Australasian Transport Research Forum 2022 Proceedings
28-30 September, Adelaide, Australia
Publication website: http://www.atrf.info

Estimating the Macroscopic Fundamental Diagram (MFD) for large-scale urban networks solely from probe vehicle trajectories

Elham Saffari[1], Mehmet Yildirimoglu[1], Mark Hickman[1]

[1] The University of Queensland, School of Civil Engineering, Brisbane, QLD 4076, Australia

Email for correspondence: m.yildirimoglu@uq.edu.au

1. Introduction

The MFD is a unimodal, low-scatter, and demand-insensitive relationship between network average flow and network average density (Geroliminis and Daganzo, 2008). MFD is used for traffic monitoring and management purposes such as, perimeter control (Ingole, Mariotte, & Leclercq, 2020), regional route guidance (Yildirimoglu, Sirmatel, & Geroliminis, 2018), demand management (Yildirimoglu & Ramezani, 2020) and control of city-scale ride-sourcing systems (Ramezani & Nourinejad, 2018). The majority of previous studies have used loop detector data and/or probe vehicle data (empirical or simulation) to estimate the MFD. Buisson and Ladier (2009) investigated the effect of loop detectors' distance from the traffic signals on density measurements. The authors concluded that this distance has a strong impact on the slope of the MFD. This was later confirmed by Courbon and Leclercq (2011) (simulation) and Ambühl et al. (2017) (empirical). In addition to the importance of loop detectors' position on a link, selection of links within the network to install loop detectors is crucial (Zockaie et al., 2018; Saffari et al., 2020). Probe vehicles on the other hand, due to their dynamic nature, seem to rectify the limitations of loop detectors. Given that the traffic information provided by probe vehicles is not complete, network average flow and density calculated using their observations will not be representative of the entire traffic stream. However, if the probe penetration rate (i.e., the portion of probe vehicles) is known, the partial observations from probe vehicles can be up-scaled to represent the full traffic conditions (Nagle and Gayah, 2013; Du et al., 2016; Paipuri et al., 2020).

The majority of the relevant literature in estimating the MFD using only probe vehicle trajectories proposed methods depended on loop detectors for estimating the penetration rate of probe vehicles. However, this method fails when loop detectors are not available or cannot provide reliable observations. The aim of this work is to estimate the MFDs based solely on probe vehicle data with an unknown penetration rate. To the best of our knowledge, this is the first study that only employs probe vehicle trajectories to estimate the MFDs for a large-scale urban network. Unlike the previous studies, our proposed framework does not rely on loop detectors to estimate the penetration rates. Clearly, the number of probe vehicles is always known given the probe vehicle data; however, finding the total number of vehicles based on only this data is very challenging. In urban areas, most intersections are signalised, where vehicles need to stop at the red lights. This provides us with the stopping positions of probe vehicles in the formed queues caused by the red light. Based on the position of the last probe vehicle in the queue, we can estimate the number of vehicles ahead of the last probe vehicle. With this information, the total number of vehicles on a link can be approximated and subsequently the penetration rate can be estimated. In other words, the observable queue length could be employed as a proxy of the total number of vehicles on a link. Moreover, probe penetration rate is a spatiotemporal variable; that is, its variability in space and time cannot be

neglected. Most previous studies assumed a constant penetration rate either in space or in time, which might result in inaccurate MFDs. Here, we account for spatial variability of the penetration rate by introducing neighbourhood penetration rates. Additionally, our framework estimates real-time penetration rates which relaxes the assumption of constant penetration rate in time.

2. Methodology

In this section, we explain the proposed methodology for estimating the MFDs based on probe vehicle trajectories considering an unknown penetration rate. Figure 1 presents an overview of the proposed framework. The first step is to estimate link-level penetration rates using the position and speed of probe vehicles. As explained before, based on the position of the last stopped probe vehicle, and by considering a constant average space headway in stopped traffic, the number of vehicles ahead of the last probe vehicle can be estimated for each cycle. In this paper, we use the estimator that was developed by Zhao et al. (2019) (Eq. 1), and build on this formulation and expand it for our study network. Based on this formulation the total number of observed vehicles on link *i* in a given cycle can be estimated as:

$$\widehat{N}^i = \left(x_p^i \frac{n_p^i + 1}{n_p^i} - 1\right) \quad (1)$$

where $\widehat{N}^i$ and n_p^i are the number of estimated vehicles and probe vehicles in the queue on link *i*, respectively. x_p^i denotes the position of the last probe vehicle in the queue on link *i* (e.g., 1st, 2nd, or 3rd vehicle in the queue). The penetration rate of link *i* then can be calculated as $\rho_i = \frac{n_p^i}{\widehat{N}^i}$. Note that vehicles that cross the intersection while the signal is green cannot be considered when calculating link penetration rates. Additionally, we do not apply the above queue-based estimation approach for the unsignalised intersections and roundabouts; their penetration rates will be estimated via a neighborhood approach (or a local averaging approach).

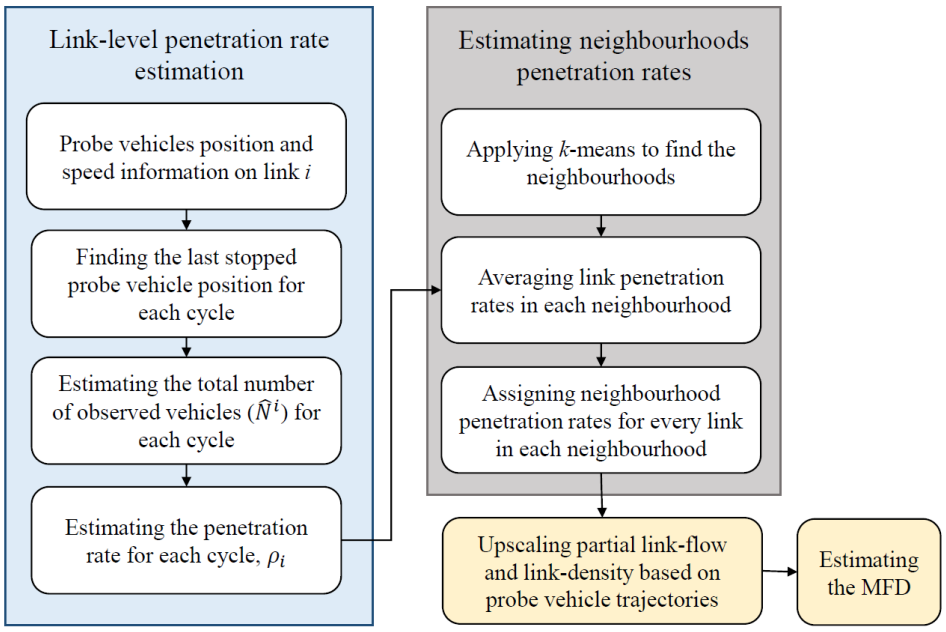

Figure 1: Flowchart of the proposed methodology

The MFD is derived for every time interval, *t* (every 60 seconds in this study), and partial observations collected from probe vehicles should be upscaled by applying the penetration rate

in every time interval. By doing so (i.e., calculating the real-time penetration rates), we account for the temporal variability of this variable. Before explaining the next step (neighbourhood penetration rates estimation), the difference between cycle lengths and time intervals should be clarified. Figure 2 depicts an example of this mismatch where a cycle length is longer than the time interval. Note that this scenario is mostly the case in the present study. As mentioned before, Eq. 1 is applied for every cycle which is not necessarily as long as the time interval that we estimate the MFD. As shown in Figure 2, the first cycle is longer than the first time interval, t_1, and it extends to a portion of the second time interval, t_2. To resolve this inconsistency between the cycle lengths and time intervals, for a given link and a given time interval, we consider the estimated penetration rate within the last cycle that ended before the given time interval. For example, in order to estimate the MFD at time A (shown in Figure 2), probe vehicle observations in time interval t_3 should be accompanied (or upscaled) with link-level penetration rates calculated in cycle 2. Note that with this method, we are not able to produce estimations for the first time interval, as the cycle length is longer than the time interval.

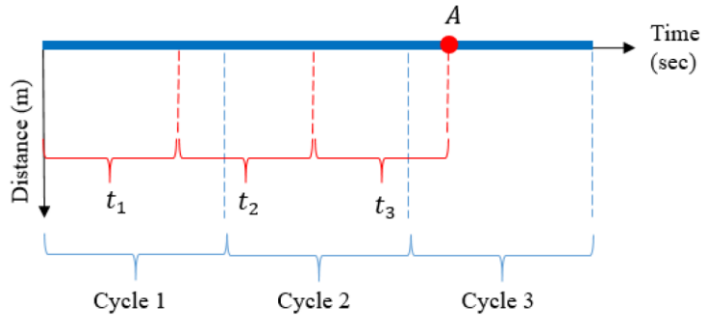

Figure 2: Cycle-based vs period-based estimations

The next step is to define neighbourhood penetration rates. This step is needed mainly because the individual link penetration rates that we estimate with the above approach may contain significant noise considering the random nature of the collected probe vehicle data. A spatial averaging procedure like the one that we propose here helps reduce the noise in the estimated penetration rates. Additionally, there are a number of unsignalized intersections in the network which need to be taken into account. For the links that are connected to these intersections, the queue-based estimation approach does not provide the penetration rates; in other words, Eq. 1 is not applicable. Furthermore, for links that are connected to signalized intersections, we only consider the stopped vehicles; however, during the green time, other vehicles might have travelled through these links. This suggests that the same vehicles may have to stop at red lights on nearby links. We can assume that during a short period of time the same vehicles are travelling in small neighbourhoods. The neighbourhood penetration rate can be calculated by averaging the penetration rate of the links within each neighbourhood.

In order to find the neighbourhoods, we apply the *k*-means algorithm, which is one of the simplest and most popular unsupervised clustering methods. Here, we aim to find groups of links which are in close vicinity (neighbourhoods). Then, we need to define the value of *k* which is rather challenging when applying the *k* -means algorithm. Here, the purpose of neighbourhoods is to define a small area where we can assume the same vehicles are travelling through in a 60-seconds time interval. Given the average speed, we can calculate the average distance travelled by vehicles. Then, considering average link length in the network, we can approximate the number of links that it travels through in every time interval. By assuming grid-like neighbourhoods, the number of links that forms each neighbourhood can then be estimated. Given the total number of links in the network, the number of neighbourhoods *k* is calculated. Note that this clustering-based approach can easily be replaced with any other rule-

based or shape-based approach (e.g., grid partitions); this is not the core of the proposed framework.

Let I_S be the number of links connected to signalized intersections within the neighbourhood *S*, $\forall S \in \{1, 2, \dots, k\}$. The penetration rate of neighbourhood *S* in time *t* can then be calculated as,

$$\rho_S(t) = \frac{\sum_{i \in I_S} \rho_i(t)}{I_S} \quad (2)$$

Note that in this equation, $\rho_i(t)$ denotes the penetration rate of link *i* calculated in the last cycle before time *t*.

Given the neighbourhood penetration rates, we are able to upscale probe vehicle observations to subsequently estimate the MFDs. To calculate link-flow and link-density from probe vehicle trajectories, Edie's formulations (Edie, 1963) can be applied. For each neighbourhood, we calculate link-flow and link-density at each time interval using the following formulas,

$$q_i(t) = \frac{\sum_p d_{ip}(t)}{n_i l_i t \rho_S(t)} \quad k_i(t) = \frac{\sum_p t_{ip}(t)}{n_i l_i t \rho_S(t)} \quad (3)$$

where q_i and k_i denote flow and density of link $i \in I_S$ in a time interval *t*, respectively. d_{ip} and t_{ip} are the distance traveled and time spent by vehicle *p* on link *i* with the length l_i and number of lanes n_i.

Once the link level flow and density values are calculated, the corresponding values in the MFD can be estimated as,

$$Q(t) = \frac{\sum q_i(t) l_i}{\sum l_i} \qquad K(t) = \frac{\sum k_i(t)\, l_i}{\sum l_i} \quad (3)$$

where $\boldsymbol{Q(t)}$ and $\boldsymbol{K(t)}$ denote average network flow and density in time interval *t*, respectively.

3. Results and Discussions

This section presents the numerical results evaluating the accuracy of the estimated the MFDs based on variable probe vehicle samples sizes (e.g., 3%, 5%, 10%, 15% and 25%). We compare the estimated MFDs with the ground-truth MFD to evaluate the accuracy of the estimated MFDs and quantify how close our estimations are to the ground-truth MFD. To derive the ground-truth MFD, all vehicle trajectories are employed. Root Mean Square Error (RMSE) for network average flow and network average density is calculated for the simulation duration *T* as follows:

$$\text{RMSE(Q)} = \sqrt{\frac{\sum_{t=1}^{T} (\hat{Q}(t) - Q(t))^2}{T}} \qquad \text{RMSE(K)} = \sqrt{\frac{\sum_{t=1}^{T} (\hat{K}(t) - K(t))^2}{T}} \quad (4)$$

where $\hat{Q}(t)$ and $\hat{K}(t)$ denote the ground-truth average flow and density in time interval *t*, respectively.

Figure 3 illustrates the estimated MFDs with respect to varying probe vehicle sample sizes. As shown in the figure, when the sample size is as small as 3% (only 3% of vehicles in the network can provide their trajectories), network average flow and average density values are underestimated. This is in fact because when there are less probe vehicles in the network, the probability of the last stopped vehicle being a probe vehicle is small. Hence, there will be more vehicles behind the last stopped probe which we are not able to detect, despite the use of Eq. 1 in estimation of queue lengths. This consequently leads to underestimating the number of total observed stopped vehicles, and therefore, overestimating the penetration rates. When applying Eq. 3, using a higher penetration rate than the true value of this parameter produces lower

values for link-flow and density and ultimately results in MFD underestimation. However, with increasing sample size, it is more likely that the last stopped vehicle would be a probe vehicle. This enables us to estimate the penetration rates more accurately. We can clearly see in Figure 3 that the estimated MFDs are getting closer to the ground-truth MFD when increasing the sample size. The numerical results in the form of RMSEs are presented in Table 1 for the sake of comparison.

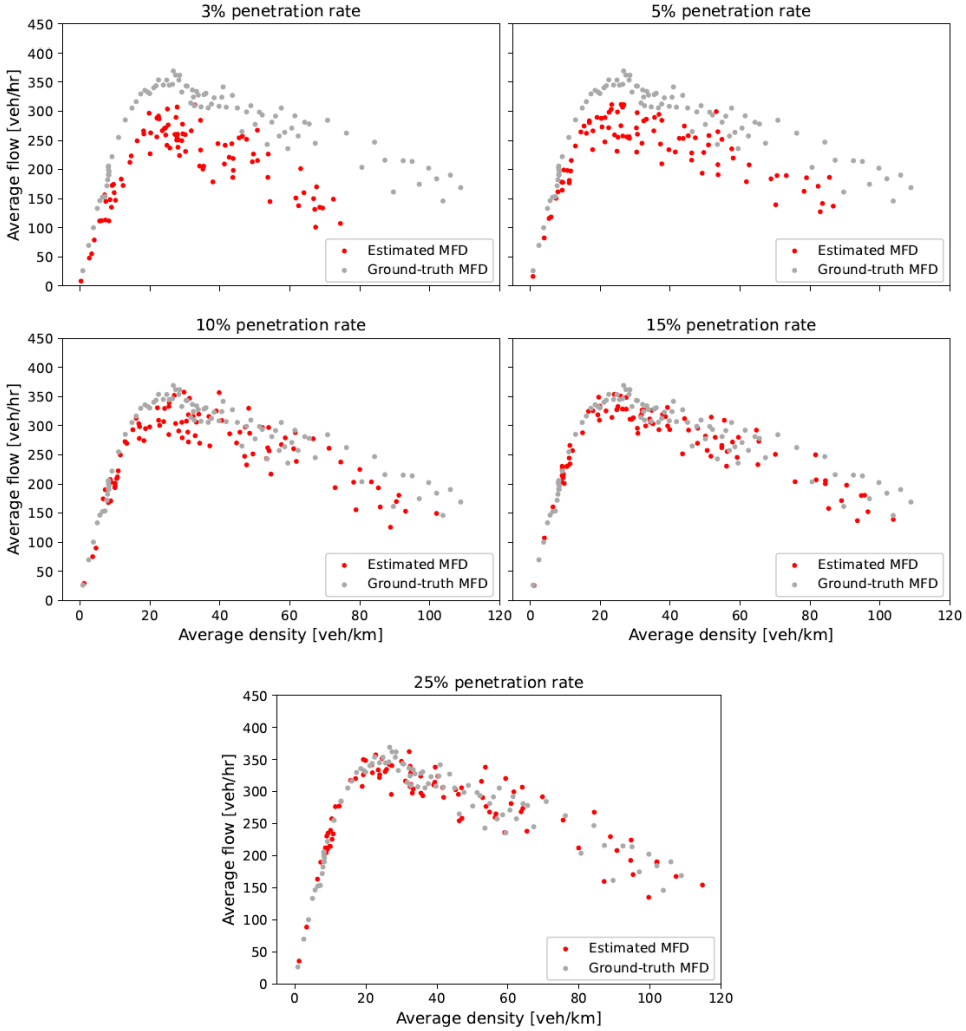

Figure 3: Estimated MFDs based on variable probe vehicle sample size

Table 1: Average flow and density RMSEs based on different probe vehicle sample sizes

	Neighbourhood method		Network penetration rate	
	RMSE(Q) [veh/km]	RMSE(K) [veh/km]	RMSE(Q) [veh/km]	RMSE(K) [veh/km]
3%	67.64	13.95	84.86	16.70
5%	44.73	10.12	68.55	12.79
10%	27.97	5.33	42.08	6.85
15%	23.62	3.93	28.34	4.68
25%	22.95	1.99	26.24	2.17

We also evaluate the estimated MFDs based on a single network penetration rate which is calculated by averaging the penetration rates of all links in the network. Doing so, we indeed dismiss the spatial variability of the penetration rate. The last two columns of Table 1 present the average flow and average density RMSEs applying the network penetration rate.

Comparing the two methods, we clearly see that the estimation errors are lower when considering the neighbourhood penetration rates to estimate the MFDs even though probe vehicles are homogeneously distributed in this scenario. This is particularly more significant in smaller probe vehicles samples. The reason perhaps is projecting a more biased penetration rate, which is estimated based on a small information provided by probe vehicles, on the entire network.

4. References

Ambühl, L., Loder, A., Menendez, M., and Axhausen, K. W. 2017. Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data. In TRB 96th Annual Meeting Compendium of Papers, pages 17–03331. Transportation Research Board.

Buisson, C. & Ladier, C. 2009. Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. Transportation Research Record: Journal of the Transportation Research Board, 2124, 127-136.

Courbon, T. & Leclercq, L. 2011. Cross-comparison of Macroscopic Fundamental Diagram Estimation Methods. Procedia - Social and Behavioral Sciences, 20, 417-426.

Du, J., Rakha, H., and Gayah, V. V. 2016. Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. Transportation Research Part C: Emerging Technologies, 66:136–149.

Edie, L. C. (1963). Discussion of traffic stream measurements and definitions. Port of New York Authority.

Geroliminis, N. & Daganzo, C. F. 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological, 42, 759-770.

Ingole, D., Mariotte, G., & Leclercq, L. (2020). Perimeter gating control and citywide dynamic user equilibrium: A macroscopic modeling framework. Transportation research part C: emerging technologies, 111, 22-49.

Nagle, A. S. and Gayah, V. V. 2013. A method to estimate the macroscopic fundamental diagram using limited mobile probe data. In 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), pages 1987–1992. IEEE.

Paipuri, M., Xu, Y., González, M. C., and Leclercq, L. 2020. Estimating mfds, trip lengths and path flow distributions in a multi-region setting using mobile phone data. Transportation Research Part C: Emerging Technologies, 118:102709.

Saffari, E., Yildirimoglu, M. & Hickman, M. 2020. A methodology for identifying critical links and estimating macroscopic fundamental diagram in large-scale urban networks. Transportation Research Part C: Emerging Technologies, 119.

Ramezani, M., & Nourinejad, M. (2018). Dynamic modeling and control of taxi services in large scale urban networks: A macroscopic approach. Transportation research part C: emerging technologies, 94, 203-219.

Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., and Liu, H. X. 2019. Various methods for queue length and traffic volume estimation using probe vehicle trajectories. Transportation Research Part C: Emerging Technologies, 107:70–91.

Zockaie, A., Saberi, M., and Saedi, R. (2018). A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. Transportation Research Part C: Emerging Technologies, 86:245–262.

Yildirimoglu, M., & Ramezani, M. (2020). Demand management with limited cooperation among travellers: A doubly dynamic approach. Transportation Research Part B: Methodological, 132, 267-284.

Yildirimoglu, M., Sirmatel, I. I., & Geroliminis, N. (2018). Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. Transportation Research Part B: Methodological, 118, 106-123