement (PeMS) [37]. The area under study is a stretch of the US-101. A schematic of the area is depicted in Figure 2 (the relative distances between detectors are out of scale). From the demand station until the downstream station, the station codes at PeMS system are 413786, 401578, 401586, and 410099. The on-ramp demand is measured based on the flow of station 409883. Between the location in which the demand is introduced and the upstream station there are two on-ramps and one off-ramp in which the net count is added to the demand station. The area between upstream and downstream detectors is an active bottleneck in the morning peaks with high occupancies (congestion) being observed on the upstream, but not at the downstream detector. Therefore, the congestion arises just at the merge. The data is pre-processed for mutual count consistency based on the method in [38].

The input demands are introduced at the on-ramp and the demand stations. We use the upstream and downstream stations to compare the results with the field data. Each detector provides the counts $c_s(i)$ and occupancy $o_s^j(i)$ at fixed steps i for each lane $j \in \{1, 2, 3\}$. The raw data collected is available at 30 seconds intervals which we aggregate into two minutes intervals. For each two minute time step, the demand is spread following an evenly spread across the interval. This follows

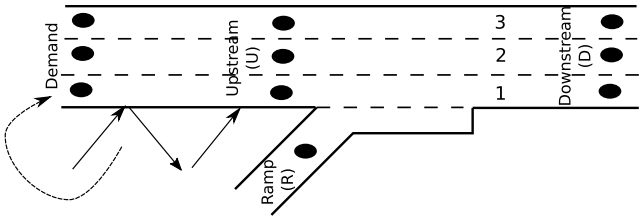


FIGURE 2. Schematic of the merge bottleneck under study.

the assumption on performing deterministic simulation in the calibration process. The Poisson arrival is a common choice, but it has been noted that in previous studies that this distribution is suitable for low traffic (under 500 veh/h) while there is no established arrival pattern for higher arrival rate [39].

The calibration approach aims to minimize the following two metrics:

$$J_1 = \frac{\sqrt{\sum_{i=1}^{I}\left(\sum_{j=1}^{3}(o_U^j(i) - \tilde{o}_U^j(i))^2\right)}}{3I},$$

(Occupancy Residual)

$$J_2 = \frac{\sqrt{\sum_{i=1}^{I}\left(N_D(i) - \tilde{N}_D(i)\right)^2}}{I},$$

(Cumulative Count Residual) where:

$$N_s(i) = \sum_{j=0}^{i} q_D(j), \quad \text{(Cumulative Count)}. \tag{11}$$

In general, related research prescribes metrics related to spacing [40], speed (e.g, [41], and flows (counts) (e.g., [42]). Ideally, a simulation model should reproduce consistently all the related metrics, but in general it is not possible. On the choice of objectives, here we distinct from previous work in two key aspects.

First, we employ a multiobjective approach as oppose to a single objective approach. This is different from using different terms in the objective function with prescribed weights. Rather, the two objectives are considered concomitantly. With this approach, we can show the trade-off between the two objectives and pick the solution which better balances the two objectives.

Second, as oppose to flow residual, we consider cumulative flow residual. One key reason is that typically CD is often analyzed through cumulative flows in oblique coordinates [43] (e.g., [2] and [44]). Among other reasons, the cumulative counts are less prone to noise and short term variations. Due to similar reasons, this curve is used to identify stationary states [45]. In addition, recent research suggesting the usage of cumulative variables for traffic flow models [46] as oppose to their derivatives (for example, cumulative flow as oppose to flow and position as oppose to speed). Previous and recent research had pointed out clearly the advantage of position with respect to speed [47], [48] for car-following models. Nevertheless, the same reasoning should apply to flow and cumulative flows. The rationale is that whenever a model calibrated against the cumulative variable (position or cumulative flow), it also lead to small errors for the variables themselves (speed, flow), but the opposite may not be true - a small speed error might be associated to large position errors for example.

Since microscopic models require several parameters to be calibrated, we perform the calibration based on the data of multiple days to increase the sample size. Specifically, we use data from 3/9/2017, 4/7/2017, 6/7/2017, and 6/19/2017. We calibrate both Gipps and IDM, each case along with the Lane-Changing parameters. Data from 2017 are used since we leveraged processed data from previous work aiming at reconcile inconsistent loop detector data [38]. Tables 1 presents all the parameter and the acceptable bounds of each variable for Gipps and IDM models as well as the Lane-Changing parameters. We keep fixed the minimum spacing for lane-changing (g) equals to 15 meters and $L^\star$ (distance for starting a mandatory lane-changing) equals to 240m and the slack for lane-changing of 0.5.

In total, for both IDM and Gipps model, there are 20 parameters to be calibrated. This is substantially more than any macroscopic model for a merge bottleneck in which the number of parameters is generally less than 10 (e.g., see [6]). Nevertheless, the increased number of parameter is common on microscopic simulation [49]. In addition, three of the aforementioned aspects on the calibration procedure is specifically suitable for this case. First, we are considering lane-by-lane occupancy data on the calibration at short time steps (2 minutes with each day amounts around 960 data samples - 240 downstream count and 720 occupancy observation.). Second, we are using data from multiple days for the calibration and validation which increase the amount of data for calibration. Third, the multiobjective approach can handle better than the single objective approach correlation between parameters (multiple parameter combination that yields similar or the same output value) since the output of the calibration are multiple parameter combinations as oppose to a single data point. We use this aspect of having multiple parameter combination in the sensitivity analysis presented on Section IV.

The goal of the calibration process is to assess whether it is possible to replicate CD through microscopic models and what are the key driving behavior to replicate it properly. Each of these questions are discussed in the following subsections.

A. CAN MICROSCOPIC MODELS REPLICATE THE CAPACITY DROP PHENOMENON?

We apply a multi-objective [50] extension of the Differential Evolution [51] algorithm for calibrating the model so as to replicate the observed data. For specific implementation details, we refer to [48]. The number of solutions (or population size N) was set to 50 and each calibration case was ran for 1000 iterations. For computing the metrics J_1 and J_2 we use moving average of 10 minutes for each measurement. Recall that counts and occupancies were aggregated in steps of 2 minutes.

TABLE 1. Model parameters and respective ranges.

Parameter	Model	Description	Unit	Range
V	Gipps	Desired speed	m/s	25-40
A	Gipps	Max. acceleration	m/s^2	0-5
B	Gipps	Max. deceleration	m/s^2	2-9
B_{avg}	Gipps	Max. deceleration	m/s^2	2-9
η	Gipps	Jam spacing	m	0-17
τ	Gipps	reaction time	s	0.5-2
A	IDM	Maximum acceleration	m/s^2	0-5
B	IDM	Comfortable deceleration	m/s^2	2-9
V	IDM	Desired Speed	m/s	25-40
δ	IDM	Free-flow exponent	-	0.2-20
t_h	IDM	Desired time gap	s	0-2.5
η	IDM	jam distance	m	0-17
A^{lc}	LC	Max. acceleration in LC	m/s^2	0-3
B^{lc}	LC	Max. deceleration in LC	m/s^2	0-3
$\Delta v_{1,2}$	LC	Speed threshold between lanes 1-2	m/s	0-10
$\Delta v_{2,3}$	LC	Speed threshold between lanes 2-3	m/s	0-10
Δv_{right}	LC	Speed threshold for right LC	m/s	0-10
γ	LC	Relaxation time after LC	s	0-2
ω	LC	minimum time for successive LC into lane	s	0-10
Δx	LC	minimum spacing for successive LC into lane	m	0-200
T	LC	Execution for LC maneuvers (see Eq. (9),(10))	s	0-2
Δt^c	LC	Cancellation time for cooperative LC	s	0-6
r^1	LC	Threshold ratio for coop. LC between lanes 1 an 2	-	1-50
r^2	LC	Threshold ratio for coop. LC between lanes 2 an 3	-	1-50
t^c	LC	Time elapsed with $v_i < r^i v_{i+1}$ to trigger coop. LC	s	0-6
ω^c	LC	minimum time for successive coop LC into lane	s	0-10
Δx^c	LC	minimum spacing for successive coop. LC into lane	m	0-200

Figure 3 depicts the Pareto Set for both car-following models with blue representing Gipps and Red the IDM model. Each point represents one solution (or element of the population) having a distinct value of each parameter of Table 1 and yielding objectives J_1 and J_2. Both models present similar behavior with a region of high occupancy error and low cumulative error that show moderate increase in cumulative flow error as the occupancy error decreases. For occupancy errors around 5%, the increase in cumulative flow error is increasingly steeper as occupancy decreases until reaching the minimum occupancy error at around 4%. Overall, both models present similar performance with a slight advantage of the IDM model.

We assessed the 50 different solutions and picked the one that qualitatively best reproduces the field data yielding a good trade-off between occupancy and cumulative flow error for both Gipps and IDM models. Figure 4 shows the detailed results for June, 6th, 2017. In all graphs, the observed data is depicted in black, the IDM simulated results in blue, and Gipps model in orange. The top left graph depicts the average occupancy over the three lanes while the top right graph shows specific occupancies for each lane with the top being the right most lane and at the bottom the left most lane. The bottom left graph depicts the downstream flow profile. Finally, the bottom right graph show the cumulative flow, referred to as T-curve, in oblique coordinates [2], [43] for baseline flow of 5400 veh/h. Generally both models resemble the observed data. With respect to occupancies, the IDM undershoots the aggregated occupancies just after 6AM (see top left graph). A closer look on the lane-based occupancies show that the key difference is the middle lane occupancy in

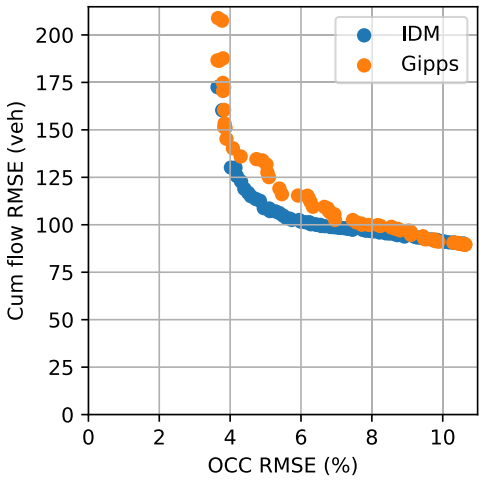


FIGURE 3. Pareto Set for IDM and Gipps Model.

which both models yields lower occupancy when congested. The data as well as the calibrated model show a pattern that has been observed in [52] with respect to the per-lane data. First, the speed distribution across different increasing from the right-most lane to the left-most lane. Second, the right-most yield considerably lower outflow (not shown) compared to the other two lanes.

The two graphs at the bottom of Figure 4 show in details the CD behavior. The T-curve (bottom right) is particular useful to distinct flow differences around capacity. The baseline flow is 5.4K veh/h therefore the T-curve increases for flows higher than that level and decreases otherwise. Clearly, both

the observed and the two models presents an increasing trend until 6:30AM and a decreasing trend after 7AM, meaning that the flow was higher than the baseline flow for a short period and smaller thereafter. In the bottom right graph, we highlight the highest flow in a period of 15 minutes (5748 veh/h) and the average flow on the congested period (5293veh/h) - resulting in a drop of around 8%. We highlight the period of highest flow in the bottom left graph. Both models reproduce the same trend with a slight advantage of IDM model that yielded highest peak flow and remained closer in the cumulative curve throughout the congested period. From the results depicted in Figure 4 that both Gipps and IDM model can reproduce CD fairly well.

Even though both models have a considerable number of parameters, both models yield similar performance across multiple days. Figure 5 depicts the time-series of occupancy (left), downstream flow (middle) and T-curve (right) for four different days, the first two days were used for calibration and the next two days were used for validation. The differences between the simulated and observed value are similar across the two models. In both models, the differences between simulated and observed values are similar. Both models were able to reproduce the peak outflow most of the time, although the peak lingered for longer in the observed field observed data.

Observing the details of the output on a specific day provides further insights on the triggering and dissipating elements of congestion. Figure 6 depicts the ramp demand in the top graph, the number of mandatory, cooperative, and discretionary lane changes that ocurred between the on-ramp and the lane drop in the middle graph, and the density contour in the bottom graph for March, 9th, 2017. The Contour plot (bottom graph) shows that the queue tail was slightly more than 5 kilometers away from its head at its maximum length.Observing the occupancy of loop detectors at California PeMS data on that specific day, it appeared that the congestion started at post mile 365.57 (401578) and reached post mile 361.81 (409859), causing 6km of congestion. Considering that congestion size is not a calibration metric, its closeness (5km vs. 6km) validates these results.

The evolution of number of lane changes over time highlights the occurrence of mandatory and cooperative lane changes as function of the bottleneck density and ramp demand. Vehicles from on-ramp attempts a discretionary lane-changing as soon as they join the mainline. Whenever the on-ramp demand is low and the bottleneck is uncongested, there is almost no mandatory lane changes as we observe very few mandatory lane changes compared to ramp density between 5AM and 6AM (blue line in the middle graph in Figure 6. Once the ramp demand exceeds 1000 veh/h, the merging vehicle are not able to find safe gaps and this leads to a larger number of mandatory and cooperative lane changes. There is an increase in the number of discretionary lane-changes; however, that occurs by vehicles taking advantage of gaps left by other vehicles that engaged in cooperative or mandatory lane changes.

Another qualitative finding from the simulation is that ramp demand has a clear impact on upstream density. From 7AM to 8AM the congestion the ramp demand exceeds 1000 veh/h (top graph at Figure 6) and the density is higher (bottom graph). Once the ramp demand decreases to 900 veh/h, the density in the merge area decreases as well.

Figure 7 illustrates the density-flow relationship at upstream (blue), merge (orange) and, downstream (green) areas. The orange line shows that congestion does not occur downstream of the bottleneck; however, upstream of the bottleneck (and merge), the flow density relationship goes to the congested branch of the fundamental diagram. In the merge area, observe the flow at around 20 vehicles per kilometer per lane to be close to 1900 vehicles per hour per lane on the congestion onset; however, when demand decreased and the congestion dissipated, the observed flow for the same density was below 1800 vehicles per hour per lane. This pattern of having for a given value of density (say, 20 veh/km/lane) yielding different flows for increasing or decreasing congestion characterizes a hysteresis and it is highlighted in the Figure. Hysteresis in traffic is commonly studied in the context of the macroscopic fundamental diagram (MFD) [53] (MFD) (e.g., see [54] and [55]), but it can also occur in highway traffic as highlighted in [56] which referred CD as a case of a clock-wise hysteresis (i.e., higher flow in increasing density). The hysteresis in the FD (or as they refer to, a critical density with two distinct speeds) was also shown in submicroscopic framework proposed in [23] (Figure 11).

B. WHAT ARE THE KEY FEATURES TO REPLICATE THE CAPACITY DROP PHENOMENON?

From the results in Figures 4 and III-A, it is clear that the CD can be properly replicated by microscopic models. We now follow the calibration approach to try to identify the key modeling features that leads to this conclusion.

Since both car-following models, IDM and Gipps, could reproduce the CD, further, we focus on the lane-changing behavior. As detailed in Section II-B, we consider three types of lane-changing: discretionary, mandatory and cooperative. The first two types are present in most, if not all, microsimulation model since the earlier efforts on lane-changing (e.g., see [29]). Cooperative lane-changing is not as established as studied compared to discretionary lane-changing although significant efforts (e.g., [27], [57], and [58]). Nevertheless, it is featured as a standard feature in commercial packages such as Aimsun [59]. Therefore, we conduct a set of experiments with this feature disabled.

Another feature we consider here is the relaxation behavior after the change of lane. This behavior is well established and had been postulated as one of potential features on the activation of merge bottlenecks [18]. Nevertheless, further studies on this aspect are scarce.

In order to understand the relative importance of drivers relaxation and cooperative lane-changing, we perform two additional calibration procedure with the Intelligent Driver Model:

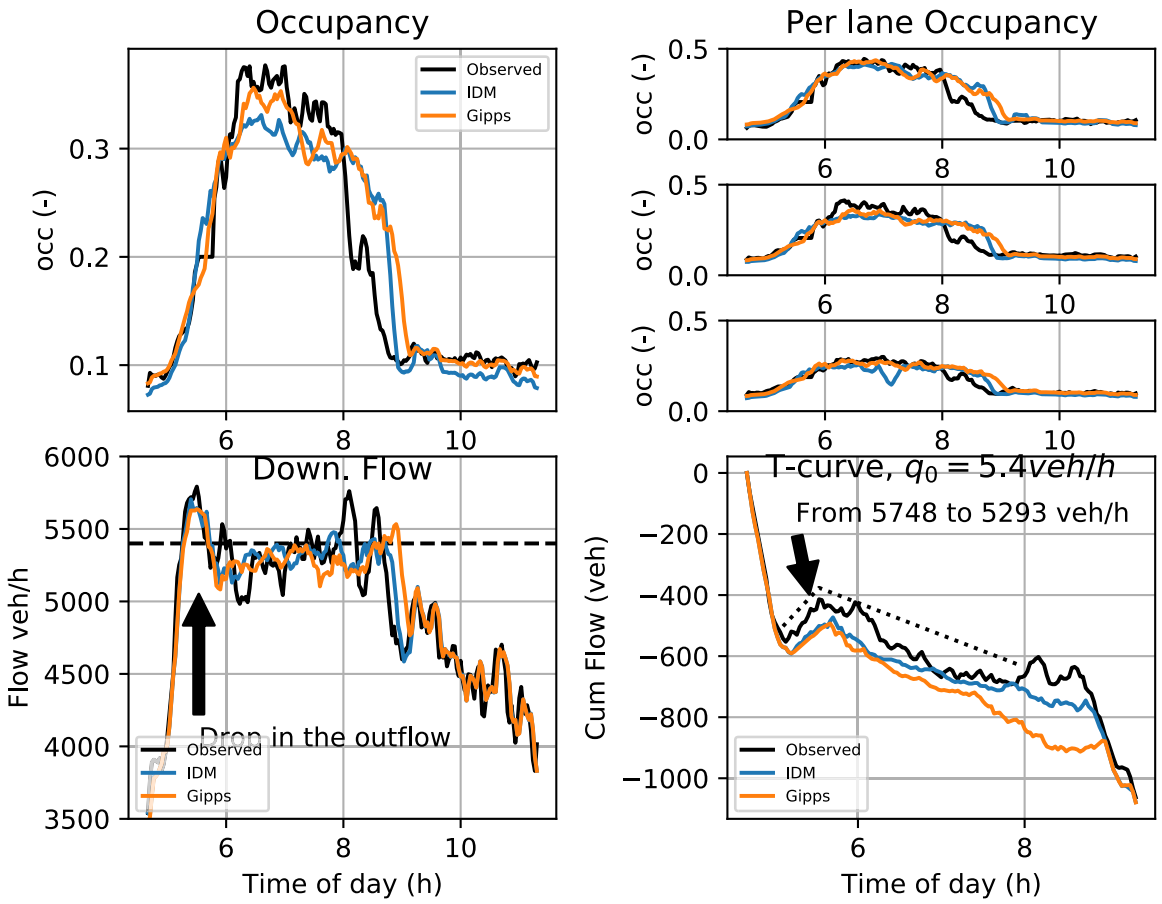


FIGURE 4. Time series of aggregated occupancy (top left), per-lane occupancy with the top graph showing the right most lane (top right), downstream flow (bottom left) and shifted Downstream Cumulative curve (T-curve - bottom right). Black, blue and orange lines represent observed data, IDM model, and Gipps models respectively.

- No cooperation: with cooperative lane-changing feature disabled and consequently not considering the parameters associated with this type of lane-changing;
- No Relaxation: disregard the relaxation following lane-changing ($\gamma = 0$).

Note that entails additional calibration runs as oppose to using the previous calibration result and allowing changes in all other parameters to better reproduce the observed data.

Figure 8 depicts the Pareto Sets for the three cases. The most salient result is the significant difference of the No Cooperation case compared to the other two cases. Specifically, the smallest occupancy RMSE is around 30% (5.3% compared to 4%) higher compared to the other two cases. Observe also that the cumulative flow RMSE is already higher for lower values of occupancy error. In all, this suggests that cooperative lane-changing is a key element around active bottlenecks. The reason for that probably is similar to the necessity of mandatory lane-changing model with explicit interactions: in congested areas there are few or no opportunities for discretionary lane-changing [27], [57]. Nevertheless, lane-changing maneuvers do occur on congested periods and this requires a combined action between follower on the target lane and the subject vehicle. This pattern can be observed on middle graph of Figure 6 which shows the number of mandatory and cooperative lane changes increasing significantly on the congested period.

On the other hand, removing the drivers relaxation following the lane-changing has little impact in the overall calibration results. It is important to point out that this result does not mean the relaxation does not occur in the real world. Rather, it means that this aspect do not influence much the observed measurements by the loop detectors upstream and downstream of the bottleneck. Particularly to this study, we did not calibrate the lane-changing model based on observed vehicle trajectories. From another point of view, this also means that the relaxation time could be calibrated separately and it would not influence the overall calibration result.

The qualitative impact of disabling these features can be observed on Figure 9. In blue it is depicted the results with the full model, in orange with no cooperative lane-changing and in green with no drivers relaxation. Clearly, the case without cooperative lane-changing cannot reproduce well the upstream occupancy significantly underestimating the occupancy in the middle and the left lane. Without driving relaxation the differences are more nuanced. Qualitatively the upstream occupancy are similar. On the other hand, the outflow in the congested is slightly underestimated compared

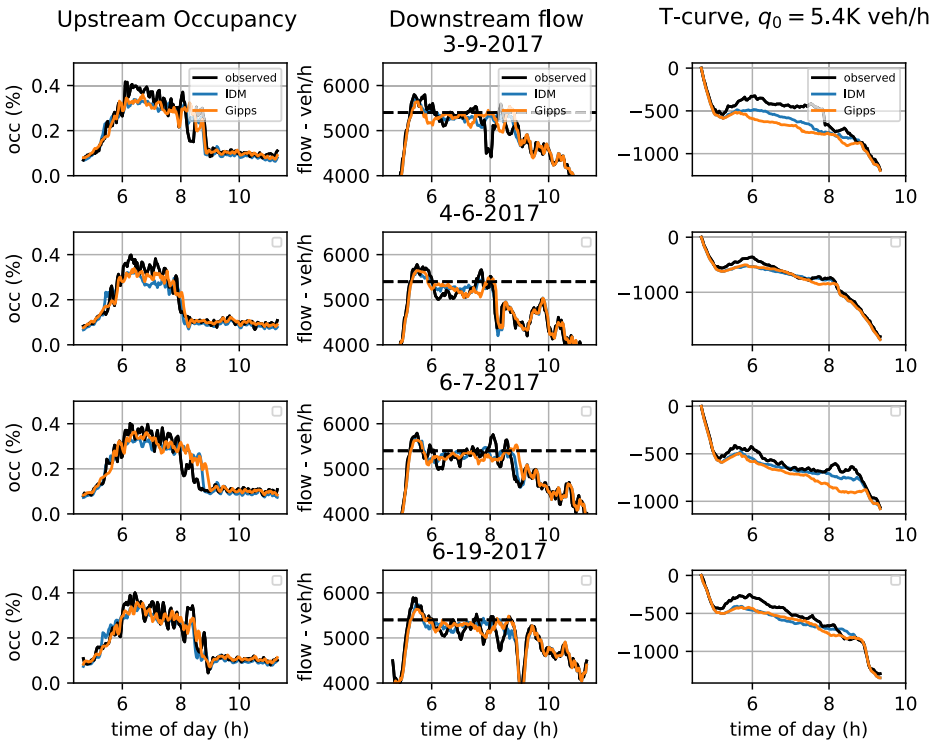


FIGURE 5. Comparison between IDM (blue) and Gipps (orange) models against observed data(black) over multiple days. Each row represents a different day. The left column depicts the upstream occupancy, the middle column downstream flow and the right column the shifted downstream cumulative curve (T-curve) for $q_0 = 5.4$ veh/h.

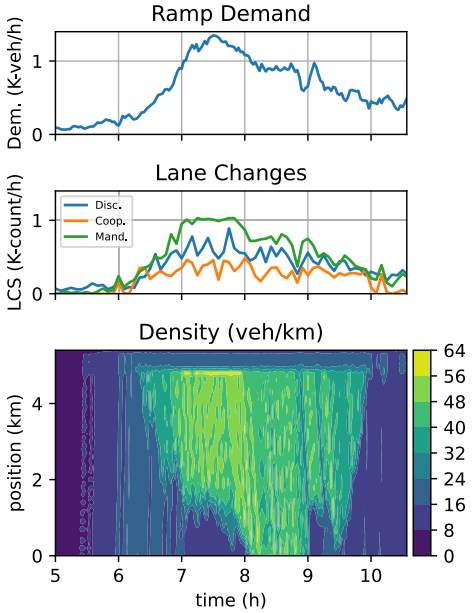


FIGURE 6. Ramp demands (top), number of lane changes at the merge for each lane changing type (middle) and density countour (bottom).

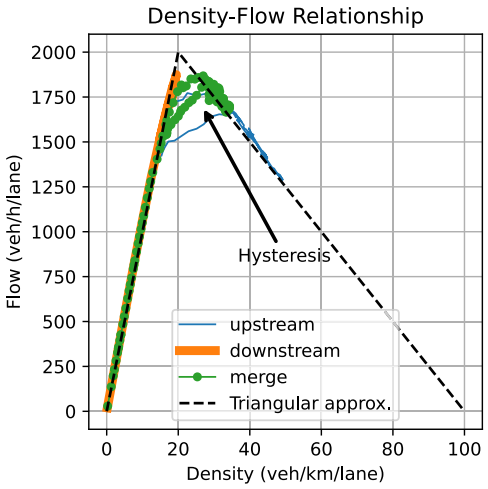


FIGURE 7. Density-flow scatter-plot on the upstream, merge and downstream areas. The blacked dashed lines represents an approximate triangular fundamental diagram for the merge area.

to the full model and the observed data as can be observed in the T-curve (bottom right) on Figure 9. Note, however, that the T-curve amplifies small differences and one can see that this difference is hard to distinct when observing the downstream flow (bottom left).

In summary, the calibration results confirm that CD can be reproduced with microscopic models and that the cooperative lane-changing is a key behavior on active bottlenecks.

IV. SENSITIVITY ANALYSIS

In this section, we perform a sensitivity analysis with respect to demand and acceleration behavior in order to assess their impacts on CD at merge bottlenecks.

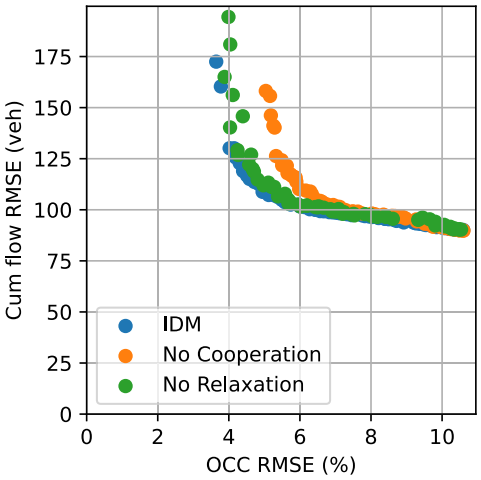


FIGURE 8. Pareto set comparison between the IDM model with all features (blue) against IDM model with no cooperative LC (orange) and no relaxation after changes of lane (green).

In Section III, we obtained multiple combinations of parameters with similar accuracy (see Figure 3), the forthcoming analysis utilizes the 15 parameter combinations with lower occupancy error to perform a sensitivity analysis. We use multiple combinations so as to avoid any bias that could be associated to an individual parameter combination. Instead, we assess whether the trends associated with each aspect analyzed holds in all combinations suggesting that it most likely hold in general.

We start with the sensitivity analysis with respect to ramp demand. We run experiments with the upstream demand increasing from 3000 veh/h until 6500 veh/h with linear increase (in steps of 2 minutes) over the course of 8 hours of simulation. Therefore, the demand will increase slowly and we can then analyze the prevailing demand on the onset of congestion and the impact in the downstream outflow when the upstream section becomes congested. For each of the 15 parameter combination, we run 5 simulations with on-ramp demand increasing from 0 to 1200 veh/h in steps of 300 veh/h.

Figure 10 depicts the results for the different ramp demands. The top graph depicts the upstream occupancy whereas the graph at the bottom depicts downstream flow. For each time step, we compute the median, lower and upper bound across the different parameters. The median trajectories are represented with continuous lines whereas the upper and lower values are represented with transparent dashed lines with the same color for each ramp demand case. The case with no ramp demand (0 veh/h) the upstream section never gets congested since the area no longer is a bottleneck since no vehicle uses the additional lane that is eventually dropped. The resulting outflow in this case is slightly around 5900 veh/h. In all other cases, the upstream sections gets eventually congested depending on the on-ramp flow. The results show that the downstream outflow decreases as the on-ramp flow reaching 5100 veh/h for on-ramp demand of 1200 veh/h.

The CD is often associated with bounded acceleration [60], [61]. To investigate its impact, we perform simulations for on-ramp demands at 600 and 900veh/h, but in those cases we vary the acceleration parameter of each combination (A on the IDM model) from -20% to $+20\%$ compared to its calibrated value. Figure 11 depicts the results. Each color represents one demand pattern and the case with higher acceleration and lower bound are marked with circles and triangles, respectively. For both ramp demands, a higher maximum acceleration rate parameter (A on IDM model) would lead to higher bottleneck outflow.

Finally, Figure 12 depicts the change in bottleneck flow (throughput) summarizes the last results. The top graph depicts the median throughput along with its upper and lower bound for varying ramp demands. For the median case, it shows that the outflow is especially sensitive for ramp demands between 300 and 900 veh/h. The bottom graph shows the result ramping flow for different changes in the acceleration rate. Observe that for no ramp demand the result is the same regardless of the acceleration rate. For higher ramp demand rates, the bottleneck throughput is higher for higher acceleration rates.

The results of sensitivity analysis are not unexpected, but two elements are worth discussing. First, the CD often associates with the upstream congestion - that is, the flow or ''capacity'' drops once queue forms upstream of the bottleneck (e.g, see [2]). The results depicted in Figure 10 shows a decrease in the outflow that is salient in the case of ramp demand of 300 and 600 veh/h (green and red lines). However, it is difficult to assess the outflow difference before and after congestion because the changes are rather small. For example, the case of 300 veh/h (green line in Figure 10) peaks at around 1870 veh/h/lane (5610 veh/h) and the observed minimum flow is around 1820 veh/h/ (5460 veh/h) lane recovering to 1850 veh/h/lane (5550 veh/h) thereafter. This means that the drop in the flow is at most 2.7% due to the upstream congestion in this case. Similar drops are observable in other cases. Therefore, these results suggest that changes in the car-following model, specifically the acceleration parameter, lead to a limited impact on the throughput before and after congestion. Nevertheless, changes in the acceleration parameter leads to significant change in the throughput during congested and uncongested periods, as can be observed in Figure 12 showing a decreasing throughput for lower acceleration rate parameter.

Second, the on-ramp demand seems to play an important role in the occurrence of the CD. In this case, the flow drops from 2000 veh/h/lane to 1750 veh/h/lane when the ramp flow increases from 0 to 1200 veh/h resulting in a relative drop in the flow of 12.5%. Although there has been models relating CD to the ramp flow, especially [13], most of the first-order models associate CD with congestion. For example, in [6] different CD models were analyzed and only 1 out of 6 relates CD to the ramp CD model based on ramp flows yield better fit

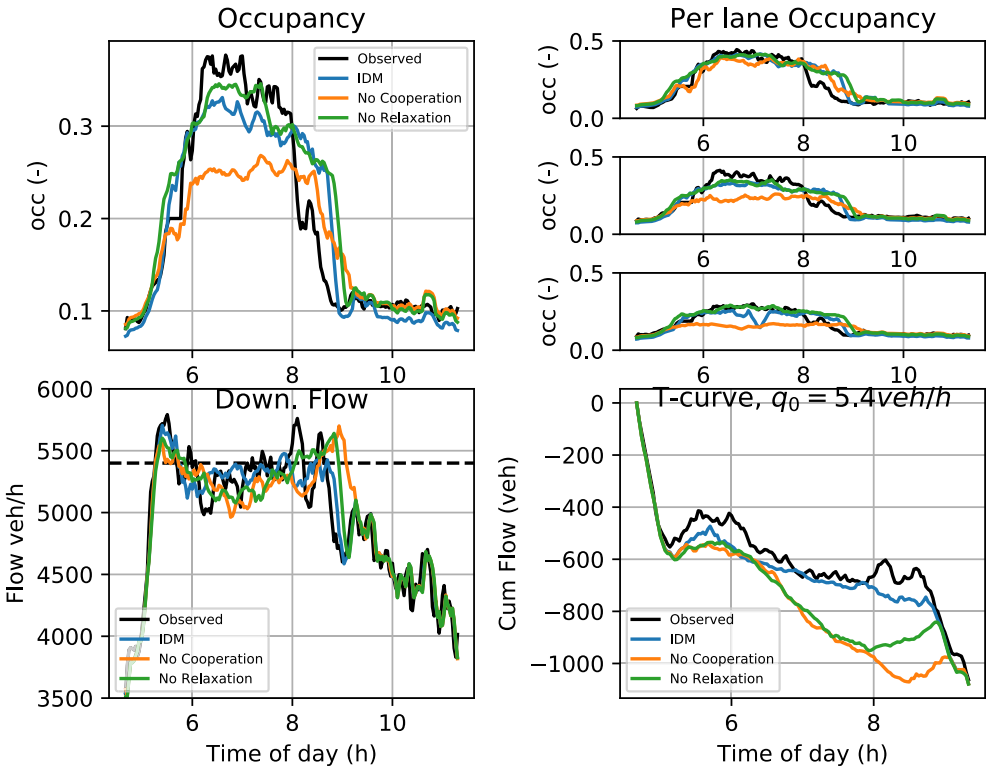


FIGURE 9. Time series comparison between the IDM model with all features (blue) against IDM model with no cooperative LC (orange) and no relaxation after changes of lane (green). The top left graph depicts occupancy, the top right the per-lane-occupancy for each lane (right most lane on the top), the bottom left graph the downstream flow and the bottom right graph the shifted downstream cumulative curve.

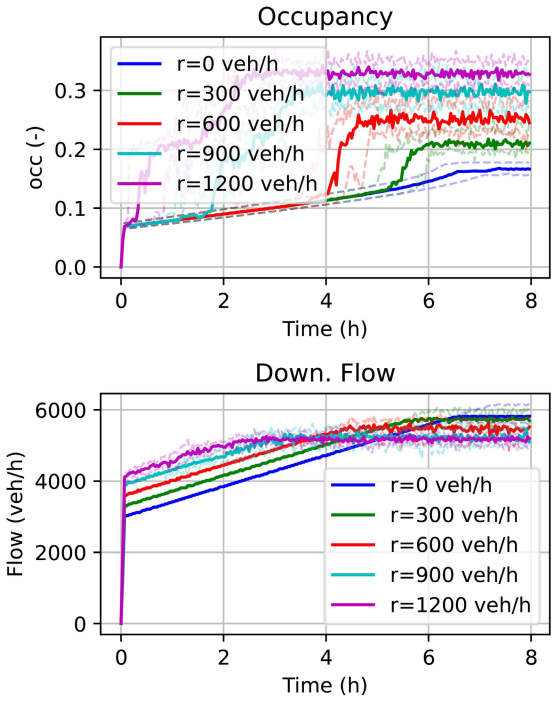


FIGURE 10. Occupancy (top) and downstream flow (bottom) for increasing upstream demand and varying (constant) ramp demand.

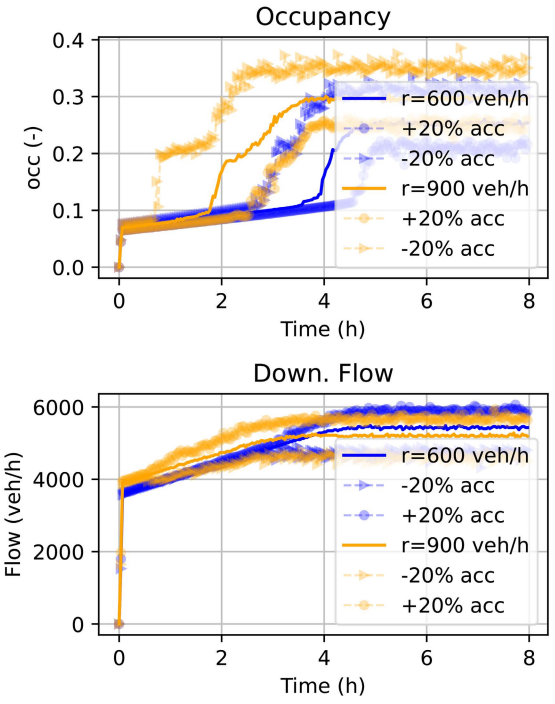


FIGURE 11. Occupancy (top) and downstream flow (bottom) for increasing upstream demand and varying (constant) ramp demand for acceleration rates varying from −20% to +20% from the calibrated values.

results compared to the models which only consider upstream congestion [62]. In all, the results shown here and in the models by [13] and [14] suggest similar mechanisms for CD which states that the capacity drop at merges is an outcome of

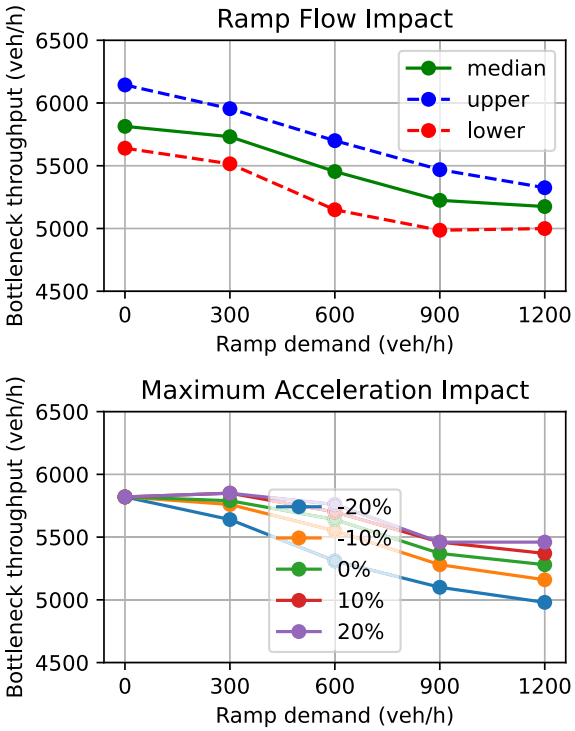


FIGURE 12. Bottleneck throughput for varying ramp demand for no changes in the acceleration parameter (top) and changes in the maximum acceleration (bottom).

lane changing behavior and bounded acceleration. Therefore, from the perspective of the simulation model, the CD arises as an interaction between the lane-changing and car-following models.

V. CONCLUSION

A multiobjective calibration approach using microscopic models was presented with the goal of unveiling the key dynamic features at merge bottlenecks associated with the CD phenomenon. The same steps applied in this paper - model, calibration and sensitivity analysis - can be used in further studies to better understand freeway traffic. These steps also establish a recipe for traffic engineers to obtain more accurate models for traffic forecasting. Based on the results, we established the following conclusions:

- Microscopic models can replicate CD. Although previous work had touched upon CD in the light of microscopic models, no previous study had shown how to replicate CD at merge bottlenecks according to real data.
- Both car-following models applied, Gipps and IDM, could replicate CD, with a slight advantage of the former. The fact that both models applied in this study well replicated CD indicates that other models can likely do so too.
- Further calibration cases, specifically removing cooperative lane-changing and relaxation after lane-changing, suggests that cooperative lane-changing is an essential feature at merge bottlenecks since calibration errors are significantly higher when that feature is not considered in the model. Although yielding different results, removing the relaxation behavior from the model does not change the performance metrics significantly.
- The sensitivity analysis shows that the key variable causing a drop in the flow is the ramp demand compared to the maximum acceleration rate. The results present a decreasing bottleneck throughput for an increasing ramp demand up to a level where the throughput is roughly constant as demand increases. This relationship goes in line with [63] that showed lower throughput when ramp demand increases, yielding a smaller merge bottleneck throughput.

The work presented here has its own limitations. First, in all simulations, we assumed homogeneous vehicles and drivers. In reality, drivers are heterogeneous and a significant portion of the vehicles are heavy duty. A potential association of heterogeneous vehicles and CD was discussed in [16]. Second, we used data from loop detectors located upstream and downstream of the bottleneck as the input data of the calibration. Loop detectors are widely used in freeways and can provide comprehensive data over multiple days. However, we did not use vehicle trajectory data which can particularly be useful for capturing the lane-changing behavior more accurately. Therefore, from the parsimony principle, we employed the fewer number of assumptions on the analysis. This reduces the number of variables and scenarios to analyze the results and also turns the simulations deterministic which facilitates the simulation-based calibration procedure. Moreover, this study shows that driver heterogeneity is not a necessary factor in replicating CD, as the CD was replicated in its absence. We could not reach this conclusion if drivers were heterogeneous in this study.

We also note that this study is limited to merge bottlenecks with high on-ramp demand triggering congestion in the mainline highway. Although we affirmed that for this particular case driver heterogeneity is not a necessary condition to replicate the capacity drop, it does not mean that driver heterogeneity does not play a role on the CD. In fact, we pointed out that the results presented in this study suggests a mechanism proposed by [13] which explains the CD as the interaction between the lane changing behavior and bounded acceleration without the presence of driver heterogeneity. Their model was later extended to account heterogeneous vehicles [15] which they find that heterogeneous vehicles have little influence on the magnitude of CD although having strong influence on the traffic dynamics at merges. Other traffic features that leads to sustained decrease throughput on freeways may also apply on the particular merge bottleneck case for example some drivers slowing down on the freeway close to an incident which have shown to lead to lower throughput [64].

As part of our future work, we intend to address some of the aforementioned limitations with more detailed datasets. Specifically, we would like to relax the limitations of driver

homogeneity in the simulation and analysis. Moreover, we are interested in applying the proposed framework in applications in which merge bottleneck is critical such as ramp metering and variable speed limit. As for ramp metering, this model can be used to investigate factors such as sensor placement and investigate cases where the on-ramp and lane drop are further away from each other, which creates control design challenges. (e.g., [65] and [66]). Similar aspects, such as distance between the actuation point and bottleneck, can also be investigated in variable speed limits. The proposed framework can also be used to investigate existing (e.g., [67] and [68]) or new strategies combining ramp metering and variable speed limits.

In addition, we are interested in analyzing the impacts of emerging technologies, such as Adaptive Cruise Control (ACC), building upon the recent work [48].

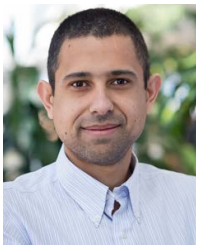
ACKNOWLEDGMENT

The authors would like to thank Irene Martínez Josemaría that provided valuable feedback on earlier versions of this manuscript.

REFERENCES

[1] R. Dowling, A. Skabardonis, J. Halkias, G. McHale, and G. Zammit, "Guidelines for calibration of microsimulation models: Framework and applications," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1876, no. 1, pp. 1–9, Jan. 2004.
[2] M. J. Cassidy and R. L. Bertini, "Some traffic features at freeway bottlenecks," *Transp. Res. B, Methodol.*, vol. 33, no. 1, pp. 25–42, Feb. 1999.
[3] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 229, no. 1178, pp. 317–345, 1955.
[4] P. I. Richards, "Shock waves on the highway," *Oper. Res.*, vol. 4, no. 1, pp. 42–51, 1956.
[5] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. Part B, Methodol.*, vol. 28, no. 4, pp. 269–287, Aug. 1994.
[6] M. Kontorinaki, A. Spiliopoulou, C. Roncoli, and M. Papageorgiou, "First-order traffic flow models incorporating capacity drop: Overview and real-data validation," *Transp. Res. B, Methodol.*, vol. 106, pp. 52–75, Dec. 2017.
[7] H. J. Payne, "Model of freeway traffic and control," in *Mathematical Models of Public Systems*. 1971, pp. 51–61. [Online]. Available: https://books.google.com.br/books?redir_esc=y&id=7I5FAAAAYAAJ&focus=searchwithinvolume&q=payne
[8] G. Whitham and R. G. Fowler, "Linear and nonlinear waves," *Phys. Today*, vol. 28, p. 55, Jan. 1975.
[9] A. Messmer and M. Papageorgiou, "METANET: A macroscopic simulation program for motorway networks," *Traffic Eng. Control*, vol. 31, no. 9, pp. 466–470, 1990.
[10] M. Papageorgiou, "Some remarks on macroscopic traffic flow modelling," *Transp. Res. A, Policy Pract.*, vol. 32, no. 5, pp. 323–329, 1998.
[11] S. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2324, pp. 63–70, Dec. 2012.
[12] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transp. Res. B, Methodol.*, vol. 40, no. 3, pp. 251–264, Mar. 2006.
[13] L. Leclercq, J. A. Laval, and N. Chiabaut, "Capacity drops at merges: An endogenous model," *Proc.-Social Behav. Sci.*, vol. 17, pp. 12–26, Jan. 2011.
[14] D. Chen and S. Ahn, "Capacity-drop at extended bottlenecks: Merge, diverge, and weave," *Transp. Res. B, Methodol.*, vol. 108, pp. 1–20, Feb. 2018.
[15] L. Leclercq, V. L. Knoop, F. Marczak, and S. P. Hoogendoorn, "Capacity drops at merges: New analytical investigations," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 171–181, Jan. 2016.
[16] D. Ngoduy, "Multiclass first-order traffic model using stochastic fundamental diagrams," *Transportmetrica*, vol. 7, no. 2, pp. 111–125, Mar. 2011.
[17] M. Treiber, A. Kesting, and D. Helbing, "Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 1, Jul. 2006, Art. no. 016123.
[18] S. Kim and B. Coifman, "Driver relaxation impacts on bottleneck activation, capacity, and the fundamental relationship," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 564–580, Nov. 2013.
[19] E. Chamberlayne, H. Rakha, and D. Bish, "Modeling the capacity drop phenomenon at freeway bottlenecks using the INTEGRATION software," *Transp. Lett.*, vol. 4, no. 4, pp. 227–242, Oct. 2012.
[20] E. R. Müller, R. C. Carlson, W. Kraus, and M. Papageorgiou, "Microsimulation analysis of practical aspects of traffic control with variable speed limits," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 512–523, Feb. 2015.
[21] J. Barceló and J. Casas, "Dynamic network simulation with AIMSUN," in *Simulation Approaches in Transportation Analysis* (Operations Research/Computer Science Interfaces Series), vol. 31, R. Kitamura and M. Kuwahara, Eds. Boston, MA, USA: Springer, 2005, doi: 10.1007/0-387-24109-4_3.
[22] K. Yuan, V. L. Knoop, and S. P. Hoogendoorn, "A microscopic investigation into the capacity drop: Impacts of longitudinal behavior on the queue discharge rate," *Transp. Sci.*, vol. 51, no. 3, pp. 852–862, Aug. 2017.
[23] F. A. Mullakkal-Babu, M. Wang, B. van Arem, B. Shyrokau, and R. Happee, "A hybrid submicroscopic-microscopic traffic flow simulation framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3430–3443, Jun. 2021.
[24] F. de Souza, M. Mosslemi, J. A. Vrugt, and W. Jin, "Microscopic simulation replicates the capacity drop phenomenon," *Proc. Comput. Sci.*, vol. 130, pp. 908–913, Jan. 2018.
[25] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, p. 1805, Aug. 2000.
[26] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transp. Res. B, Methodol.*, vol. 15, no. 2, pp. 105–111, 1981.
[27] P. Hidas, "Modelling vehicle interactions in microscopic simulation of merging and weaving," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 1, pp. 37–62, 2005.
[28] Z. Zheng, "Recent developments and research needs in modeling lane changing," *Transp. Res. B, Methodol.*, vol. 60, pp. 16–32, Feb. 2014.
[29] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transp. Res. B, Methodol.*, vol. 20, no. 5, pp. 403–414, Oct. 1986.
[30] A. Kesting and M. Treiber, "How reaction time, update time, and adaptation time influence the stability of traffic flow," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 23, no. 2, pp. 125–137, 2008.
[31] M. Treiber and A. Kesting, "Car-following models based on driving strategies," in *Traffic Flow Dynamics*. Berlin, Germany: Springer, 2013, doi: 10.1007/978-3-642-32460-4_11.
[32] P. Hidas, "Modelling lane changing and merging in microscopic traffic simulation," *Transp. Res. C, Emerg. Technol.*, vol. 10, nos. 5–6, pp. 351–371, Oct./Dec. 2002.
[33] L. Leclercq, N. Chiabaut, J. Laval, and C. Buisson, "Relaxation phenomenon after lane changing: Experimental validation with NGSIM data set," *Transp. Res. Rec.*, vol. 1999, no. 1, pp. 79–85, 2007.
[34] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model MOBIL for car-following models," *Transp. Res. Rec.*, vol. 1999, no. 1, pp. 86–94, Jan. 2007.
[35] C. F. Choudhury, M. E. Ben-Akiva, T. Toledo, G. Lee, and A. Rao, "Modeling cooperative lane changing and forced merging behavior," in *Proc. 86th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2007, pp. 1–23.
[36] M. Rahman, M. Chowdhury, Y. Xie, and Y. He, "Review of microscopic lane-changing models and future research opportunities," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1942–1956, Dec. 2013.
[37] T. Choe, A. Skabardonis, and P. Varaiya, "Freeway performance measurement system: Operational analysis tool," *Transp. Res. Record: J. Transp. Res. Board*, vol. 1811, no. 1, pp. 67–75, Jan. 2002.
[38] F. D. Souza, "Freeway loop detector data reconciliation based on vehicle conservation," *Proc. Comput. Sci.*, vol. 151, pp. 321–326, Jan. 2019.
[39] Q. Meng and H. L. Khoo, "Self-similar characteristics of vehicle arrival pattern on highways," *J. Transp. Eng.*, vol. 135, no. 11, pp. 864–872, Nov. 2009.
[40] V. Punzo and F. Simonelli, "Analysis and comparison of microscopic traffic flow models with real traffic microscopic data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1934, pp. 53–63, Jan. 2005.

[41] 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," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 184–193, Feb. 2015.
[42] R. Balakrishna, C. Antoniou, M. Ben-Akiva, H. N. Koutsopoulos, and Y. Wen, "Calibration of microscopic traffic simulation models: Methods and application," *Transp. Res. Rec.*, vol. 1999, no. 1, pp. 198–207, 2007.
[43] M. J. Cassidy and J. R. Windover, *Methodology for Assessing Dynamics of Freeway Traffic Flow*, no. 1484. Transportation Research Record, 1995. [Online]. Available: https://onlinepubs.trb.org/Onlinepubs/trr/1995/1484/1484-010.pdf
[44] S. Oh and H. Yeo, "Estimation of capacity drop in highway merging sections," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2286, no. 1, pp. 111–121, Jan. 2012.
[45] M. J. Cassidy, "Bivariate relations in nearly stationary highway traffic," *Transp. Res. B, Methodol.*, vol. 32, pp. 49–59, Jan. 1998.
[46] V. Punzo and M. Montanino, "Speed or spacing? Cumulative variables, and convolution of model errors and time in traffic flow models validation and calibration," *Transp. Res. B, Methodol.*, vol. 91, pp. 21–33, Sep. 2016.
[47] A. Kesting and M. Treiber, "Calibrating car-following models by using trajectory data: Methodological study," *Transp. Res. Rec.*, vol. 2088, pp. 148–156, Mar. 2008.
[48] F. de Souza and R. Stern, "Calibrating microscopic car-following models for adaptive cruise control vehicles: Multiobjective approach," *J. Transp. Eng., A, Syst.*, vol. 147, no. 1, Jan. 2021, Art. no. 04020150.
[49] Y. Hollander and R. Liu, "The principles of calibrating traffic microsimulation models," *Transportation*, vol. 35, no. 3, pp. 347–362, May 2008.
[50] T. Robič and B. Filipič, "DEMO: Differential evolution for multiobjective optimization," in *Evolutionary Multi-Criterion Optimization. EMO 2005* (Lecture Notes in Computer Science), vol. 3410, C. A. C. Coello, A. A. Hernández, and E. Zitzler, Eds. Berlin, Germany: Springer, 2005, doi: 10.1007/978-3-540-31880-4_36.
[51] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
[52] K. Yuan, V. L. Knoop, L. Leclercq, and S. P. Hoogendoorn, "Capacity drop: A comparison between stop-and-go wave and standing queue at lane-drop bottleneck," *Transportmetrica B, Transp. Dyn.*, vol. 5, no. 2, pp. 145–158, Apr. 2017.
[53] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transp. Res. B, Methodol.*, vol. 42, no. 9, pp. 759–770, Nov. 2008.
[54] N. Geroliminis and J. Sun, "Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks," *Proc.-Social Behav. Sci.*, vol. 17, pp. 213–228, Jan. 2011.
[55] V. V. Gayah and C. F. Daganzo, "Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability," *Transp. Res. B, Methodol.*, vol. 45, no. 4, pp. 643–655, 2011.
[56] C. F. Daganzo, "On the macroscopic stability of freeway traffic," *Transp. Res. B, Methodol.*, vol. 45, no. 5, pp. 782–788, Jun. 2011.
[57] K. I. Ahmed, "Modeling drivers' acceleration and lane changing behavior," Sc.D. thesis, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 1999. [Online]. Available: https://dspace.mit.edu/handle/1721.1/9662?show=full
[58] M. Sarvi and M. Kuwahara, "Microsimulation of freeway ramp merging processes under congested traffic conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 470–479, Sep. 2007.
[59] J. Casas, J. L. Ferrer, D. Garcia, J. Perarnau, and A. Torday, "Traffic simulation with Aimsun," in *Fundamentals of Traffic Simulation* (International Series in Operations Research & Management Science), vol. 145, J. Barceló, Ed. New York, NY, USA: Springer, 2010, doi: 10.1007/978-1-4419-6142-6_5.
[60] J. P. Lebacque, "Two-phase bounded-acceleration traffic flow model: Analytical solutions and applications," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1852, no. 1, pp. 220–230, Jan. 2003.
[61] M. M. Khoshyaran and J. P. Lebacque, "Capacity drop and traffic hysteresis as a consequence of bounded acceleration," *IFAC-PapersOnLine*, vol. 48, no. 1, pp. 766–771, 2015.
[62] F. de Souza, "Calibration procedure for traffic flow models of merge bottlenecks," in *Proc. 6th Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2019, pp. 1–7.
[63] L. Xiao, M. Wang, W. Schakel, and B. van Arem, "Unravelling effects of cooperative adaptive cruise control deactivation on traffic flow characteristics at merging bottlenecks," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 380–397, Nov. 2018.
[64] D. Chen, S. Ahn, J. Laval, and Z. Zheng, "On the periodicity of traffic oscillations and capacity drop: The role of driver characteristics," *Transp. Res. B, Methodol.*, vol. 59, pp. 117–136, Jan. 2014.
[65] Y. Wang, E. B. Kosmatopoulos, M. Papageorgiou, and I. Papamichail, "Local ramp metering in the presence of a distant downstream bottleneck: Theoretical analysis and simulation study," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2024–2039, Oct. 2014.
[66] F. de Souza and W. Jin, "Integrating a Smith predictor into ramp metering control of freeways," 2017. [Online]. Available: https://trid.trb.org/view/1439497
[67] R. C. Carlson, I. P. M. Papamichail, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transp. Sci.*, vol. 44, no. 22, pp. 238–253, 2010.
[68] H. W. Cho and J. A. Laval, "Combined ramp-metering and variable speed limit system for capacity drop control at merge bottlenecks," *J. Transp. Eng., A, Syst.*, vol. 146, no. 6, Jun. 2020, Art. no. 04020033.



FELIPE DE SOUZA received the Bel.Eng. degree in control and automation engineering and the M.E. degree in automation and systems engineering from the Federal University of Santa Catarina, Brazil, in 2008 and 2012, respectively, and the Ph.D. degree in transportation systems engineering from the University of California at Irvine, Irvine, in 2018.

Since 2018, he has been a Postdoctoral Appointee with the Argonne National Laboratory, Vehicle and Mobility Systems Group. His research interests include traffic modeling and control, transportation systems simulation, and dispatching algorithms for on-demand mobility services.

Dr. de Souza was a recipient of Science Without Borders Scholarship for postgraduate studies abroad from the CAPES Foundation, Ministry of Education, Brazil, from 2014 to 2018, and the Best Paper Award at IEEE CASE 2010 Conference. In addition, he was a recipient of the 2017 ITS California Scholarship for continued research in intelligent transportation systems.

MARJAN MOSSLEMI received the M.S. degree in water studies from the University of Bergen, Norway, and the M.S. degree in transportation engineering from Linköping University, Sweden. She is currently pursuing the Ph.D. degree in transportation systems engineering with the University of California at Irvine, Irvine.

From 2009 to 2015, she worked in the transportation engineering sector, both in research and industry, first in Norway and later in Canada, which led to several collaborative publications. She focuses on the supply and demand analysis of shared autonomous vehicle services. Her research interests include traffic modeling, transportation planning, and operation research for shared autonomous vehicle services.

Ms. Mosslemi was awarded the NORAM Scholarship issued by Norway-America Association, in 2019, which offers limited scholarships annually to outstanding Norwegian graduate students in the USA.