Determining Desired Speeds from Vehicle Trajectory Data

Marvin V. Baumann[1] , **Claude M. Weyland**[1] , **Jan Ellmers**[1] ,
Lea Fuchs[1] , **Josephine Grau**[1] , and **Peter Vortisch**[1]

Abstract

Desired speed or free speed distributions are important input parameters for microscopic traffic flow simulations. Whereas driven speeds can be measured, desired speeds are not detectable for all vehicles because of vehicles constraining each other. An established approach to estimating desired speed distributions is the modified Kaplan–Meier approach, which estimates desired speed distributions based on single vehicle data from stationary detectors. We propose a novel approach to determine desired speeds based on vehicle trajectory data. The proposed approach, as well as the modified Kaplan–Meier approach, is applied to a trajectory dataset recorded on a German freeway. With the modified Kaplan–Meier approach, we observed that the resulting desired speed distributions vary by approximately 5 km/h depending on the position of the stationary detectors. The desired speed distributions obtained from the trajectory-based approach are approximately 5 to 10 km/h higher than those estimated by the modified Kaplan–Meier approach. This difference in results is probably because the trajectory-based approach observes each vehicle over a longer distance rather than just at stationary points. Nevertheless, it can be concluded that the estimation of desired speed distributions is subject to a certain degree of inaccuracy. The analysis of the vehicle trajectory data revealed a notable intra-vehicle instability in desired speeds, with a difference of 5 to 7 km/h observed for 40% of the vehicles between different periods of free driving. These findings should be considered in the context of calibrating microscopic traffic flow simulations.

Keywords

operations, calibration/validation, microscopic traffic simulation, vehicle trajectory, free speed, desired speed, modified Kaplan-Meier approach

The term "desired speed" (or "free speed") refers to the maximum speed a driver intends to reach. If necessary, the driver performs a lane change to overtake preceding vehicles traveling at lower speeds. The desired speed is individual to each driver, resulting in a distribution of desired speeds across the collective of drivers on a particular road segment. Whereas driven speeds can be measured, desired speeds are not detectable for all vehicles. The relation between driven speed and desired speed can only be assumed for vehicles that are moving freely without being constrained by preceding vehicles. As traffic density increases, the interaction between road users intensifies, leading to fewer freely moving vehicles and consequently a greater disparity between desired and driven speeds. Furthermore, road users with higher desired speeds exhibit a higher likelihood of being constrained compared with those with lower desired speeds (*1*). Relying solely on free speeds of a subset of vehicles to infer the desired speed distribution for all road users would result in a systematic underrepresentation of high desired speeds. However, this approach is used in certain studies to determine desired speeds on highways (*2*, *3*). The most established method for estimating desired speed distributions is the modified Kaplan–Meier approach (*4*, *5*). This approach can be applied to single vehicle data recorded by stationary detectors, consisting of both constrained and unconstrained vehicles.

Desired speed distributions serve as an input parameter for microscopic traffic flow simulations and significantly influence the traffic dynamics within the simulation.

[1]Institute for Transport Studies, Karlsruhe Institute of Technology, Karlsruhe, Germany

Corresponding Author:
Marvin V. Baumann, marvin.baumann@kit.edu

Numerical simulations conducted by Lipshtat indicate that the width of the distribution has a significant impact on the stability of traffic flow because it leads to larger variations in desired speeds among individual drivers (*6*). These differences result in more drivers being constrained by those with lower desired speeds, inevitably leading to an increased amount of overtaking attempts. Additionally, Farzaneh and Rakha have found that the desired speed distribution also influences the speed at capacity and, therefore, the shape of the fundamental diagram (*7*). Thus, it is crucial to calibrate the desired speed distributions in microscopic traffic flow models.

The advancements in drone technology and automatic image processing now offer the possibility to record and analyze vehicle trajectories for traffic flow investigations. Utilizing vehicle trajectories to estimate desired speeds provides significant benefits because it enables the examination of each vehicle over an extended distance. Moreover, it provides flexibility in recording as vehicle trajectories can be captured at almost any desired location, while stationary detectors for recording single vehicle data need to be installed along the observed road segment.

Within the scope of this study, we propose a method to determine desired speed distributions based on vehicle trajectory data. The approach entails identifying intervals within each trajectory where the vehicle travels at free speed. Our evaluations are based on a vehicle trajectory dataset of a 1,200-m-long segment of a German freeway. To examine the validity of the proposed method, we apply the established modified Kaplan–Meier approach to the dataset by synthesizing stationary detector data from the vehicle trajectories. Further, the intra-vehicle stability of desired speed and the overall concept of one global desired speed per vehicle are discussed. To the best of our knowledge, there are currently no published approaches for determining desired speeds from trajectory data.

Literature Review

Desired Speed Estimation

Approaches for estimating desired speed distributions described in the literature are based on product-limit methods. One commonly used method is the estimator by Kaplan and Meier, originally applied in the context of clinical studies (*8*). This method features the estimation of survival probabilities over a specific time period, for example, after an operation. Some patients pass away at a specific point in time after the operation because of post-operative complications. Other patients experience different causes of death or drop out of the study. Therefore, all observed lifetimes contribute to the estimation as either a death event or a loss, forming an incomplete observation. The Kaplan–Meier approach is a nonparametric approach, meaning that the resulting

distribution does not conform to a specific mathematical function type.

In traffic engineering, desired speed distributions draw an analogy to survival probabilities in medicine. Instead of a time period, we consider speeds, and instead of the events of death and loss, we consider whether a vehicle is unconstrained (free driving) or constrained by preceding vehicles. In the context of desired speeds, the application of the Kaplan–Meier approach was initially introduced by Botma (*1*). Hoogendoorn (*4*, *9*) and Geistefeldt (*5*, *10*) have further developed this approach for estimating desired speeds. In addition, the Kaplan–Meier approach is used in traffic flow analysis to determine distribution functions of freeway capacity (*10–12*). Also analogous to survival probabilities in medicine, the procedure involves estimating the survival probability of the traffic flow, whereby the capacity represents the lifetime of the traffic flow.

To estimate desired speed distributions, individual vehicle data from stationary detector recordings are used. For each vehicle j passing through the stationary detector, the speed v_j and the time headway t_j are known. Vehicles are initially categorized as constrained or unconstrained based on their headway. For constrained vehicles, the desired speed cannot be easily determined, but it is greater than or equal to the driven speed. In contrast, for unconstrained vehicles, the driven speed can be equated with the desired speed.

A significant disadvantage of the Kaplan–Meier approach is that the resulting desired speed distribution strongly depends on the chosen threshold value of the headway. Thus, Hoogendoorn (*4*) further developed the Kaplan–Meier approach by including partially constrained observations to the classification as constrained and unconstrained observations. For this modified Kaplan–Meier approach, the probability θ_j is determined, indicating the likelihood that a vehicle with a specific headway t_j is constrained or unconstrained. This modification is based on a threshold process (*4*). Hoogendoorn's second modification addresses the estimation of desired speed distributions on multi-lane highways (*9*). The author points out that his initial approach is not directly applicable to highways. First, small headways on highways are often caused by overtaking maneuvers, but in such cases the vehicles are not constrained by each other. Second, distinguishing between constrained and unconstrained vehicles based solely on headways is not suitable in congested situations. Consequently, in his modified Kaplan–Meier approach for highways, Hoogendoorn introduces a two-part probability function $\theta(d_j, \Delta v_j)$, which takes into account both the distance to the lead vehicle d_j and the relative speed Δv_j. A vehicle is considered more unconstrained, the greater its distance and speed difference compared with the lead vehicle. A high speed difference indicates an

overtaking maneuver. Hoogendoorn defines the probability θ_j of a vehicle being constrained as indicated in Equations 1 to 3 (*9*).

$$\theta_j = \theta_{gap_j} \cdot \theta_{speed_j} \quad (1)$$

$$\theta_{gap_j} = \begin{cases} 1, & \text{for } d_j \leqslant a_1 \\ 1 + \frac{a_1}{a_2} - \frac{d_j}{a_2}, & \text{for } a_1 < d_j < a_1 + a_2 \\ 0, & \text{for } d_j \geqslant a_1 + a_2 \end{cases} \quad (2)$$

$$\theta_{speed_j} = \begin{cases} 1, & \text{for } |\Delta v_j| \leqslant b_1 \\ 1 + \frac{b_1}{b_2} - \frac{|\Delta v_j|}{b_2}, & \text{for } b_1 < |\Delta v_j| < b_1 + b_2 \\ 0, & \text{for } |\Delta v_j| \geqslant b_1 + b_2 \end{cases}$$

$$(3)$$

where

d_j = gap to the lead vehicle (m),
a_1 = parameter (m), typical value by Hoogendoorn (*9*) = 20,
a_2 = parameter (m), typical value by Hoogendoorn (*9*) = 150,
Δv_j = relative speed between vehicle j and its lead vehicle (m/s),
b_1 = parameter (m/s), typical value by Hoogendoorn (*9*) = 2.5, and
b_2 = parameter (m/s), typical value by Hoogendoorn (*9*) = 2.5.

In this study, Equation 4, as proposed by Hoogendoorn, is utilized for the modified Kaplan–Meier approach for highways (*9*).

$$F(v^0) = 1 - \prod_{j=1}^{n_{v^0}} \left(\frac{n - j - 1}{n - j - \theta_j} \right) \quad (4)$$

where

v^0 = speed (m/s),
j = position of the considered vehicle j in the dataset (for v_j ascending and constrained before unconstrained vehicles),
v_j = driven speed of the vehicle j,
n_{v^0} = number of vehicles j with driven speed $v_j \leqslant v^0$,
n = number of recorded vehicles (all observations), and
θ_j = probability, that vehicle j is constrained.

Trajectory Data in the Context of Microscopic Traffic Simulation

Driven by technological advances in the development of camera drones and automatic image processing, various freely available trajectory datasets have been collected and processed to support the development of trajectory processing and the use of trajectories in the context of

Table 1. Overview Over Open Trajectory Datasets of Freeway Traffic

Dataset	Year	Segment length	Source
NGSIM	2006	600 m	Kovvali et al. (*14*)
highD	2018	420 m	Krajewski et al. (*15*)
HIGH-SIM	2021	2,400 m	Shi et al. (*16*)
Automatum	2021	240–280 m	Spannaus et al. (*17*)
I-24 MOTION	2022	4 mi	Gloudemans et al. (*18*)
A2 dataset	2022	1,200 m	This paper

simulation models. Table 1 shows a selection of currently available datasets. Research has been conducted on subjects such as traffic flow analysis and modeling, traffic-related estimation and prediction, traffic flow model calibration, vehicle trajectory data cleaning, and vehicular ad hoc network-related studies based on the NGSIM dataset (*13*).

Hamdar et al. (*19*) and Hao et al. (*20*) calibrated and validated their proposed car-following models based on trajectories of the NGSIM dataset by comparing speed, acceleration, time headway, and space headway. Przybyla et al. (*21*) analyzed trajectories to identify driver behavior and characteristics to propose a data-driven car-following model, whereas Li et al. (*22*) examined the distribution of headways collected from NGSIM trajectories. To calibrate Wiedemann's car-following model in PTV Vissim, Durrani et al. estimated the model parameters directly from the NGSIM trajectories (*23*). The study showed that vehicle-following behavior is significantly different among various vehicle classes, and recommends setting parameter values as distributions instead of fixed values. Hale et al. propose a method to calibrate a microscopic traffic flow simulation of a freeway segment by comparing and minimizing the difference between observed and simulated trajectories (*24*). To ensure comparing sufficiently similar trajectories, both observed and simulated trajectories have been binned into specific groups. Li et al. (*25*) and Zhong et al. (*26*) propose approaches similar to Hale et al. (*24*) to calibrate the intelligent driver model (IDM) with NGSIM trajectories, whereas Kurtc (*27*) demonstrated that the IDM is capable of reproducing naturalistic vehicle trajectories based on the highD dataset.

The literature review shows that vehicle trajectories are a suitable data source for the calibration of microscopic traffic flow simulations, but most studies focus on driving behavior. To the best of our knowledge, there are currently no published approaches for determining desired speeds from trajectory data. Nonetheless, the studies mentioned above indicate that trajectory data is capable of providing the required information to estimate desired speeds, such as speed, acceleration, as well as time and space headways to the surrounding vehicles.
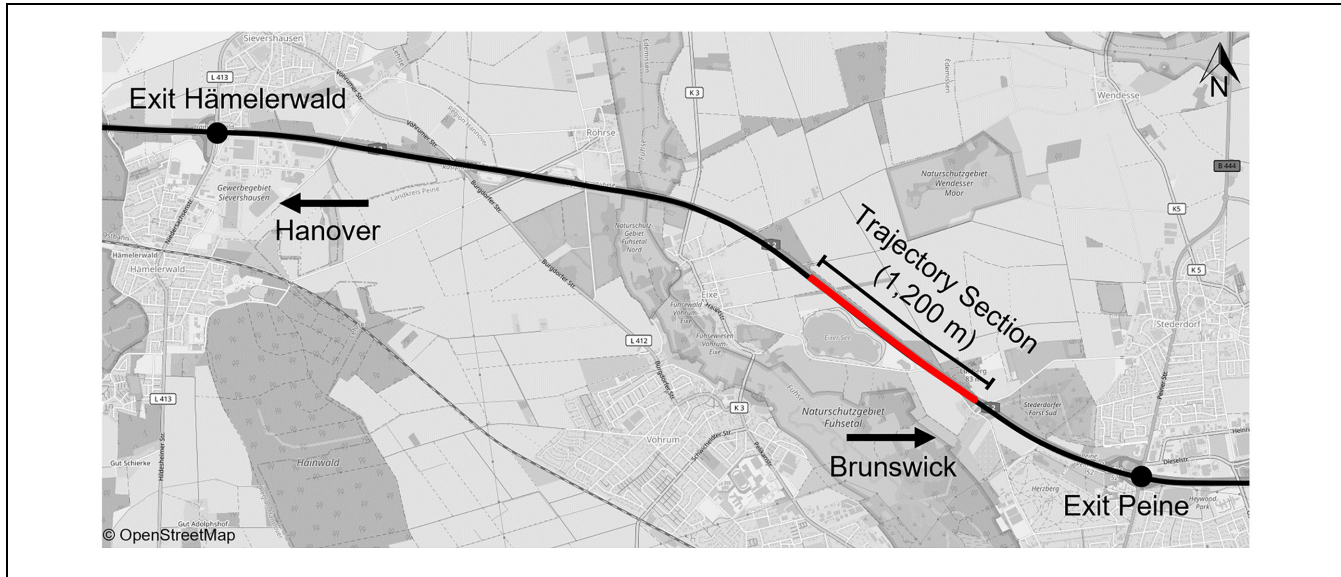
Figure 1. Location of the A2 dataset.

Method

Used Data

To estimate desired speeds from vehicle trajectories, it is beneficial to have a trajectory dataset with wide spatial coverage. As the length of the recorded trajectory increases, it is more likely to identify time periods in which the vehicle is not constrained by surrounding vehicles and thus is able to attain its desired speed. Considering the available datasets presented in Table 1, the recorded trajectories, except for the HIGH-SIM and the I-24 MOTION dataset, are too short to fully utilize the advantages of vehicle trajectories. However, the HIGH-SIM dataset only includes a single recording of 40 min containing 2,182 vehicles for one driving direction, whereas longer recordings and more captured vehicles would be beneficial. The I-24 MOTION dataset is not yet available at the time of this study.

For this reason, we use a recently recorded vehicle trajectory dataset that was acquired through a collaboration with a commercial provider. This dataset focuses on a segment of the German A2 freeway between the cities of Hanover and Brunswick, and is referred to as the A2 dataset in the following. It is located 1,000 m west of the exit Peine and 5 km east of the exit Hämelerwald (see Figure 1). The segment is composed of three lanes per driving direction and has no speed limit. The drone recording was carried out on Tuesday, November 1, 2022, during the morning peak hour, and covered a distance of 1,200 m in both directions. Overall, 11,488 vehicles (passenger cars and heavy vehicles) have been captured. Because of technical limitations (battery capacity of the camera drone), the data collection resulted in
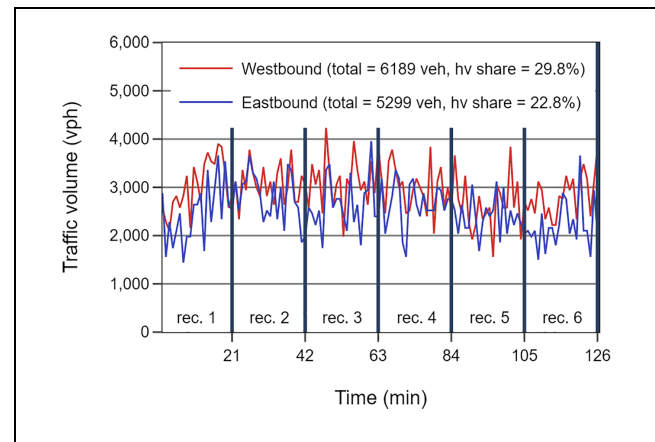

Figure 2. Traffic volume of the A2 dataset.
Note: vph = vehicles per hour; veh = vehicle; hv = heavy vehicle share; rec. = recording.

six separate recordings of 21 min. Figure 2 illustrates the traffic volume (passenger cars and heavy vehicles), and Figure 3 illustrates the spatiotemporal evolution of the average passenger car speed aggregated to 50-m and 1-min intervals. In the westbound driving direction, there are slightly higher traffic volumes than in the eastbound direction and no noticeable traffic instabilities. Whereas in the eastbound driving direction, congestion is beginning to form in Recording 5. This breakdown is most probably caused by an oversaturation of the exit Peine, which is located 1,000 m downstream of the recorded area.

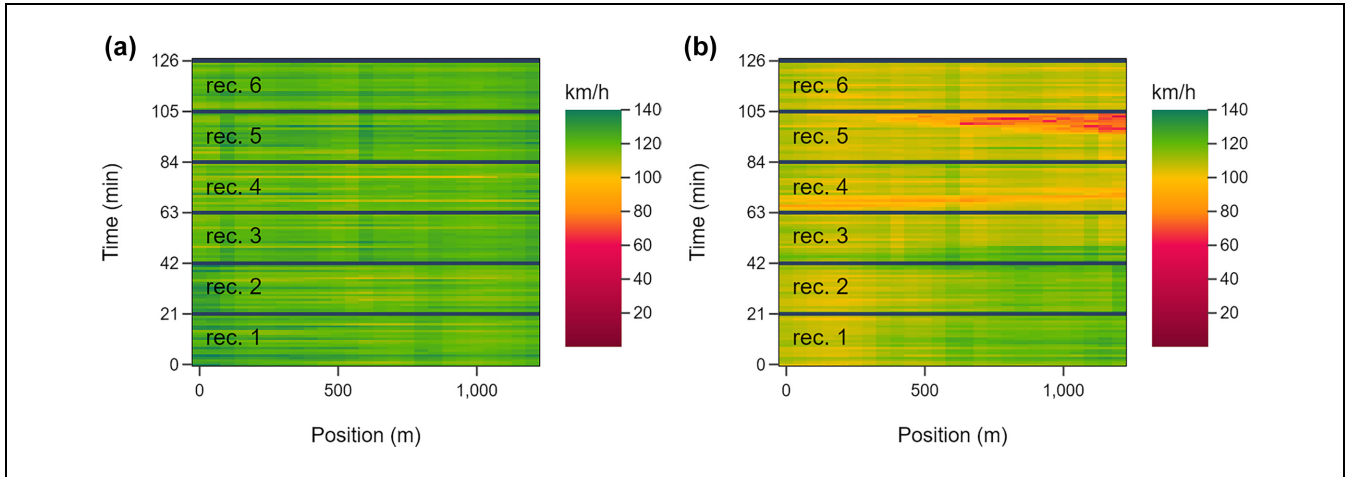The data preprocessing conducted for the A2 dataset is described in the following sections.

Figure 3. Spatiotemporal evolution of average passenger car speed of the A2 dataset (*a*) westbound, and (*b*) eastbound.
Note: rec. = recording.

Data Preprocessing

Lane Matching. The recorded vehicle trajectories are initially projected onto world coordinates. To be able to determine time headways, we perform a projection onto roadway coordinates and a lane matching. Automatic extraction of map lane boundaries is not used because of the distorted positions of the vehicles. Instead, we carry out the determination of the lane boundaries manually. To support this process, we generate cross-sections at intervals of 10 m, and create histograms of the lateral positions of the intersecting trajectories. The lane boundaries are then placed as close as possible to the minima of the histograms.

Subsequently, the lane boundaries are iteratively improved. For this purpose, we introduce a metric to determine vehicles overtaking each other within the same lane based on the current lane boundaries. The objective is to minimize this metric by slightly adjusting the lane limits.

Vehicle Dynamics Calculation. We compute the speed of a vehicle $v(t)$ as the travel speed between the positions of the vehicle at times $t - 0.5$ s and $t + 0.5$ s. The acceleration is then calculated based on the instantaneous rate of change of velocity. Figure 4 illustrates the distribution of all acceleration values in the A2 dataset. The resulting accelerations are comparable with the distribution of acceleration values that Montanino and Punzo reconstructed for the NGSIM dataset (*28*).

Headway Calculation. At each time step, we determine the preceding and following vehicles of a vehicle. To achieve this, we cluster the vehicles into groups based on their lane for each time step. Each group is then sorted based
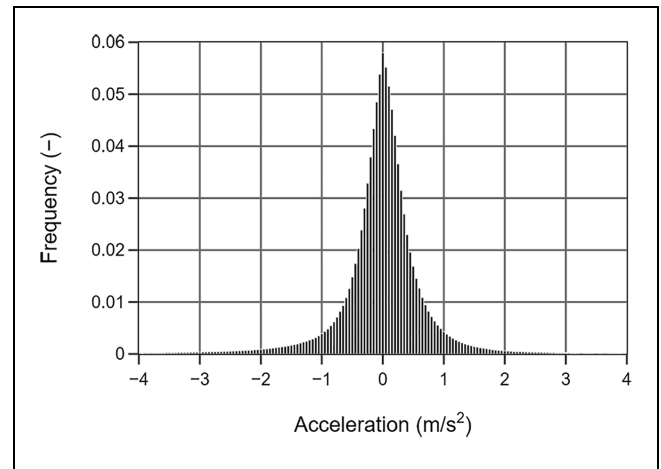

Figure 4. Distribution of acceleration values after data processing.

on the distance traveled, allowing for retrieval of the preceding vehicles.

Then, we calculate the distances between each vehicle and the surrounding vehicles and, based on these distances, we compute space and time headways. The time headway is based on the speed of the following vehicle.

Application of the Modified Kaplan–Meier Approach

Applying the modified Kaplan–Meier approach requires individual vehicle data from stationary detectors. These data need to be synthesized from the vehicle trajectory datasets and are further referenced as synthetic stationary detectors. To do this, the trajectories are cut every 50 m along the recorded segment (see Figure 5), and for each of these stationary detectors, the current speed of every vehicle and its space headway to the preceding vehicle, as well
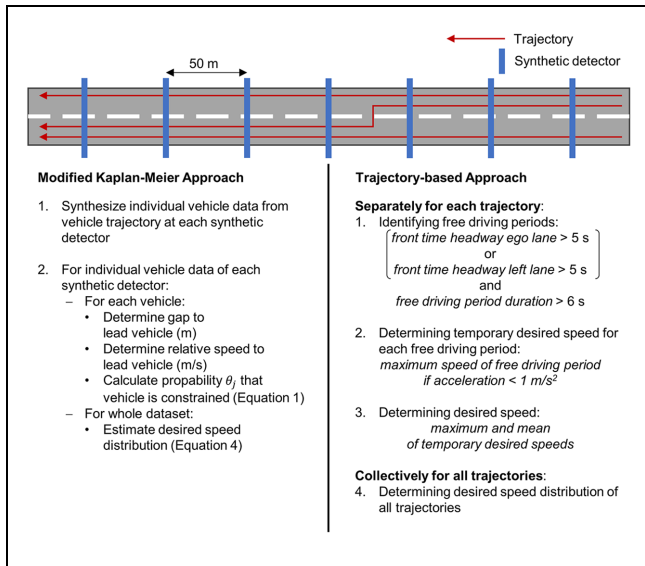
Figure 5. Estimating desired speed distributions with the modified Kaplan–Meier approach and determining desired speeds based on vehicle trajectory data.

as the speed difference to the preceding vehicle, are calculated. Based on these values, we determine for each vehicle the probability that it is constrained while driving (see Equation 1). The modified Kaplan–Meier estimation is carried out individually for each of these synthetic stationary detectors, resulting in one desired speed distribution every 50 m along the freeway segment, as well as for the combined data of all synthetic detectors of each direction.

Determining Desired Speeds Based on Vehicle Trajectory Data

The modified Kaplan–Meier approach is limited to observing a vehicle only at a single point in space and time. In contrast, the method proposed in this study utilizes vehicle trajectories to determine desired speeds as we examine a vehicle over a specific time duration and distance. In the first step, we identify free driving periods during which an ego vehicle is driving without being constrained by preceding vehicles. We define being constrained as having a front time headway of less than a threshold of 5 s. Other studies use smaller thresholds (e.g., Geistefeldt [5]), but to ensure that the vehicles are unconstrained we opt for this conservative definition. This assessment considers both the preceding vehicle in the same lane and in the left lanes of the ego vehicle, whereas the headway to at least one of these preceding vehicles must be higher than the threshold. To ensure that the driver is conscious about a free driving period, only periods exceeding a minimum free driving period duration of 6 s are considered further. This represents a duration in which a driver has enough time

to perceive their current situation and potentially adjust their speed to match their desired speed. This parameter value is a best guess that we find reasonable but have not empirically investigated so far. For this reason, its impact on the proportion of vehicles for which a desired speed can be determined will be discussed later as part of a sensitivity analysis.

In the next step, we determine a temporary desired speed for each free driving period. For this purpose, we identify the highest speed attained during the period and verify whether the vehicle's acceleration, by the time it reaches this maximum speed, remains below a threshold of $1 \frac{m}{s^2}$. This criterion ensures that the vehicle has achieved its desired speed within the given period and that the acceleration toward the desired speed is not prematurely interrupted because of the presence of a preceding vehicle.

If multiple free driving periods are identified for a vehicle, various temporary desired speeds may result. To determine a desired speed distribution for the collective of vehicles on the segment, these temporary desired speeds need to be aggregated into a single desired speed value for each vehicle. Several approaches can be used for this aggregation, such as selecting the maximum temporary desired speed, the mean of the temporary desired speeds, or the temporary desired speed that occurs most frequently. Since this study is based on trajectory datasets with a length of 1,200 m, and thus only a limited amount of free driving periods are expected to be identified per vehicle, we carry out the estimation with the mean and maximum temporary desired speed and discuss the results. The differences between these temporary desired speeds and the stability of the desired speed over the course of the road segment are examined further.

Results

Modified Kaplan–Meier Approach

We applied the Kaplan–Meier approach separately to the data from each individual synthetic detector, allowing the estimation of desired speed distributions at 50-m intervals for both driving directions. Figure 6 shows the resulting desired speed distributions. The color shades illustrate the different synthetic detectors, with the colors darkening along the driving direction. In the westbound direction, the median of the distributions is approximately 15 km/h higher than in the eastbound direction. This difference could potentially be attributed to the presence of a freeway exit about 1,000 m downstream and the occurrence of merging processes toward the end of the trajectories in the eastbound direction.

Figure 7 depicts the variation of desired speed distributions in 50-m increments along the segment. Each line corresponds to the evolution of a percentile, with the percentiles plotted at 5% intervals. In the westbound
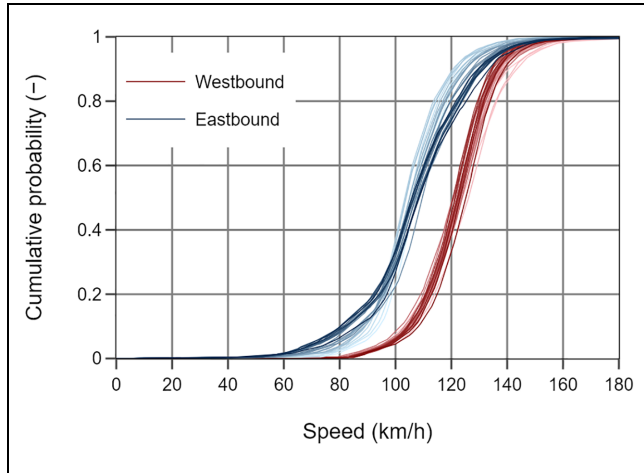
Figure 6. Desired speed distributions of passenger cars for every synthetic detector based on the Kaplan–Meier approach.

direction, the desired speed distributions remain relatively constant, with percentile values as well as the median fluctuating in an interval of 5 km/h, with the median remaining between 120 and 125 km/h. In contrast, in the eastbound direction, it is evident that the dispersion of the desired speed distribution increases with position, particularly as the distance to the exit decreases. Despite this variation, the median of the distribution remains approximately constant at 105 km/h along the trajectory. Interestingly, both the lower and upper percentiles widen equally. This could be because, on the one hand, vehicles may reduce their speed without direct influence from other vehicles, and on the other hand, there tend to be fewer vehicles on the left lane near the exit, providing more space for vehicles staying on the freeway to drive faster.

Trajectory-Based Approach

Estimating the desired speed from vehicle trajectories enables the examination of the stability of a vehicle's desired speed throughout the duration of its trajectory. As Figure 8 illustrates, we could only identify one free driving period for most of the vehicles because of the 1,200 m length of the vehicle trajectory dataset. To examine the intra-vehicle stability of the desired speed, we focused only on vehicles for which more than one free driving period could be identified. To assess the intra-vehicle stability of the desired speed, the difference between a vehicle's highest and lowest temporary desired speeds is determined. While other statistical metrics like the standard deviation of the temporary desired speeds are conceivable, they have been omitted because of the limited amount of free driving periods available for evaluation. This maximum desired speed delta exceeds 5 to 7 km/h for 40% of the vehicles, depending on the driving direction.

Figure 9 displays the desired speed distributions estimated using the trajectory-based approach. Similar to the results of the modified Kaplan–Meier approach, the westbound desired speed distribution is approximately 15 km/h higher in the median range. As we mentioned above, this difference could potentially be attributed to the presence of a freeway exit about 1,000 m downstream and the occurrence of merging processes toward the end of the trajectories in the eastbound direction. We have considered two variants of the method concerning the determination of the desired speed for each vehicle based on the temporary desired speeds: the maximum and mean of the temporary desired speeds. These results were compared with the estimations obtained with the modified Kaplan–Meier approach, which were computed based on the entire set of synthetic detector data for each driving direction. Because of the limited amount of vehicles for which we have been able to identify multiple free driving periods, the results of the maximum and mean of the temporary desired speed mostly align, as they yield the same result for vehicles with only one free driving period.

We apply non-parameterized bootstrapping with 1,000 iterations separately for each driving direction to further examine the significance of the obtained results. Table 2 presents the results for the distributions' mean, standard deviation, and median for different sample sizes. Regardless of the sample size, the median in the westbound driving direction yields a standard error of about 1.0 km/h, and in the eastbound driving direction, it is lower than 1.5 km/h. Comparing the results of the trajectory-based approach with the modified Kaplan–Meier approach, we observed similar outcomes in both driving directions: the shapes of the desired speed distributions largely align, but the values at the median level are approximately 5 to 10 km/h higher with the trajectory-based approach.

Discussion

Comparing the proposed trajectory-based approach with the established modified Kaplan–Meier approach indicates that the achieved results are reasonably plausible because they fall within a similar range as those of the modified Kaplan–Meier approach, and the shapes of the desired speed distributions align. The discrepancies can be explained as follows. Since the modified Kaplan–Meier approach relies on stationary detector data, each vehicle is considered only once at a single point in space and time. Therefore, it is possible that a vehicle has not yet reached its desired speed at that particular point and continues to accelerate after passing the stationary detector. Additionally, it is conceivable that apart from the free driving period
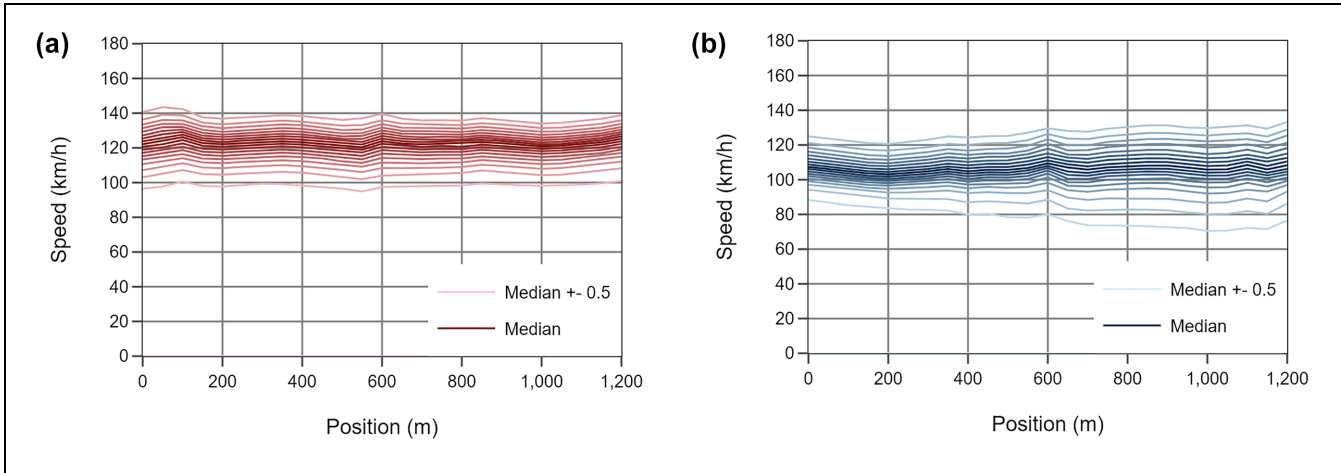
Figure 7. Percentiles of the desired speed distributions of passenger cars estimated with the modified Kaplan–Meier approach (*a*) westbound and (*b*) eastbound.
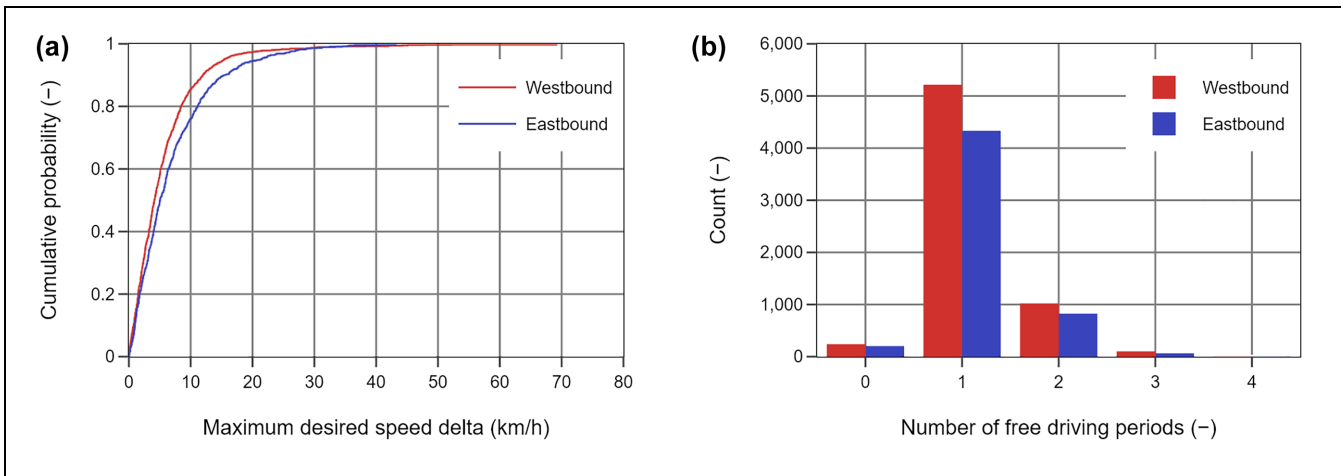

Figure 8. (*a*) Distribution of maximum desired speed distribution delta for passenger cars with multiple free driving periods, and (*b*) number of free driving periods per passenger car.

recorded by the stationary detector, there are other free driving periods for the same vehicle with higher temporary desired speeds. Under these circumstances, it is reasonable that the desired speed distribution obtained through the trajectory-based approach displays higher speeds than the modified Kaplan–Meier approach.

Considering the discrepancies observed both in the application of the modified Kaplan–Meier approach to multiple synthetic detectors and between the results of the modified Kaplan–Meier approach and the proposed trajectory-based approach, it can be concluded that the estimation of desired speed distributions is subject to a certain degree of inaccuracy, ranging between 5 and 10 km/h for the A2 dataset. This should be taken into account when calibrating microscopic traffic flow simulation models.
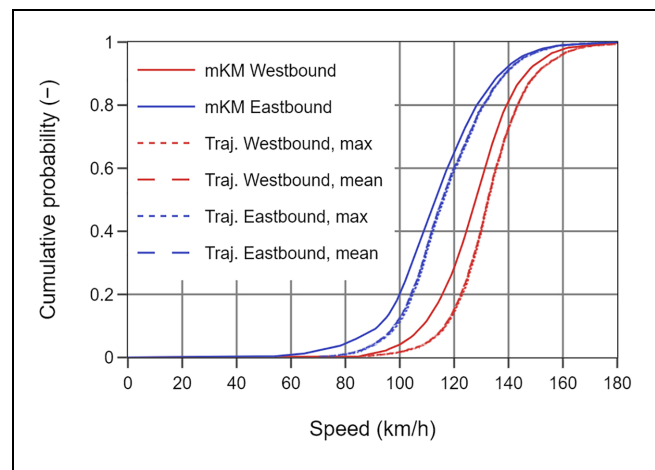

Figure 9. Desired speed distribution of passenger cars estimated with trajectory-based approach (traj.) and modified Kaplan–Meier approach (mKM).

Table 2. Standard Error of Estimation Statistics for Different Bootstrap Sample Sizes (1,000 Iterations)

| Statistic standard error (km/h) | Bootstrap sample size | | | | | | | |
| | Westbound | | | | Eastbound | | | |
	250	500	1,000	2,000	250	500	1,000	2,000
Mean	1.01	1.01	0.99	1.02	1.17	1.20	1.18	1.17
Standard deviation	0.92	0.98	0.98	0.98	0.99	0.93	0.94	0.92
Median	1.01	0.96	0.98	1.00	1.46	1.48	1.47	1.42
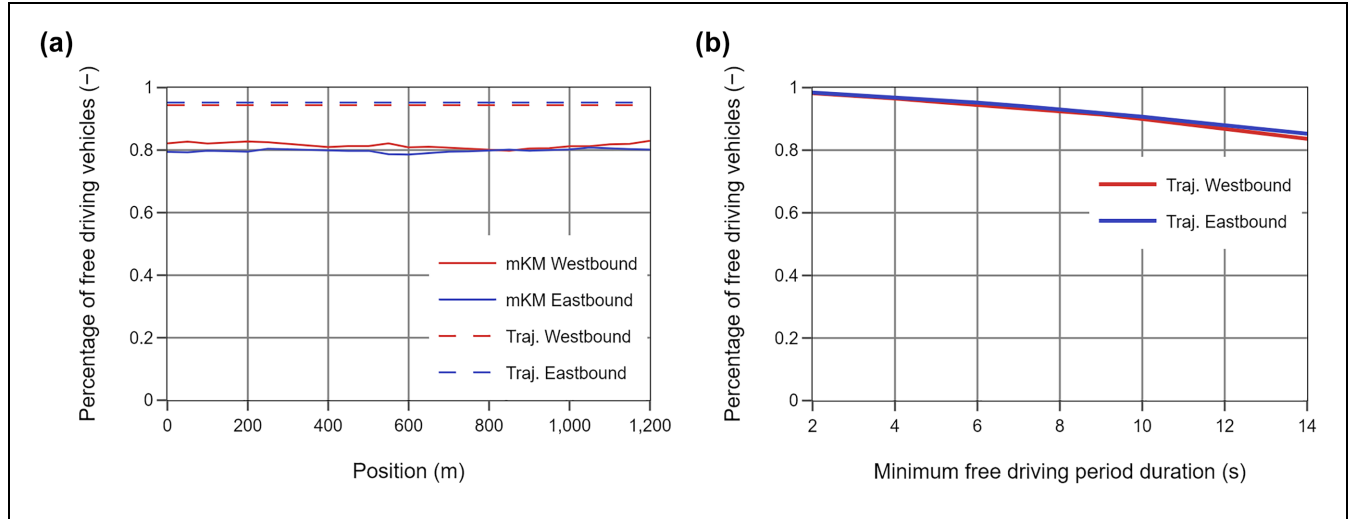
Figure 10. Percentage of free driving passenger cars depending on the used approach (*a*) and the min. free driving period duration (*b*).
Note: traj = trajectory-based approach; mKM = modified Kaplan–Meier approach.

The question of which method provides the "more accurate" results cannot be easily answered based solely on empirical evidence from the vehicle trajectory dataset. Without knowledge of the actual ground truth, that is, the true desired speed of each vehicle, we lack a definitive benchmark for comparison. Simulation experiments, where a specific desired speed distribution is defined as an input parameter and the presented approaches are used to attempt to reproduce this desired speed distribution from the resulting simulation outcomes, can provide valuable insights.

Nevertheless, this raises the question of whether the concept of a fixed desired speed over time and space accurately reflects reality. As shown in Figure 8, the temporary desired speeds can exhibit considerable variations across different free driving periods (in the A2 dataset, the maximum desired speed delta exceeds 5 to 7 km/h for 40% of the vehicles). Current microscopic traffic flow simulation tools do not consider these intra-vehicle instabilities of desired speed in their models. The extent to which these instabilities influence traffic flow cannot be conclusively determined here and must be addressed in further investigations.

Based on the A2 dataset, Figure 10*a* illustrates that depending on the position of the synthetic detector and driving direction, approximately 80% of the vehicles are unconstrained according to the modified Kaplan–Meier approach, allowing for direct determination of desired speed. As we do not know the desired speeds of the constrained vehicles, the desired speeds are estimated by applying the product-limit method. In contrast to the modified Kaplan–Meier approach, the trajectory-based approach lacks a mechanism to determine desired speeds for vehicles without a free driving period. This mechanism is not necessarily required here as we were able to identify a free driving period for 96% of the vehicles in both directions (see Figure 10*a*), allowing for the determination of desired speed.

The percentage of free driving vehicles is influenced by the choice of the minimum free driving period duration parameter. A sensitivity analysis reveals that the percentage of free driving vehicles decreases linearly with increasing the minimum free driving duration (see Figure 10*b*). We use 6 s as a threshold in this study.

Conclusion and Further Research

Desired speed distributions are crucial input parameters in microscopic traffic flow simulations. However, their determination is not straightforward, especially in situations of high traffic densities, where most vehicles are constrained by surrounding vehicles and cannot attain their desired speeds. The established method for estimating desired speed distributions is the modified Kaplan–Meier approach, which estimates the desired speed distribution based on single vehicle data from stationary detectors using the product-limit method. Nonetheless, the application of this method is limited to a stationary perspective, where each vehicle is only captured at a single point in space and time. In contrast, vehicle trajectory data provides the opportunity to observe each vehicle over a certain distance. In this paper, we propose a novel method for estimating desired speeds from vehicle trajectory data, as no similar approach has been found in the existing literature. This trajectory-based approach is compared with the established modified Kaplan–Meier approach, which is applied to synthesized stationary detector data derived from the vehicle trajectories. Furthermore, we evaluated the intra-vehicle stability of desired speeds using the outcomes obtained from both methods. To conduct our evaluations, we used vehicle trajectories covering a length of 1,200 m on a German freeway. The trajectory dataset covers 2 h of recording time, encompassing both driving directions and a total of 11,488 vehicles.

Applying the modified Kaplan–Meier approach on synthetic detectors at 50-m intervals along the recorded vehicle trajectories revealed that the desired speed distribution widens as the freeway exit is approached, with lower percentiles experiencing a notable decrease in speed. Analyzing the trajectory data demonstrated that 40% of the vehicles exhibited a delta of temporary desired speeds exceeding 5 to 7 km/h between different free driving periods. These findings suggest that the desired speed of a vehicle does not remain constant along its route, although the obtained results are based on only 1,200-m-long trajectories, and longer trajectories are necessary to verify them further.

Comparing our proposed method based on vehicle trajectory data with the established modified Kaplan–Meier approach, it becomes evident that the obtained results are reasonably plausible. The estimated desired speed distributions lie within a comparable range to those of the modified Kaplan–Meier approach, and the shapes of the distributions align. For both approaches, the westbound desired speed distribution is approximately 15 km/h higher in the median range, most probably because of the presence of a freeway exit about 1,000 m downstream and the occurrence of merging processes toward the end of the trajectories in the eastbound

direction. The desired speed distributions obtained from the trajectory-based approach are approximately 5 to 10 km/h higher than those estimated by the modified Kaplan–Meier approach. This difference in results is possibly because the trajectory-based approach observes each vehicle over a longer distance rather than just at stationary points. While this finding holds true for both driving directions in the A2 dataset, this investigation should be conducted in the future for additional road segments to validate the findings. Nevertheless, it can be concluded that the estimation of desired speed distributions is subject to a certain degree of inaccuracy. This should be considered in the context of calibrating microscopic traffic flow simulations.

The analysis of the trajectory data enables the identification of a free driving period for almost every vehicle in the A2 dataset, facilitating the determination of desired speeds. However, at higher traffic densities than those present in the current dataset, this approach may not be as straightforward. Future studies should explore and refine the proposed trajectory-based approach to deliver plausible results even when free driving periods cannot be identified for all vehicles. Moreover, all analyses of desired speeds in this study pertain to passenger cars. The extent to which the share of heavy vehicles influences these desired speeds and whether the approach can also be used to estimate desired speeds for heavy vehicles needs to be further investigated in subsequent studies. To extrapolate empirical findings and compensate for the lack of a ground truth, microscopic traffic flow simulations should be employed. Using these simulations, the proposed approach can be applied to various scenarios and traffic densities, allowing for an examination of how well the desired speed distributions defined as input parameters in the simulation can be reproduced by the approach based on the simulation results. Further investigations into the intra-vehicle stability of desired speed also require additional vehicle trajectory datasets, ideally covering longer stretches, such as the I-24 MOTION project (*27*), whose data was not available at the time of this study. In this context, it is important to explore the potential impact of intra-vehicle instability of desired speed on traffic flow and to assess whether and how such variability should be represented in microscopic traffic flow simulations.

Author Contributions

Declaration of Conflicting Interests

Funding

ORCID iDs

Marvin V. Baumann https://orcid.org/0000-0001-5952-7112
Claude M. Weyland https://orcid.org/0000-0002-2936-1067
Jan Ellmers https://orcid.org/0009-0007-7402-715X
Lea Fuchs https://orcid.org/0009-0004-3222-8218
Josephine Grau https://orcid.org/0009-0001-2852-2551
Peter Vortisch https://orcid.org/0000-0003-1647-2435

References

1. Botma, H., The Free Speed Distribution of Drivers: Estimation Approaches. In *Five Years 'Crossroads of Theory and Practice'* (P. Bovy, ed.), Delft University Press, Netherlands, 1999, pp. 1–22.
2. Wiedemann, R., H. Hubschneider, and H. Fritsche. *Zusammenhang zwischen der Wunschgeschwindigkeit und relevanten Kenngrößen des Verkehrsablaufs auf Autobahnen—Abschlussbericht* (Correlation Between Desired Speed and Relevant Traffic Flow Parameters on Freeways). Institut für Verkehrswesen—Universität Karlsruhe, Germany, 1982.
3. Dong, J., N. Hawkins, A. Houchin, N. Shafieirad, and C. Lu. *Vissim Calibration for Urban Freeways: Final Report*. Institute for Transportation, Iowa State University, Ames, 2015.
4. Hoogendoorn, S. P., Unified Approach to Estimating Free Speed Distributions. *Transportation Research Part B: Methodological*, Vol. 39, No. 8, 2005, pp. 709–727.
5. Geistefeldt, J. Empirical Estimation of Desired Speed Distributions for Microscopic Traffic Simulation. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2490: 41–46.
6. Lipshtat, A. Effect of Desired Speed Variability on Highway Traffic Flow. *Physical Review E*, Vol. 79, No. 6, 2009, p. 066110.
7. Farzaneh, M., and H. Rakha. Impact of Differences in Driver-Desired Speed on Steady-State Traffic Stream Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2006. 1965: 142–151.
8. Kaplan, E. L., and P. Meier. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, Vol. 53, No. 282, 1958, pp. 457–481.
9. Hoogendoorn, S. P. Vehicle-Type and Lane-Specific Free Speed Distributions on Motorways: A Novel Estimation Approach Using Censored Observations. *Transportation Research Record: Journal of the Transportation Research Board*, 2005. 1934: 148–156.
10. Geistefeldt, J. *Verkehrsablauf und Verkehrssicherheit auf Autobahnen mit vierstreifigen Richtungsfahrbahnen* (*Traffic Flow and Traffic Safety on Four-Lane Freeways*). PhD thesis. Ruhr-Universität Bochum, 2007.
11. Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th International Symposium on Transportation and Traffic Theory*, Elsevier, Amsterdam, College Park, MD, 2005, pp. 125–144.
12. Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2007. 2027: 1–8.
13. He, Z. Research Based on High-Fidelity NGSIM Vehicle Trajectory Datasets: A Review. *Research Gate*, 2017, pp. 1–33. DOI: 10.13140/RG.2.2.11429.60643/1
14. Kovvali, V. G., V. Alexiadis, and L. Zhang, *Video-Based Vehicle Trajectory Data Collection*. Transportation Research Board, Washington, D.C., 2007.
15. Krajewski, R., J. Bock, L. Kloeker, and L. Eckstein. The HighD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. *Proc., 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, New York, 2018, pp. 2118–2125.
16. Shi, X., D. Zhao, H. Yao, X. Li, D. K. Hale, and A. Ghiasi. Video-Based Trajectory Extraction with Deep Learning for High-Granularity Highway Simulation (HIGH-SIM). *Communications in Transportation Research*, Vol. 1, 2021, p. 100014.
17. Spannaus, P., P. Zechel, and K. Lenz. Automatum Data: Drone-Based Highway Dataset for the Development and Validation of Automated Driving Software for Research and Commercial Applications. *Proc., 2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, New York, 2021, pp. 1372–1377.
18. Gloudemans, D., Y. Wang, J. Ji, G. Zachar, W. Barbour, and D. B. Work. I-24 MOTION: An Instrument for Freeway Traffic Science. *arXiv Preprint arXiv:2301.11198*, 2023.
19. Hamdar, S. H., H. S. Mahmassani, and M. Treiber. From Behavioral Psychology to Acceleration Modeling: Calibration, Validation, and Exploration of Drivers' Cognitive and Safety Parameters in a Risk-Taking Environment. *Transportation Research Part B: Methodological*, Vol. 78, 2015, pp. 32–53.
20. Hao, H., W. Ma, and H. Xu. A Fuzzy Logic-Based Multi-Agent Car-Following Model. *Transportation Research Part C: Emerging Technologies*, Vol. 69, 2016, pp. 477–496.
21. Przybyla, J., J. Taylor, J. Jupe, and X. Zhou. Estimating Risk Effects of Driving Distraction: A Dynamic Errorable Car-Following Model. *Transportation Research Part C: Emerging Technologies*, Vol. 50, 2015, pp. 117–129.
22. Li, L., X. Chen, and Z. Li. Asymmetric Stochastic Tau Theory in Car-Following. *Transportation Research Part*

F: Traffic Psychology and Behaviour, Vol. 18, 2013, pp. 21–33.

23. Durrani, U., C. Lee, and H. Maoh. Calibrating the Wiedemann's Vehicle-Following Model Using Mixed Vehicle-Pair Interactions. *Transportation Research Part C: Emerging Technologies*, Vol. 67, 2016, pp. 227–242.

24. Hale, D. K., A. Ghiasi, F. Khalighi, D. Zhao, X. Li, and R. M. James. Vehicle Trajectory-Based Calibration Procedure for Microsimulation. *Transportation Research Record: Journal of the Transportation Research Board*, 2023. 2677: 1764–1781.

25. Li, L., X. M. Chen, and L. Zhang. A Global Optimization Algorithm for Trajectory Data Based Car-Following Model Calibration. *Transportation Research Part C: Emerging Technologies*, Vol. 68, 2016, pp. 311–332.

26. Zhong, R. X., K. Y. Fu, A. Sumalee, D. Ngoduy, and W. H. K. Lam. A Cross-Entropy Method and Probabilistic Sensitivity Analysis Framework for Calibrating Microscopic Traffic Models. *Transportation Research Part C: Emerging Technologies*, Vol. 63, 2016, pp. 147–169.

27. Kurtc, V. Studying Car-Following Dynamics on the Basis of the HighD Dataset. *Transportation Research Record: Journal of the Transportation Research Board*, 2020. 2674: 813–822.

28. Montanino, M., and V. Punzo. Making NGSIM Data Usable for Studies on Traffic Flow Theory: Multistep Method for Vehicle Trajectory Reconstruction. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2390: 99–111.