Australasian Transport Research Forum 2021 Proceedings
8-10 December, Brisbane, Australia
Publication website: http://www.atrf.info

Combining multiple traffic data sources to estimate Macroscopic Fundamental Diagram in large-scale urban networks

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

Since the concept of the Macroscopic Fundamental Diagram (MDF) has been introduced by (Geroliminis and Daganzo, 2008), many studies have investigated the existence and characteristics of the MFD using empirical and simulation data. MFD is a powerful and efficient model for monitoring and managing large-scale urban networks. For instance, 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). Nevertheless, estimating the MFD for large-scale networks faces important challenges; monitoring resources are often limited in such networks. Furthermore, common sensors that are used to collect traffic data (i.e., loop detectors and probe vehicles), have limitations of their own. For instance, loop detectors are fixed sensors and cannot provide accurate density measurements (Buisson and Ladier, 2009, Courbon and Leclercq, 2011). On the other hand, to estimate the MFD using probe vehicle data, the probe penetration rate must be known a priori. Given that the individual sensors cannot provide complete and accurate traffic measurements, combining the traffic data from multiple sources may improve the estimation of the MFD (Ambühl and Menendez, 2016, Beibei et al., 2016, Ji et al., 2018, Leclercq et al., 2014).

Our aim in this study is to develop a data fusion method that takes advantage of both (limited number of) loop detectors and probe vehicles, which may or may not be homogeneously distributed in the network. This study builds on the premise that full-scale traffic data (i.e., covering all links in the network), albeit approximate, is available for the network, which is produced as a result of our earlier work (Saffari et al., 2020). Very briefly, the previous study identifies a small number of critical links in the network where loop detectors should be installed, and produces an approximation of complete traffic variables (i.e., flow and density) for all links. In this study, in addition to loop detector measurements from the critical links, we assume that real-time probe vehicle data with an unknown penetration rate is available. These two data sets are the inputs to our fusion algorithm.

2. Methodology

In this section, we present the proposed methodology to combine real-time probe vehicle data, with an unknown probe penetration rate, and approximate full-scale traffic data (resulted from our earlier work). Applying the proposed fusion method, we can calculate link-flow and link-density for all the links in the network. Figure 1 presents the main steps of the proposed methodology. There are two main parts shown in the flowchart; 1) calculating the local penetration rates, and 2) applying the Bayesian fusion method. In the first part, we start with probe vehicle and loop detector observations on the critical links; this allows us to calculate the penetration rate for each critical link where a loop detector is installed. We then find the *k*-

nearest critical links for each link in the network, and calculate the average penetration rate of these k links. This allows us to estimate a local penetration rate for each link, which may vary across the network. In the second part of the algorithm, we upscale probe vehicle observations, applying the estimated local penetration rates. This data is one of the inputs to the Bayesian fusion model. As shown in Figure 1, two traffic sources (i.e., approximate full-scale traffic data and upscaled probe vehicle measurements) are combined applying the proposed Bayesian data fusion model. The output of the model is fused link-flow and link-density values which we later use to estimate the MFD for the network.

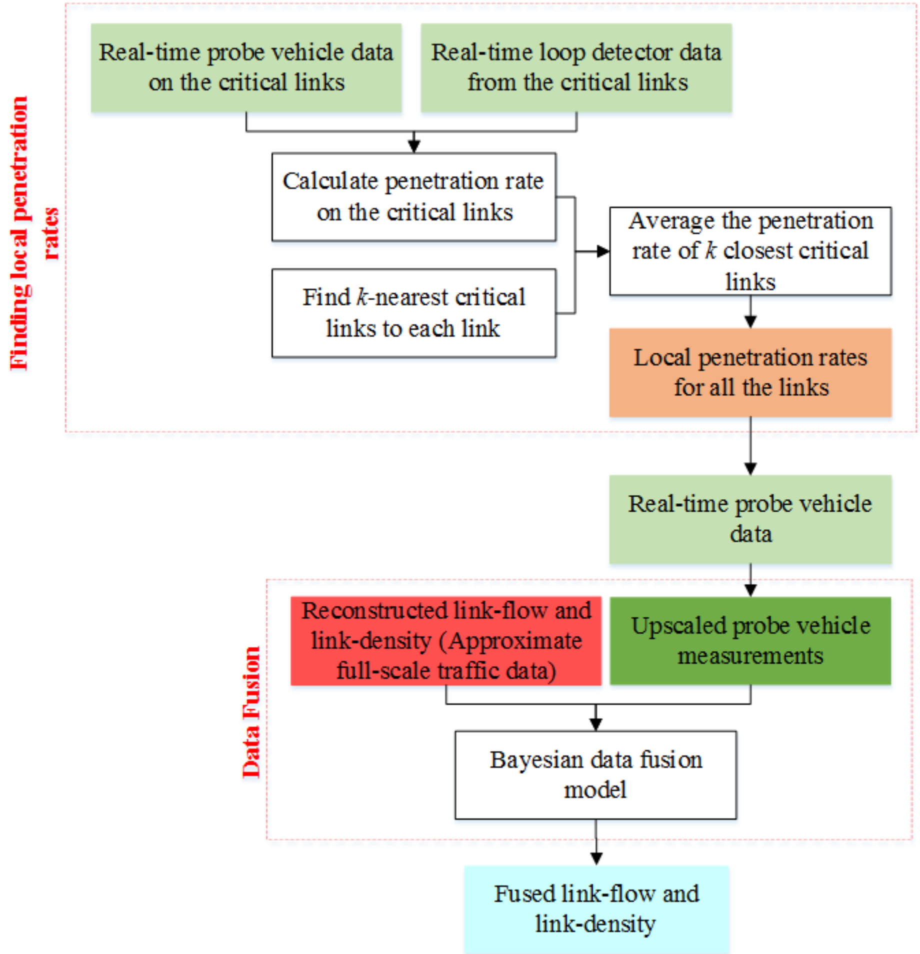

Figure 1: Flowchart of the proposed methodology

The data fusion method that we adopt in this study is based on Bayesian inference. As mentioned, in our problem, there are two sets of data, real-time probe vehicle and approximate full-scale traffic data, that we aim to combine by applying a Bayesian data fusion model. Let $q_a^{i,t}, q_p^{i,t}, k_a^{i,t}$ and $k_p^{i,t}$ denote approximate flow, flow based on probe vehicles, approximate density and density based on probe vehicles of link i in time of day t, respectively. n_a and n_p are the number of approximate traffic data observations and the number of probe vehicle observations, respectively. Let us assume M days of traffic measurements are available (i.e., M replications simulated in Aimsun). Thus, the number of observations on each link in each time of day interval can vary between zero and $M (0 \leq n_a, n_p \leq M)$. Based on Bayes theorem:

$$P\left(\mu_q^{i,t} \middle| q_a^{i,t}, q_p^{i,t}\right) = \frac{P(\mu_q^{i,t}, q_a^{i,t}, q_p^{i,t})}{P(q_a^{i,t}, q_p^{i,t})} \qquad P\left(\mu_k^{i,t} \middle| k_a^{i,t}, k_p^{i,t}\right) = \frac{P(\mu_k^{i,t}, k_a^{i,t}, k_p^{i,t})}{P(k_a^{i,t}, k_p^{i,t})} \tag{1}$$

where $\mu_q^{i,t}$ and $\mu_k^{i,t}$ denote the value of the fused flow and fused density of link *i* in time *t*, respectively. We omit *i* and *t* in the notation in the following equations for the sake of brevity.

Note that to avoid duplication and due to space limitation, we write the final equation only for calculating fused link-flow. The same formula will be applied for density to calculate the fused density values for the network links. Here, we assume that μ_q, q_a, q_pfollow normal distributions, $N(\mu_0, \sigma_0^2)$, $N(\mu_q, \sigma_a^2)$ and $N(\mu_q, \sigma_p^2)$, respectively. μ_0 and σ_0^2 denote the mean and variance of the prior distribution, respectively. We can find the mean of fused link-flow (μ_f) as:

$$\mu_f = \frac{1}{\frac{n_a}{{\sigma_a}^2} + \frac{n_p}{{\sigma_p}^2} + \frac{1}{{\sigma_0}^2}} \left(\frac{\sum_{r=1}^{n_a} q_a^r}{{\sigma_a}^2} + \frac{\sum_{s=1}^{n_p} q_p^s}{{\sigma_p}^2} + \frac{\mu_0}{{\sigma_0}^2} \right) \tag{2}$$

As mentioned earlier, we omit the link *i* and time *t* from the equations for simplification purposes. In other words, by applying Eq. 2, we calculate the fused flow value on link *i* in time interval *t*. Therefore, to find link flow values for all links in every time interval, Eq. 2 needs to be applied $N \times T$ times, where N is the total number of links in the network and T is the total number of time intervals. Once link-flow and link-density values are calculated, we can find the MFD parameters of network average flow $Q(t)$, and network average density, $K(t)$, using the following formulas:

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

where $q_i(t)$ and $k_i(t)$ are fused link-flow and density measurements from link *i* in time interval *t*, respectively. The length of each link is denoted by l_i.

3. Bayesian data fusion results

In this section, we investigate the performance of the proposed fusion algorithm based on a heterogeneous probe vehicle distribution and different subsets of critical links. The network of the study is a large-scale urban network of Eixample district in Barcelona, Spain, which is modelled in Aimsun, a well-known traffic simulation package. In order to generate probe vehicles and extract their trajectories, we use Aimsun API (Application Programming Interface) in a micro-simulation environment. The simulation period represents a 90-minute morning peak time. We consider seven replications of the explained micro-simulation model, representing '7 days' (i.e. the maximum number of observations is $\boldsymbol{M = 7}$). Note that the number of critical links shows the number of links with loop detectors placed on them. We explore subsets of 20, 40, 60 and 80 links that represent approximately 2%, 3%, 5% and 7% of the links, respectively. For each subset of critical links, we first apply *k*-NN to find the three nearest critical links for all the links in the network. Then, the penetration rate on each link is calculated by averaging the penetration rate of the three nearest critical links. Incorporating the penetration rates, we can upscale partial probe vehicle observations to complete traffic measurements (link-flow and link-density). Note that this is obviously an approximation of the complete traffic measurements. The upscaled real-time probe vehicle data set is one of the two inputs to the Bayesian fusion algorithm. The second input, as explained before, is the approximate full-scale traffic data. The next step is to apply the Bayesian data fusion method (Eq. 2) and combine the two aforementioned data sets and calculate the fused link-flow and link-density values.

The next step after finding the link level estimations is to use Eq. 3 to find the network average flow and network average density and estimate the MFDs using this fusion method. Figure 2 illustrates the estimated MFDs with respect to different subsets of critical links along with the ground-truth MFD which is calculated from the trajectories of all vehicles in the network. One expected observation is that having more critical links, which subsequently means having more loop detectors in the network, results in a better fit and less scatter. Having more loop detectors spread out in the network leads to better local penetration rate estimations which essentially improves the MFD estimations.

To evaluate our estimations, we compare the estimated network average flow and network average density with the ground-truth values by applying Eq. 4.

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

where $\hat{Q}(t)$ and $Q(t)$ stand for the ground-truth and estimated average network flow in time interval t, respectively; and $\widehat{K}(t)$ and $K(t)$ are the ground-truth and estimated average network density in time interval t, respectively.

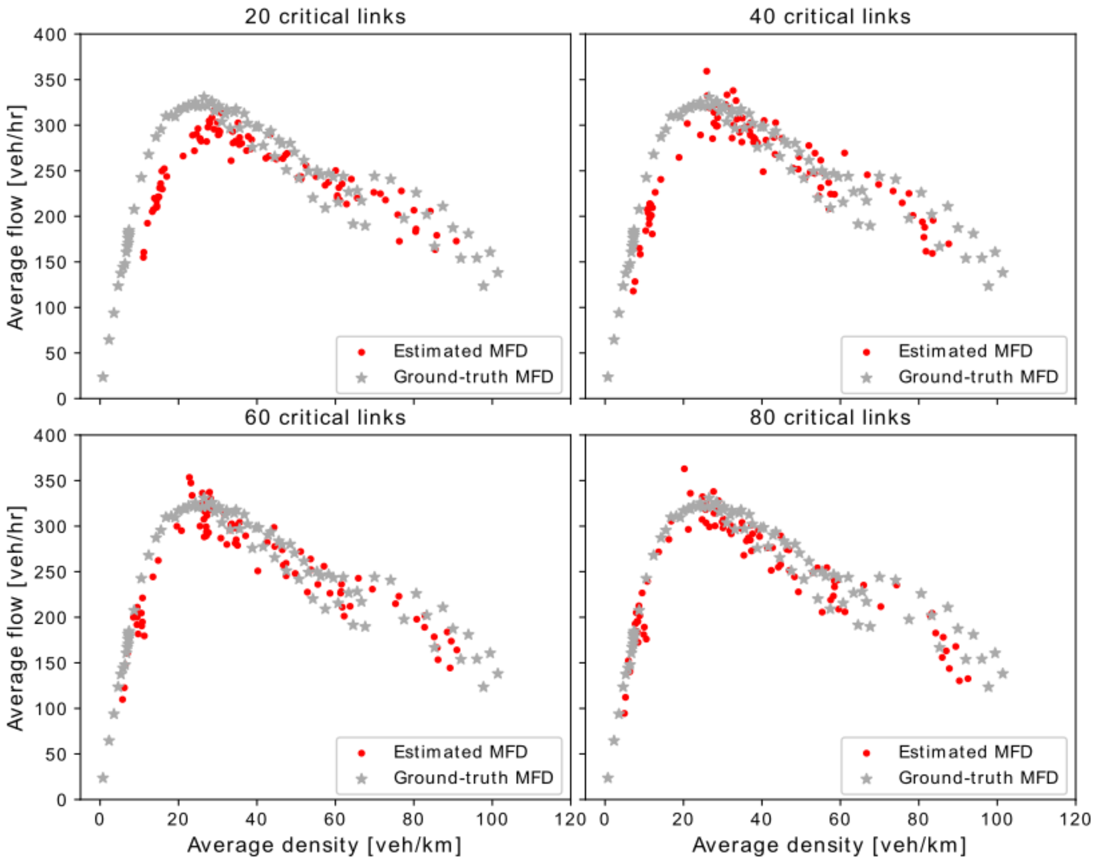

Figure 2: Estimated MFDs with respect to different subsets of critical links

The results of this calculation are presented in Table 1. This table also compares the estimation errors of the Bayesian fusion method and the baseline method which is applied in our earlier study. As we can see in Table 1, the Bayesian data fusion method improves the average flow and the average density estimations in most of the scenarios. Although we do not see a significant difference between RMSE(Q) from the fusion and the baseline method, we clearly observe a great improvement in all RMSE(K) values when applying the fusion model. The improvement in RMSE(K) ranges from 23% to 46%, whereas the improvement in RMSE(Q) is at most 7%. Note that, loop detector observations form the basis of the approximate full-scale traffic data. While the resulting flow estimations are fairly accurate and up to par with the fusion algorithm, the density estimations are significantly worse. These results can confirm the fact that loop detectors cannot provide accurate density measurements in congested signalized traffic sections, while they provide reasonably accurate flow measurements. Therefore, incorporating probe vehicle observations using the proposed fusion algorithm can significantly decrease the bias in loop detector density measurements and result in more accurate density estimations. Additionally, one possible reason that we do not see a considerable improvement in flow estimations, when incorporating probe vehicle observations, is that loop detector measurements do not significantly differ from the probe vehicle measurements. In other words, adding probe vehicle measurements to the loop detector measurements may not provide more information about the traffic state on the links.

Table 1: Estimation error with respect to different subsets of critical links

	Bayesian data fusion method		**Baseline method**		**Percentage improvement (%)**	
#Cririal links	RMSE(Q) [veh/km]	RMSE(K) [veh/km]	RMSE(Q) [veh/km]	RMSE(K) [veh/km]	RMSE(Q) [veh/km]	RMSE(K) [veh/km]
20	35.86	6.65	37.00	12.30	3	46
40	25.30	5.58	27.19	7.27	7	23
60	22.25	4.52	22.66	7.23	2	37
80	18.68	4.41	18.75	5.96	0	26

4. References

AMBÜHL, L. & MENENDEZ, M. 2016. Data fusion algorithm for macroscopic fundamental diagram estimation. Transportation Research Part C: Emerging Technologies, 71, 184-197.

BEIBEI, J., VAN ZUYLEN, H. & SHOUFENG, L. Determining the macroscopic fundamental diagram on the basis of mixed and incomplete traffic data. TRB 95th Annual Meeting Compendium of Papers, 2016.

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.

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

JI, Y., XU, M., LI, J., VAN ZUYLEN, H. J. J. P.-T. & TRANSPORTATION 2018. Determining the macroscopic fundamental diagram from mixed and partial traffic data. 30, 267-279.

LECLERCQ, L., CHIABAUT, N. & TRINQUIER, B. 2014. Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. Transportation Research Part B: Methodological, 62, 1-12.

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.

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.

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.

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.