Estimating the fundamental diagram using moving observers

van Erp, Paul; Thoen, Sebastiaan; Knoop, Victor; Hoogendoorn, Serge

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Estimating the fundamental diagram using moving observers

1ˢᵗ Paul B.C. van Erp
Transport & Planning
TU Delft
Delft, The Netherlands
p.b.c.vanerp@tudelft.nl

2ⁿᵈ Sebastiaan Thoen
Transport & Planning
TU Delft
Delft, The Netherlands
s.thoen@student.tudelft.nl

3ʳᵈ Victor L. Knoop
Transport & Planning
TU Delft
Delft, The Netherlands
v.l.knoop@tudelft.nl

4ᵗʰ Serge P. Hoogendoorn
Transport & Planning
TU Delft
Delft, The Netherlands
s.p.hoogendoorn@tudelft.nl

Abstract—The fundamental diagram (FD) describes the relation between the flow and density in equilibrium conditions. In this paper, we propose an estimation approach to estimate the FD based on data from moving observers. This approach consists of two main steps: (1) estimate flow and density for space-time areas based on trajectories of moving observers and the times and locations they are overtaken or being overtaken and (2) estimate the FD based on the {flow,density}-estimates. To evaluate and gain a deeper understanding of the proposed approach, a simulation study was conducted. This study shows that the {flow,density}-estimates provide valuable information to estimate the FD. Furthermore, the FDs belonging to the simulated traffic flow are estimated accurately. We realize that the second step is expected to be less accurate for traffic that behaves stochastic. Therefore, we provide a potential solution path to extend the second step in future work.

Index Terms—moving observers, Edie's definitions, fundamental diagram

I. Introduction

The fundamental diagram (FD) describes the relation between the macroscopic traffic flow variables flow q and density k when traffic is in equilibrium. In combination with $q = ku$, where u is the mean speed, it describes the relation between all three macroscopic variables in equilibrium. The (parameters of the) FD are valuable for traffic state estimation and prediction [1], and in traffic control measures, e.g., [2].

This paper proposes an alternative approach to estimate the FD based on traffic sensing data collected by moving observers. This method consist of two main steps. First, we use the individual passings observed by moving observers to estimate q and k for a set of areas in space-time using Edie's definitions [3]. Using these data in combination with Edie's definitions is new to FD

estimation. Secondly, we estimate the FD based on the {q, k}-estimates. Our methodology takes into account the representativeness of the {q, k}-estimates for the FD and does not require prior knowledge or assumptions related to any parameters of the FD. In this way it differs from existing methodologies that are discussed in the next section.

This paper is organized as follows. In Section II a literature overview is given that serves as background information for the proposed approach. Sections III and IV respectively explain the methodologies designed for the first and second step of the approach. Next, in Section V, the simulation study to test the proposed approach is explained, after which the results of the study are reported in Section VI. Conclusions and an outlook are provided in Section VII.

II. Background on fundamental diagram estimation

This section covers the important topics related to exposing the fundamental relation between the macroscopic traffic flow variables, i.e., the fundamental diagram (FD). For this purpose, we discuss: (1) the macroscopic description of traffic flow, (2) the traffic sensing data that is used to estimate the fundamental diagram and (3) the important characteristics of the proposed FD estimation methodology.

Traffic flow can be described on a macroscopic level using flow q and density k. Edie [3] provided the generalized definitions for these two variables for an space-time area:

$$q = \frac{\sum_i d_i}{A} \tag{1}$$

$$k = \frac{\sum_i ts_i}{A} \tag{2}$$

where the sum of the distance traveled and time spent by all vehicles with the area, i.e., $\sum_i d_i$ and $\sum_i ts_i$ are respectively the total travel distance (TTD) and total time spent (TTS). Furthermore, size of the space-time area is given by A. Given that flow and density are known, the mean speed for the space-time area is also known given the relation $u = \frac{q}{k}$.

Edie's definitions provide the true aggregated traffic conditions for the full space-time area. However, within this area different traffic states can be present. Cassidy [4] argues that there is no reason to expect Edie's q, k and u to lie on a reproducible function, as the values will be merely a weighted average of the different traffic states. If traffic behaves according to a concave FD (e.g., the triangular FD) the weighted average of different traffic states will lie on (if both states are on the same linear function) or below (otherwise) the FD. Furthermore, if one traffic state holds for the full space-time area, the traffic state will lie on the FD.

To estimate the FD we rely of observations of traffic flow, i.e., traffic sensing data. The dominant source of sensing data that are used to estimate the FD are stationary detectors that observe flow and speed aggregated over a time-period. For instance, Dervisoglu et al. [5] and Knoop and Daamen [6] use loop-detector data to fit/calibrate the FD. However, there are multiple important problems related to using stationary detectors for FD estimation: For instance, (1) density is underestimated (even when observing harmonic mean speeds) if traffic is non-stationary during the aggregation period, (2) the detectors do not detect traffic at standstill [6], and (3) FD estimation is limited to those road sections where detectors are installed [1]. Alternatively, it has been proposed to use probe data for FD estimation, e.g., Seo et al. [7]. The general downsides of probe data are that include individual driving behavior, e.g., aggressive vs. timid driver are likely to behave differently in the same situation, and that it is difficult to accurate estimate flow and density. To overcome the latter problem, Seo et al. [7] assume that the jam density is known.

Due to technological advances, an increasing number of vehicles is equipped with sensing equipment that can observe other road-users [8]. These vehicles can be used to collect new types of traffic sensing data. For instance, they can serve as moving observers that observe passings of other road-users with respect to their own position over time (trajectory) [8], [9]. Such new data may be used to estimate the FD.

The important characteristics of the proposed methodology are as follows. Our methodology does not require
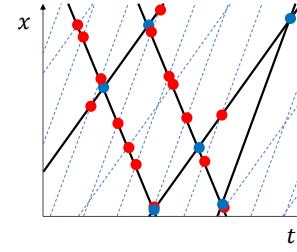
(a) Vehicle trajectories on a road stretch (blue dashed lines) and moving observers (black lines)
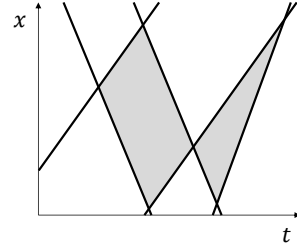
(b) Two space-time areas (grey) that are enclosed by moving observers

Fig. 1: Moving observers and relevant space-time areas.

prior knowledge or assumptions related to any parameters of the FD, e.g., [5], [6] and [7] respectively define a minimum value for the free-flow speed, the value for the wave speed and the value of jam density. Furthermore, we estimate q and k for areas in space-time based on Edie's definitions. Thereby we overcome the discussed downsides of stationary detector and probe data. However, as discussed above, some estimates, which include multiple traffic states, may not be representative for the FD. Therefore, in the proposed methodology, we use the expectation that traffic behaves according to a concave FD and thus that these estimates lie on or below the FD.

III. ESTIMATING FLOW AND DENSITY FOR SPACE-TIME AREAS USING MOVING OBSERVERS

In this study, we use relative flow data that are collected by moving observers. The path in space-time over which the moving observer travels is denoted as the observation path. We assume that each individual vehicle passing (and its direction) is observed.

Figure 1a shows the observation paths of five moving observers (black lines) and the observed individual passings (dots). In this example, respectively three and two moving observers travel in the driving and in the

2

opposite-driving direction of the observed traffic flow. There are multiple advantages of observing the opposite-driving direction. For instance, these moving observers report more vehicle passings. However, for this paper the most important advantage is that there are more intersections (indicated with the blue dots) of the observation paths, as will be discussed below.

A set of intersecting observation paths can form areas in space-time for which all boundaries are observed, i.e., all boundaries are part of an observation path. Figure 1b shows the areas for which this holds given the same set of observation paths that are shown in Figure 1a. Given that we observe all passings over the observation paths, we observe each individual vehicle that enters or leaves the space-time area.

Based on the individual passings and their direction (in- or outflow of the area), we can determine the TTD and TTS. To determine the TTD and TTS, the trajectories of the observers traveling in driving direction have to be taken into account. A weight of 0.5 is assigned to these trajectories as the boundaries separate two adjacent areas. Furthermore, the size of the space-time area, i.e., A, can be determined based on the spatial-temporal characteristics of the observation paths. This provides the required information to estimate q and k using Edie's definitions, i.e., equations (1) and (2), thereby yielding a $\{q, k\}$-estimate for each enclosed area.

IV. ESTIMATING THE FUNDAMENTAL DIAGRAM PARAMETERS

The methodology explained in this section can be used to fit a triangular FD to $\{q, k\}$-estimates obtained in Section III. A triangular FD is a simple two-variate FD that consists of two connected linear branches, which we refer to as the free-flow and congested branch. According to Seo et al. [1] the triangular FD is popular due to its simplicity, theoretically preferable features and some empirical evidence. Figure 2 schematically visualizes a triangular FD with the important parameters, i.e., free-flow speed v^f, wave speed ω, capacity q^C, critical density k^{cr}, jam density k_j and passing rate r. The triangular FD is described by the following function:

$$\begin{aligned} q &= v^f k && \text{if } k \leq k^{cr}, \\ q &= r - \omega k && \text{if } k > k^{cr}. \end{aligned} \qquad (3)$$

The free-flow and congested branch are connected at $\{k^{cr}, q^C\}$, i.e., $q^C = v^f k^{cr} = r - \omega k^{cr}$. Therefore, three parameters suffice to define the triangular fundamental diagram (e.g., v^f, ω and k^{cr}).
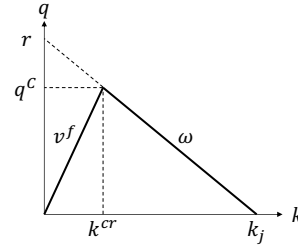

Fig. 2: Triangular fundamental diagram with important parameters.

To estimate the FD based on combined estimates of q and k, i.e., $\{q, k\}$-estimates, we need to know which $\{q, k\}$-estimates are representative for the free-flow branch and which are representative for the congested branch. As explained before, the $\{q, k\}$-estimates, that are obtained using Edie's definitions, will be the weighted average of the different traffic states in the area. Therefore, if traffic is deterministic and its behavior is correctly described by a triangular FD, the $\{q, k\}$-estimate either lie on or below the FD.

The critical density k^{cr} is the location of a structural break in describing q based on k. Estimates for which $k \leq k^{cr}$ can be representative for the free-flow branch, but not for the congested branch. Alternatively, estimates for which $k > k^{cr}$ be representative for the congested branch. However, in both cases, the observation might still include traffic states from both traffic phases, which are expected to lie below the line (branch) of the FD. If we known the location of the structural break, i.e., if we know k^{cr}, we can choose the remaining two parameters (defining the FD) such that *all* observations lie on or below the FD. To this end, we need an approach to find this unknown structural break.

Finding unknown structural breaks is studied extensively in the field of time-series analysis. The Quandt-Andrews breakpoint test [10], [11] is commonly applied approach to (1) find the location of the 'largest' structural break and (2) test if this structural break is significant. This test examines the different possible locations of the structural break and applies a Chow test [12] for each location. For each Chow-break-test, the F-statistic is calculated and the location with the maximum (supreme) F-statistic is selected as the structural break location. Depending on the F-statistic at this location and in the case that all assumptions related to the error distribution hold, we can say whether there is a significant structural break.

The proposed methodology is based on the Quandt-

3

Andrews breakpoint test; however, there are two important differences. Firstly, in contrast to time-periods (which are used as explanatory variables in time-series analysis), k is a continuous and will (most-likely) have varied intervals. This limits the accuracy in finding the correct k^{cr}, as the correct k^{cr} may lie in between two consecutive k^{cr} that are considered. However, by reducing the step-size in the k^{cr} that are evaluated, we can improve the accuracy in finding the structural break. Secondly, the assumptions related to the error distribution will not hold as all 'errors' will be negative (lie below the fitted line). The term 'errors' is used as there can still be differences between the fitted FD and $\{q, k\}$-estimate while the fitted FD is correct. Therefore, in the remainder of this article we refer to differences instead of errors. However, similar to an error we still minimize a difference statistic (see FD estimation sequence) as an erroneous FD induces larger differences. Dealing with differences instead of errors prevents us from testing whether the structural break is significant. However, this is not the purpose of the proposed methodology, as we always assume that there is a structural break in the FD.

Given the explanations and assumptions discussed above, the following FD estimation sequence is designed (note that steps 2b and 2d are a result of the expectation that all $\{q, k\}$-estimate either lie on or below the FD). In the estimation sequence, the full set of $\{q, k\}$-estimates is given by $\mathcal{O}$. To find the location of the structural break we minimize a difference statistic between the estimated FD and $\{q, k\}$-estimates. The choice for the sum of squared differences (see below) in not crucial. Other statistics, e.g., the mean absolute difference, may be chosen without having a large effect on the final result.

FD estimation sequence

1) Define a set of n to-be evaluated k^{cr}, i.e., $\mathbf{k}^{cr} = [k_1^{cr} \cdots k_n^{cr}]$
2) For each k_i^{cr}:
   a) Define subsets of free-flow observations, i.e., $\mathcal{O}_{\mathcal{F}} = \{o \in \mathcal{O} | k_o \leq k_i^{cr}\}$, and congested observations, i.e., $\mathcal{O}_{\mathcal{C}} = \{o \in \mathcal{O} | k_o > k_i^{cr}\}$.
   b) Find the free-flow speed v_i^f:

   $$v_i^f = \max_{o \in \mathcal{O}_{\mathcal{F}}} \left(\frac{q_o}{k_o} \right) \tag{4}$$

   c) Calculate capacity q_i^C:

   $$q_i^C = v_i^f k_i^{cr} \tag{5}$$

   d) Find the wave speed ω_i:

   $$\omega_i = \max_{o \in \mathcal{O}_{\mathcal{C}}} \left(\frac{q_o - q_i^C}{k_o - k_i^{cr}} \right) \tag{6}$$

   e) Calculate passing rate r_i:

   $$r_i = q_i^C - \omega_i k_i^{cr} \tag{7}$$

   f) Calculate the Sum of Squared Differences SSD_i:

   $$SSD_i = \sum_{o \in [k_o \leq k_i^{cr}]} \left(q_o - v_i^f k_o \right)^2$$
   $$+ \sum_{o \in [k_o > k_i^{cr}]} (q_o - (r_i + \omega_i k_o))^2 \tag{8}$$

3) Find the location of the structure break, i.e., find i for which $\min_i = SSD_i$, and select the related parameters as the estimate for the FD.

V. SIMULATION STUDY

A simulation study is conducted to evaluate the performance of the proposed FD estimation approach. This section contains the following elements: explanation of (1) the two applied (microscopic) car-following models and the simulation traffic conditions and data characteristics and (2) the evaluations performed to test and understand the methodology.

A. Microscopic simulation of traffic

Traffic is simulated using two different car following models: Newell's car following model [13] and the Intelligent Driver Model (IDM) [14]. The main difference between the models lies in the presence of transient phases. In Newell's model speed changes occur instantaneously and hence no transient phases are found. The IDM includes acceleration and deceleration towards a desired speed and hence includes transient phases. Table I contains the parameters that we use for simulating traffic using Newell's model and IDM.

Both models are deterministic, which means that the traffic flow simulated with both models results in a deterministic FD. However, the shape of the FD differs for the two models. When using Newell's model, the FD is triangular and is given by:

$$\begin{aligned} q &= v^f k & \text{if } k \leq k^{cr}, \\ q &= \frac{1}{\tau} - \frac{s_{jam}}{\tau} k & \text{if } k > k^{cr}. \end{aligned} \tag{9}$$

4

TABLE I: Parameters used for Newell's model and IDM

Model	Parameter	Value	Unit
Newell's	v^f	33.33	m/s
	s_{jam}	6.00	m
	τ	0.90	s
IDM[1]	v_0	33.33	m/s
	T	1.20	s
	δ	4.00	-
	a	0.80	m/s^2
	b	1.25	m/s^2
	s_0	1.00	m
	l	5.00	m

[1]) Default values from Treiber et al. [15]

When using IDM, the FD is a smooth continuous function. We can describe the spacing s as a function of speed u for equilibrium conditions, i.e., the acceleration is zero:

$$s = \sqrt{\frac{(s_0 + l + uT)^2}{1 - \left(\frac{u}{v_0}\right)^{\delta}}} \tag{10}$$

Based on this function we can plot the FD, where $k = 1/s$ and $q = u/s$.

In this simulation study, we want to estimate the FD for a road. For this purpose, traffic is simulated in two directions and data is collected from moving observers that are driving on this road and those that are driving in opposite direction. In both directions 250 veh are simulated over a period of 600 s with time-steps of 0.1 s. We only estimate the FD for one direction. Traffic in this direction is denoted as the 'observed traffic', while the opposing direction is denoted as the 'observing traffic'. The initial positions of the most upstream vehicle is respectively 0 m and 30,000 m for the observed and observing traffic. To determine the spacing between each combination of consecutive vehicles, we take a random draw from an exponential function with mean 10 m plus $s_{jam} + u\tau$ (for Newell's model) or $s_0 + l + uT$ (for IDM).

As we need to observe both free-flow as congested states in order to estimate the FD, a bottleneck is simulated in the observed traffic flow. In the simulation with Newells model, a bottleneck is simulated on the observed stream by reducing the speed of the first vehicle to 0 m/s for 30 s $\leq t < $ 150 s, to 5 m/s for 150 s $\leq t < $ 300 s, and increasing it back to v^f at $t = 300$ s until the end of the simulation. In the IDM simulation, the first vehicle decelerates with a m/s^2 to standstill starting at $t = 30$ s, and accelerates back to v^f with a m/s^2 starting at
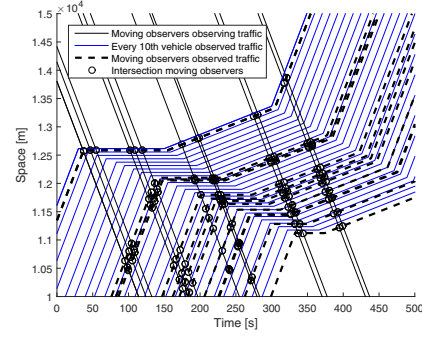
(a) Newell's model
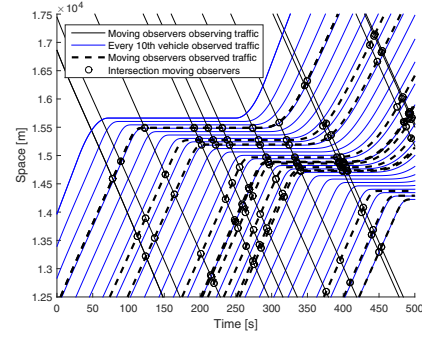
(b) IDM

Fig. 3: Simulated traffic flow and moving observers.

$t = 250$ s. Congested states are found upstream of this vehicle. No bottlenecks are included in the observing traffic flow, which thus has a constant speed of v^f.

It is assumed that the moving observers are able to observe all vehicles that they pass in opposite direction, i.e., the moving observers that are part of the 'observing traffic' observe the 'observed traffic'. For both directions 5 % of the vehicles is a moving observer. Following the principles explained in Section III, the trajectories of these vehicles are used to construct areas in space-time for which we can obtain $\{q, k\}$-estimates.

Figure 3 shows the trajectories of the observed traffic and the moving observers in space-time for Newell's model and IDM.
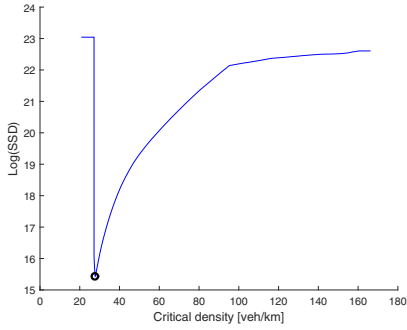
B. Evaluation of the proposed methodology

The objective of this simulation study is to evaluate the performance of the proposed methodology in correctly estimating the FD. Therefore, we compare the estimated FD with the true FD.

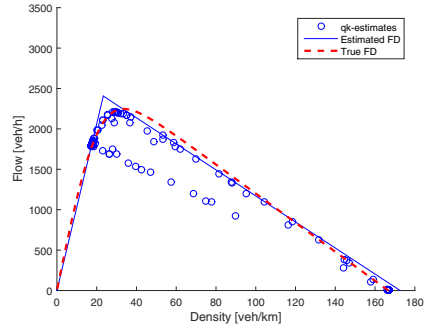For both traffic models (Newell's model and IDM) we show (1) the $\{q, k\}$-estimates, (2) the estimated
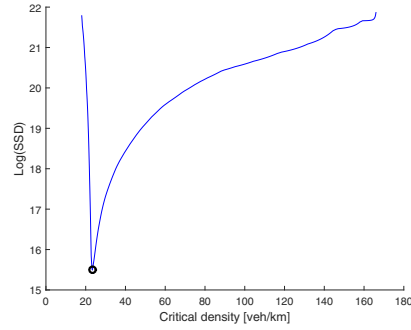
5
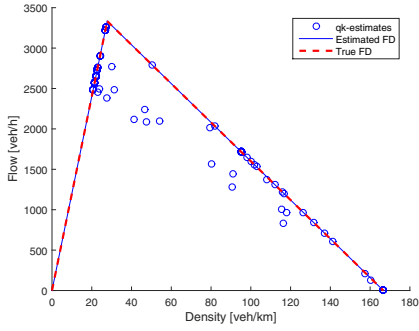
(a) Estimated FD

(b) SSD for different k^{cr}

Fig. 4: Estimated FD for Newell's model.

(a) Estimated FD

(b) SSD for different k^{cr}

Fig. 5: Estimated FD for IDM.

triangular FD and (3) the true FD. Furthermore, for each of the evaluated k^{cr} the Sum of Squared Differences, i.e., SSD is reported. Due to the extreme variation in SSD, we report the $\log(SSD)$. These plots allow us to evaluate the performance of the proposed methodology and discuss in more detail how it works.

VI. RESULTS

Figures 4 and 5 respectively show the estimated triangular FDs for traffic that is simulated using Newell's model and the Intelligent Driver Model (IDM). Furthermore, they depict the sum of squared differences between the estimated FD and $\{q, k\}$-estimates. For both car following models, we observe $\{q, k\}$-estimates that lie approximately on a line between a traffic state on the free-flow branch and the jam density traffic state, see Figures 4a and 5a. These estimates are weighted averages of the free-flow and jam density traffic states within the space-time area related to these estimates. Below, we separately discuss the accuracy of the estimated FDs for traffic simulated with both models.

The proposed methodology accurately estimates the FD that holds for the traffic simulated with Newell's model. The estimated (blue lines) and true (red dotted line) FD almost perfectly overlap in Figure 4a. In case a k^{cr} is considered that is below the true k^{cr}, there might be observations in the free flow branch which are now (incorrectly) assigned to the congested branch. This changes the direction of the congested branch, giving a positive wave speed ω. This explains the instant decrease in $\log(SSD)$ in Figure 4b between $k^{cr} = 27.77$ veh/km and $k^{cr} = 27.78$ veh/km.

The estimation of the IDM FD is not as accurate as for Newell's triangular FD. The proposed methodology makes two important assumptions, i.e., (1) the FD has a triangular shape and (2) all $\{q, k\}$-estimates lie on or below the FD. As shown in Figure 5a the true FD (red dotted line) does not have a triangular shape. However, given the restrictive shape, the estimated FD (blue line) still seems to be a good fit. The sum of squared differences SSD (Figure 5b) shows a less extreme change prior to the optimum k^{cr} than for the Newell's model-

6

based FD. This is a result of the smoothed top of the true IDM-FD. The combination of the two assumptions may lead to a larger difference between the true and estimated FD if other states on the FD are observed. For instance, at lower density k a larger speed u can be observed. If such a state is part of the set of $\{q, k\}$-estimates, the proposed methodology would estimate a larger v^f. Therefore, instead of the maximum observed u, we may want to consider all $\{q, k\}$-estimates that are relevant for the free-flow branch. In this case, the challenge lies in defining which estimates are relevant.

VII. Conclusions and outlook

This paper proposes an approach to estimate the fundamental diagram (FD) based on relative flow data collected with moving observers. This approach consists of two steps: (1) estimate flow and density based on the sensing data for areas in space-time using Edie's definitions and (2) estimate a triangular FD based on the flow and density estimates, and their theoretical relation to the FD.

The proposed approach works well. The flow and density estimates, i.e., $\{q, k\}$-estimates, provide valuable information to find the FD. With a simple algorithm, which assumes that the $\{q, k\}$-estimates lie on or below the FD, we were able to accurately expose the FD.

In reality the $\{q, k\}$-estimates can lie above the desired FD. For instance, if traffic behaves stochastic and we may observe states that lie above the desired mean FD. Furthermore, we may have to deal with observation errors, which yield $\{q, k\}$-estimates that do not perfectly describe the true traffic conditions. In this case, the approach proposed in this work can still be followed, i.e., obtain $\{q, k\}$-estimates using Edie's definitions and estimate the FD based on these $\{q, k\}$-estimates. However, one should alter the second step and estimate each branch based on the $\{q, k\}$-estimates relevant for that branch. This is more complex than selecting the maximum value for each branch and will therefore lead to a more extensive methodology.

References

[1] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura. Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*, 43:128–151, 2017.

[2] M. Papageorgiou, H. Hadj-salem, and J.M. Blosseville. ALINEA : A Local Feedback Control Law for On-Ramp Metering. *Transportation Research Record*, 1320:58–64, 1991.

[3] L. C. Edie. Discussion of traffic stream measurements and definitions. In *2nd Int. Symp. On the Theory of Traffic Flow*, Paris, 1965. OECD.

[4] M.J. Cassidy. Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B: Methodological*, 32(1):49–59, 1998.

[5] G. Dervisoglu, G. Gomes, J. Kwon, A. Muralidhuran, P. Varaiya, and R. Horowitz. Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity. In *Transportation Research Board 88th Annual Meeting*, pages 1–13, 2009.

[6] V. L. Knoop and W. Daamen. Automatic fitting procedure for the fundamental diagram. *Transportmetrica B: Transport Dynamics*, 5(2):129–144, 2017.

[7] T. Seo, T. Kusakabe, and Y. Asakura. Calibration of fundamental diagram using trajectories of probe vehicles: Basic formulation and heuristic algorithm. In *Transportation Research Procedia*, volume 21, pages 6–17, 2017.

[8] R. Florin and S. Olariu. On a Variant of the Mobile Observer Method. *IEEE Transactions on Intelligent Transportation Systems*, 18(2):441–449, 2017.

[9] P. B. C. Van Erp, V. L. Knoop, and S. P. Hoogendoorn. Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, 114:281–299, 2018.

[10] R. E. Quandt. Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association*, 55 (290):324–330, 1960.

[11] D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.

[12] G. C. Chow. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3):591–605, 1960.

[13] G.F. Newell. A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, 36:195–205, 2002.

[14] M. Treiber, A. Hannecke, and D. Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62 (2):1805–1824, 2000.

[15] M. Treiber, A. Hennecke, and D. Helbing. Microscopic Simulation of Congested Traffic. In *Traffic and Granular Flow '99*, pages 365–376, 2000.

7