

The 15th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 23-25, 2024, Hasselt, Belgium

How well can we really estimate desired speeds? - A simulation study on the modified Kaplan-Meier approach

Marvin V. Baumann^{a,*}, Claude M. Weyland^a, Josephine Grau^a, Lea Fuchs^a, Peter Vortisch^a

^a*Institute for Transport Studies, Karlsruhe Institute of Technology, Otto-Ammann-Platz 9, D-76131 Karlsruhe, Germany*

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 due to 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. Due to a lack of a ground truth, it is not possible to validate the estimation results based on empirical data. Therefore, we employ a microscopic traffic simulation model as a virtual testbed to generate a synthetic ground truth. Our results highlight that when applying the modified Kaplan-Meier approach, it is crucial to ensure that the underlying empirical data are adequately filtered to exclude periods where the traffic density on the left lane exceeds 10 veh/km. Otherwise, the estimation may not yield valid results.

© 2024 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the Conference Program Chairs

Keywords: desired speed; free speed; Kaplan-Meier approach; microscopic traffic flow simulation; simulation study; PTV Vissim

1. Introduction

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 move 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 more significant disparity between desired and driven speeds. Furthermore, road users with higher desired speeds are more likely to be constrained than those with lower desired speeds [1]. Relying solely on the free speeds of a subset of vehicles to infer the desired speed distribution for all road users would result in a

* Corresponding author. Tel.: +49 721 608 47784

E-mail address: marvin.baumann@kit.edu

systematic underrepresentation of high desired speeds. However, this approach is used in certain studies to determine desired speeds on highways [11, 2]. The most established method for estimating desired speed distributions is the modified Kaplan-Meier approach [5, 4]. This approach can be applied to single vehicle data recorded by stationary detectors, consisting of both constrained and unconstrained vehicles.

The accuracy of these methods in estimating the desired speeds of the entire driver collective cannot be fully validated through empirical investigations. A definitive ground truth is not available, as it would necessitate surveying all drivers retrospectively to ascertain their desired speeds. Hence, validating the estimation of desired speeds can only be done based on synthetic data. Microscopic traffic flow simulations offer a practical solution for creating this data source, allowing for a realistic depiction of traffic flow, with the desired speed distribution as a configurable input parameter. In this study, we utilize a simulation model of a German freeway as a case study to investigate under which conditions (e.g. traffic density) it is possible to reconstruct the input desired speed distribution from synthetic data using the modified Kaplan-Meier approach as formulated by Hoogendoorn [6]. Furthermore, we conduct a sensitivity analysis regarding the parameters of Hoogendoorn's formulation.

2. 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 [7], originally applied in the context of clinical studies. This method features the estimation of survival probabilities over a specific time period, e.g., after an operation. 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 traffic, the application of the Kaplan-Meier approach was initially introduced by Botma [1]. Hoogendoorn [5, 6] and Geistefeldt [3, 4] have further developed this approach.

Individual vehicle data from stationary detector recordings are used to estimate desired speed distributions. 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. Therefore, Hoogendoorn [5] 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 [5]. Hoogendoorn's second modification addresses the estimation of desired speed distributions on multi-lane highways [6]. The author points out that his initial approach is not directly applicable to highways. Firstly, small headways on highways are often caused by overtaking maneuvers, but in such cases, the vehicles are not constrained by each other. Secondly, distinguishing between constrained and unconstrained vehicles based solely on headways is unsuitable in congested situations. Therefore, 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 to the lead vehicle. A high speed difference indicates an overtaking maneuver. Hoogendoorn [6] defines the probability θ_j of a vehicle being constrained as indicated in equations (1) to (3).

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

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

$$\theta_{speed_j} = \begin{cases} 1, & \text{for } |\Delta v_j| \leq b_1 \\ 1 + \frac{b_1}{b_2} - \frac{|\Delta v_j|}{b_2}, & \text{for } b_1 < |\Delta v_j| < b_1 + b_2 \\ 0, & \text{for } |\Delta v_j| \geq b_1 + b_2 \end{cases} \quad (3)$$

with d_j : gap to the lead vehicle [m]
 a_1 : parameter [m], typical value by Hoogendoorn [6] = 20
 a_2 : parameter [m], typical value by Hoogendoorn [6] = 150
 Δv_j : relative speed between vehicle j and its lead vehicle [m/s]
 b_1 : parameter [m/s], typical value by Hoogendoorn [6] = 2.5
 b_2 : parameter [m/s], typical value by Hoogendoorn [6] = 2.5

In this study, equation (4), as proposed by Hoogendoorn [6], is utilized for the modified Kaplan-Meier approach for freeways.

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

with 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 $v_j \leq v^0$
 n : number of recorded vehicles (all observations)
 θ_j : probability, that vehicle j is constrained

3. Method

3.1. Simulation Setup

To generate synthetic data for the subsequent analyses, we construct a model in PTV Vissim 2024 [9], serving as a fictional case study. This model is designed to replicate a segment of a German freeway, spanning a total length of 10 km, featuring a three-lane configuration without merging or diverging segments. We place a detector in the middle

of the segment to record both single-vehicle data and aggregated macroscopic traffic flow metrics for our evaluations. The simulation duration for each simulation run is seven hours, with the initial 60 minutes serving as a lead time with a traffic volume of 1000 vehicles per hour, not considered in the evaluation. Subsequently, we increase the traffic volume in intervals of 30 minutes by 500 vehicles per hour until reaching a maximum traffic volume of 6500 vehicles per hour. The vehicle mix in this study is exclusively limited to passenger cars to avoid interactions between different vehicle classes with distinct desired speed distributions.

Accordingly, all vehicles in the simulation follow the same desired speed distribution. Therefore, we use a truncated normal distribution with a median of 120 km/h and a standard deviation of 30 km/h, ranging from 50 km/h to 220 km/h, which is realistically representing traffic flow on a German freeway [10]. This input desired speed distribution will undergo a sensitivity analysis later in this study. For the calibration of driving behavior, we utilize lane-specific parameter sets from Weyland [10], shown in Table 1 and Table 2. Further explanations on the driving behavior parameters and abbreviations can be found in the PTV Vissim software manual [9].

Table 1. Calibrated car following behavior for cars in Vissim.

lane	CC0 [m]	CC1 ¹ [s]	CC2 [m]	CC3 [s]	CC4 [m/s]	CC5 [m/s]	CC6 [1/(m·s)]	CC7 [m/s ²]	CC8 [m/s ²]	CC9 [m/s ²]
1	1.5	$\mu=0.75$	4	-8	-0.35	0.35	11.44	0.25	3.5	1.5
2	1.5	$\mu=0.60$	4	-8	-0.35	0.35	11.44	0.25	3.5	1.5
3	1.5	$\mu=0.90$	4	-8	-0.35	0.35	11.44	0.25	3.5	1.5

¹ please refer to Weyland [10] for the actual distributions of *CC1*

Table 2. Calibrated lane change behavior for cars in Vissim.

lane	MaxDecelOwn [m/s ²]	MaxDecelTrail [m/s ²]	DecelRedDistOwn [m per -1 m/s ²]	DecelRedDistTrail [m per -1 m/s ²]	AccDecelOwn [m/s ²]	AccDecelTrail [m/s ²]	MinClear [m]	FreeDriveTm [s]	SafeDistRedFact [-]	CoopDecel [m/s ²]	AdvMerg [yes/no]	VehRouteLookAhead [yes/no]	CoopLnChg [yes/no]	CoopLnChgSpeedDiff [km/h]	CoopLnChgCollTm [s]
1	-4	-3	300	200	-1.0	-0.75	0.5	15	0.6	-3	ja	ja	ja	10.8	10
2	-4	-3	300	200	-1.0	-0.75	0.5	5	0.6	-3	ja	ja	ja	10.8	10
3	-4	-3	300	200	-1.0	-0.75	0.5	5	0.6	-3	ja	ja	ja	10.8	10

Overall, we carry out ten simulation runs with different random seeds. Figure 1 illustrates the resulting aggregated macroscopic traffic flow metrics of the synthetic simulation data. The plots depict a realistic traffic flow on a German freeway with a right-hand traffic rule and no speed limit: While at low traffic volumes, the majority of vehicles travel on the right lane (lane 1), from around 4000 veh/h onwards, most traffic is accommodated on the left lanes (Figure 1, top right). Above 4000 veh/h, the highest traffic density is observed on the left lanes (Figure 1, bottom left). The initially larger differences in average speeds between the lanes at low traffic volumes decrease as traffic volumes increase (Figure 1, bottom right).

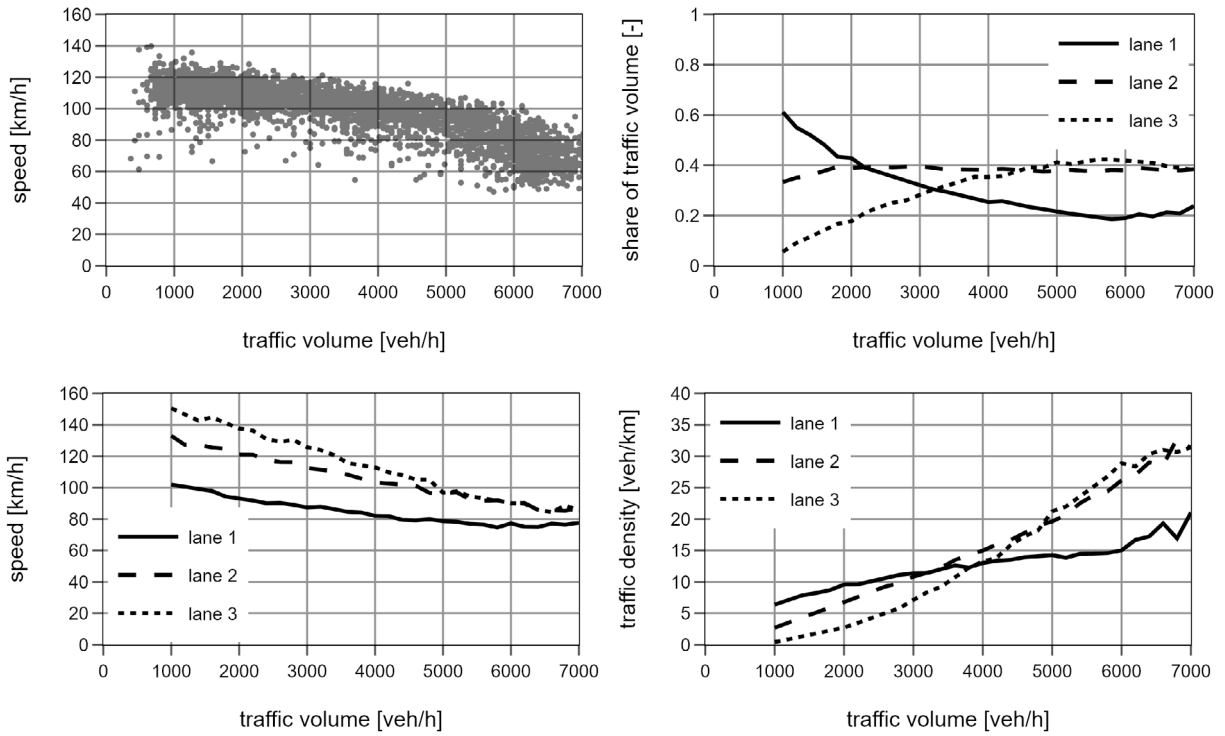


Fig. 1. Aggregated macroscopic traffic flow metrics of the synthetic simulation data (lane 1 = right lane).

3.2. Evaluation

Based on the synthetic single-vehicle data recorded at the detector in the simulation, the modified Kaplan-Meier approach is employed to estimate the desired speed distribution. We use the most established variant of the modified Kaplan-Meier approach by Hoogendoorn [6] with the parameter set recommended by the author, as introduced in Section 2. The data collected from all simulation runs is utilized to simulate observations over multiple days. Furthermore, the recorded data is classified based on the current traffic density in intervals of 5 veh/km. This classification only considers the traffic density in the left lane, as, in accordance with the right-hand traffic rule in Germany, its occupancy indicates overtaking opportunities. For each interval we calculate the average constraint per vehicle θ_{mean} according to equation (5). Subsequently, we compare the desired speed distribution estimated with the modified Kaplan-Meier approach with the input desired speed distribution. We utilize the Population Stability Index (PSI) [12], a symmetric derivation of the Kullback-Leibler Divergence [8], to quantify the deviations between the input desired speed distribution and the desired speed distribution estimated with the modified Kaplan-Meier approach. As a rule of thumb, a PSI below 0.1 indicates an appropriate model, while a PSI between 0.1 and 0.25 necessitates closer scrutiny due to the elevated PSI [12]. A PSI exceeding 0.25 suggests that the model should be discarded.

$$\theta_{mean} = \frac{\sum_{j=1}^n \theta_j}{n} \quad (5)$$

with n : number of detected vehicles [-]
 θ_j : probability of constrain [-], see equation (4)

During a sensitivity analysis, we examine the influence of the parameterization of the modified Kaplan-Meier approach. Regarding the parameter sets under investigation, we follow the sensitivity analysis conducted by Hoogendoorn [6]. Analogous to our previously described approach, we calculate the PSI for each parameter set. Subsequently, we explore the sensitivity regarding the input desired distribution by systematically varying the distribution's median and variance.

4. Results and Discussion

Figure 2 illustrates the estimated desired speed distributions depending on the traffic density on the left lane and demonstrates that the resulting desired speed distribution, and consequently the performance of the estimation, strongly depends on the traffic density and thus the traffic situation. The higher the traffic density and, therefore, the average constraint per vehicle θ_{mean} , the lower speeds the estimated desired speed distribution exhibits. The evaluation of the PSI (Table 3) reveals that a plausible estimation of the desired speed distribution is only feasible at traffic densities on the left lane below 10 veh/km. At higher traffic densities, there are either too few free-flowing vehicles or too much influence, leading to challenges in the estimation process. The joint analysis of all intervals (all traffic densities) also results in an inadequate estimation. Even though a sufficient number of unconstrained vehicles would be available, too many constrained vehicles distort the estimation.

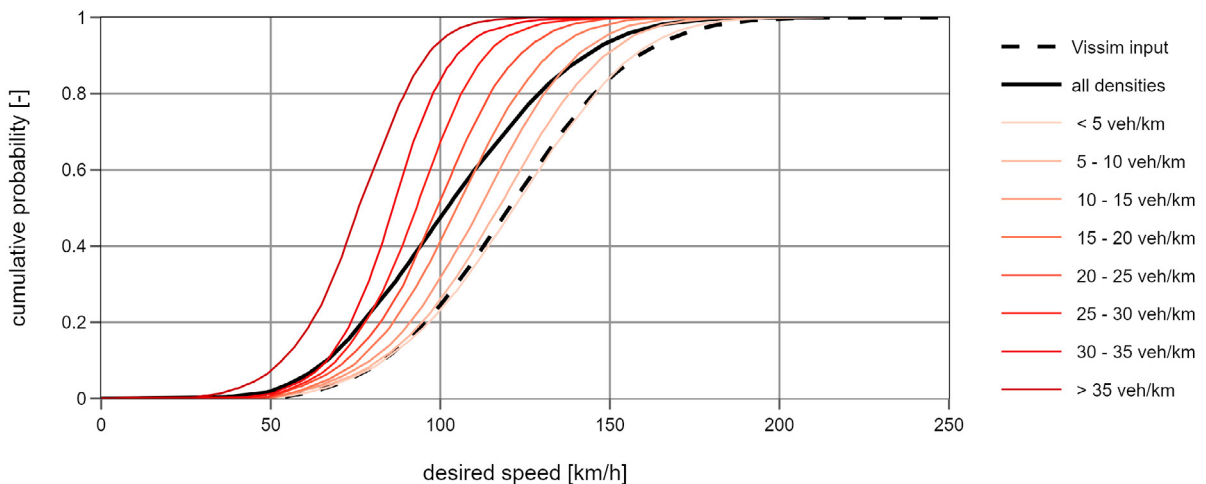


Fig. 2. Input desired speed distribution and estimated desired speed distributions depending on the traffic density on the left lane.

The sensitivity analysis regarding the parameterization of the modified Kaplan-Meier approach indicates that, with the tested parameter sets, the performance of the estimation cannot be significantly improved. Nevertheless, some interesting insights emerge: An increase in the parameter value leads to a rise in the average constraint (θ_{mean}) for all parameters. This is logical as higher parameter values make vehicles more constrained at higher following distances or speed differentials. The speed differential (b_1 and b_2) has a greater impact than the following distance (a_1 and a_2). This increased influence affects the performance of the estimation differently depending on the traffic density on the left lane. In general, it is observed that higher parameter values and a higher θ_{mean} for the same data sample (e.g., density lower than 5 veh/km) lead to a desired speed distribution with higher speeds. For lower traffic densities (< 5 veh/km), the performance of the estimation deteriorates because a larger proportion of vehicles is (incorrectly) assumed to be unable to achieve their desired speed, leading to their desired speed being estimated higher than the measured speed. For high traffic densities (e.g., 15 - 20 veh/km), the performance of the estimation can be improved with higher parameter values because overestimating the desired speeds of supposedly unconstrained vehicles counteracts the general underestimation of the desired speed distribution at high traffic densities. However, unrealistic parameter values would have to be chosen to achieve meaningful estimation results for high traffic densities, which is not advisable.

The sensitivity analysis regarding the input desired speed distribution is carried out separately for the median and the variance. An increase in the median results in a decrease in the average constraint (θ_{mean}), which, as seen before, leads to an improvement in the performance of the estimation for low traffic densities and a deterioration for high traffic densities. The same pattern holds for an increasing variance of the input desired speed distribution.

Table 3. Results of the sensitivity analysis.

input desired speed distribution		modified Kaplan-Meier parameters				density left lane [veh/km]	PSI	θ_{mean}	n
median	variance	a_1 [m]	a_2 [m]	b_1 [m/s]	b_2 [m/s]				
120	30	20	150	2.50	2.50	all	0.30	0.31	223468
120	30	20	150	2.50	2.50	< 5	0.02	0.15	37614
120	30	20	150	2.50	2.50	5 - 10	0.07	0.22	27201
120	30	20	150	2.50	2.50	10 - 15	0.22	0.25	24979
120	30	20	150	2.50	2.50	15 - 20	0.50	0.29	19136
120	30	20	150	2.50	2.50	20 - 25	0.99	0.31	18609
120	30	20	150	2.50	2.50	25 - 30	1.61	0.35	21674
120	30	20	150	2.50	2.50	30 - 35	2.39	0.41	22137
120	30	20	150	2.50	2.50	> 35	3.58	0.44	20850
120	30	0	150	2.50	2.50	< 5	0.01	0.12	37614
120	30	50	150	2.50	2.50	< 5	0.03	0.19	37614
120	30	0	150	2.50	2.50	15 - 20	0.56	0.24	19136
120	30	50	150	2.50	2.50	15 - 20	0.44	0.33	19136
120	30	20	100	2.50	2.50	< 5	0.01	0.12	37614
120	30	20	200	2.50	2.50	< 5	0.02	0.18	37614
120	30	20	100	2.50	2.50	15 - 20	0.54	0.26	19136
120	30	20	200	2.50	2.50	15 - 20	0.47	0.31	19136
120	30	20	150	0.00	2.50	< 5	0.01	0.06	37614
120	30	20	150	5.00	2.50	< 5	0.04	0.23	37614
120	30	20	150	0.00	2.50	15 - 20	0.77	0.10	19136
120	30	20	150	5.00	2.50	15 - 20	0.32	0.44	19136
120	30	20	150	2.50	1.25	< 5	0.01	0.13	37614
120	30	20	150	2.50	5.00	< 5	0.03	0.19	37614
120	30	20	150	2.50	1.25	15 - 20	0.56	0.25	19136
120	30	20	150	2.50	5.00	15 - 20	0.40	0.37	19136
100	30	20	150	2.50	2.50	< 5	0.03	0.18	33796
100	30	20	150	2.50	2.50	15 - 20	0.27	0.35	21652
110	30	20	150	2.50	2.50	< 5	0.02	0.16	34286
110	30	20	150	2.50	2.50	15 - 20	0.37	0.31	21033
130	30	20	150	2.50	2.50	< 5	0.02	0.14	39364
130	30	20	150	2.50	2.50	15 - 20	0.66	0.26	20110
140	30	20	150	2.50	2.50	< 5	0.01	0.13	43086
140	30	20	150	2.50	2.50	15 - 20	0.97	0.24	20914
120	20	20	150	2.50	2.50	< 5	0.05	0.21	46159
120	20	20	150	2.50	2.50	15 - 20	0.29	0.34	25832
120	25	20	150	2.50	2.50	< 5	0.02	0.17	40654
120	25	20	150	2.50	2.50	15 - 20	0.39	0.30	19919
120	35	20	150	2.50	2.50	< 5	0.03	0.14	34961
120	35	20	150	2.50	2.50	15 - 20	0.58	0.27	19101
120	40	20	150	2.50	2.50	< 5	0.03	0.13	33251
120	40	20	150	2.50	2.50	15 - 20	0.66	0.27	19610

5. Conclusion

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. Unfortunately, there is no empirical ground truth regarding the desired speed distribution, and therefore, examining the estimation's performance is not possible based on empirical data.

This study employed a microscopic traffic simulation model as a virtual testbed to generate synthetic data. Within the simulation model, the desired speed distribution can be defined as an input parameter and subsequently estimated based on the simulation results, enabling an evaluation of the performance of the estimation conducted with the modified Kaplan-Meier approach. The results indicate that the performance of the estimation is highly dependent on the traffic density on the left lane, as its occupancy, in accordance with the right-hand traffic rule in Germany, indicates overtaking opportunities. A plausible estimation of the desired speed distribution is only feasible at traffic densities on the left lane below 10 veh/km. The conducted sensitivity analysis reveals that even by adjusting the modified Kaplan-Meier approach's parameters, the estimation's performance cannot be significantly improved to justify its application at higher traffic densities. These findings are largely independent of the underlying input desired speed distribution.

The synthetic ground truth examined in this study is limited to a single vehicle class (passenger cars) with one specific input desired distribution. Future studies should explore how a diverse vehicle mix and the presence of heavy vehicles affect the estimation and how accurately multiple input desired speed distributions can be reconstructed separately from each other.

When applying the modified Kaplan-Meier approach, it is essential to ensure that the underlying empirical data are sufficiently filtered to exclude periods where the traffic density on the left lane exceeds 10 veh/km. Furthermore, the question arises regarding the potential advancement of the modified Kaplan-Meier approach to achieve valid estimation results at higher traffic densities. However, it becomes crucial to consider whether the concept of desired speeds remains meaningful when the majority of vehicles are constrained at high traffic densities.

References

- [1] Botma, H., 1999. The Free Speed Distribution of Drivers: Estimation Approaches, in: Bovy, P. (Ed.), Five years 'Crossroads of theory and practice', Delft University Press, Delft. pp. 1–22.
- [2] Dong, J., Hawkins, N., Houchin, A., Shafieirad, N., Lu, C., 2015. Vissim Calibration for Urban Freeways: Final Report. Technical Report. Institute for Transportation, Iowa State University.
- [3] Geistefeldt, J., 2007. Verkehrsablauf und Verkehrssicherheit auf Autobahnen mit vierstreifigen Richtungsfahrbahnen (eng.: Traffic Flow and Traffic Safety on Four-Lane Freeways). Ph.D. thesis. Ruhr-Universität Bochum.
- [4] Geistefeldt, J., 2015. Empirical estimation of desired speed distributions for microscopic traffic simulation. Transportation Research Record 2490, 41–46. Publisher: SAGE Publications Sage CA: Los Angeles, CA.
- [5] Hoogendoorn, S.P., 2005a. Unified approach to estimating free speed distributions. Transportation Research Part B: Methodological 39, 709–727. Publisher: Elsevier.
- [6] Hoogendoorn, S.P., 2005b. 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, 148–156 URL: <http://dx.doi.org/10.3141/1934-16>, doi:10.1177/0361198105193400116.
- [7] Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American statistical association 53, 457–481. Publisher: Taylor & Francis.
- [8] Kullback, S., Leibler, R.A., 1951. On information and sufficiency. The annals of mathematical statistics 22, 79–86. URL: <https://www.jstor.org/stable/2236703>. publisher: JSTOR.
- [9] PTV, 2023. PTV Vissim 2024 - Manual. 2023 ed., Karlsruhe.
- [10] Weyland, C.M., 2023. Mikroskopische Verkehrsflusssimulation von Autobahnen mit Streckenbeeinflussungsanlagen (eng.: Microscopic Traffic Flow Simulation of Freeways with Dynamic Freeway Control Systems). Ph.D. thesis. Karlsruhe Institute of Technology (KIT). doi:10.5445/IR/1000162768.
- [11] Wiedemann, R., Hubschneider, H., Fritsche, H., 1982. Zusammenhang zwischen der Wunschgeschwindigkeit und relevanten Kenngrößen des Verkehrsablaufs auf Autobahnen - Abschlussbericht (eng.: Correlation between Desired Speed and Relevant Traffic Flow Parameters on Freeways). Institut für Verkehrswesen - Universität Karlsruhe, Karlsruhe.
- [12] Yurdakul, B., 2018. Statistical properties of population stability index. Western Michigan University. URL: <https://search.proquest.com/openview/55a8dd6d5aeb966a4dacf5cf06b6d7b7/1?pq-origsite=gscholar&cbl=18750>.