Assessing the Accuracy of Link Accumulation Estimation

Anna TAKAYASU[1], Ludovic LECLERCQ[1] and Nikolas GEROLIMINIS[2]

Keywords: Traffic State Estimation, Streaming Data-Driven Estimation, Accuracy Assessment

1. Introduction

In this study, we assess the accuracy of link accumulation estimation by using real-time traffic data, e.g. loop detector data and probe vehicle data. We consider the technical and scaling issues including penetration of each traffic data.

Accurate traffic state estimation is crucial for network monitoring and management, e.g. signal control, perimeter control and providing traffic information. Therefore, over the past decades, the methods to improve accuracy and/or to save computational time have attracted a great deal of attention. Seo, et al. [1] shows three categories of traffic state estimation methods: *model driven approach*, i.e. based on physical traffic model, *data driven approach*, i.e. relying on historical data to estimate simple model, and *streaming data driven approach*, i.e. requiring only real-time data combined with simple physical. Especially, the calculation of *streaming data driven approach* is very fast and quite robust for uncertain phenomenon and unpredictable incidents compared to other approaches. On the other hand, this approach may require a large amount of real-time data and the accuracy may not be as high as the other two approaches. It is important for validity assessment to qualify the required data quantity and estimate accuracy of outputs. However, certain data requirements to estimate accurate traffic state are not quantified. Therefore, in this paper, we focus on the accuracy assessment of traffic state estimation by streaming a data driven approach.

When working on traffic state estimation with real-time data, it is important to design an adequate method. Velocity and accumulation are fundamental variables. It is already known that velocity can be estimated quite accurately by probe vehicle data. On the other hand, traffic accumulation estimation has several technical and scaling issues. There are two levels of traffic accumulation depending on the space coverage: local and link level. By using loop detector data, local level accumulation is usually derived from occupancy, i.e. the percent time that the detector is occupied by vehicles, and the vehicle length. In many researches, the vehicle length is assumed as a constant value at every road links or during a whole day as direct estimation of vehicle length from loop detector data is not possible. However several researchers mention that this assumption may lead a huge error when estimating traffic density (Zefreh, et al.[2], Leclercq, et al.[3], Kockelman[4]). Link level accumulation is important to estimate traffic state of a whole

[1] Université Gustave Eiffel, Université Lyon, ENTPE, Lyon, France
[2] EPFL, Lausanne, Switzerland

link or/and a network. To derive accumulation at the link level by using loop detector data, local accumulation estimation should be scaled-up. Leclercq[5] and Knoop, et al.[6] clarified theoretically and empirically that traffic accumulation estimation inside loop detection area is not adequate to estimate accumulation over the whole link. On the other hand, by using probe vehicle data, scaling-up from local to link level is not necessary to estimate traffic accumulation. However, it exhibits some scaling issues as the relation between the total accumulation and the probe vehicle accumulation is given by the penetration rate, and there is strong variance in this estimator for low values of penetration rate. This factor is hard to estimate in general and can experience spatial and/or temporal variation. According to several issues mentioned above, the accuracy of traffic accumulation estimation in streaming data driven approach should be investigated in a more comprehensive manner.

Our contribution is 1) providing clear confidence intervals for each estimation method and 2) proposing the best balance of data quantity for accurate estimation and low computational time. Additionally, the result of this study can be useful to clarify the important online parameter for real-time traffic state estimation and for the model-based approach.

2. Network and Traffic Data

Lyon is the third largest city and has the second largest urban area in France. Around 2 million residents live in the metropolitan area. *GrandLyon*, a French territorial collectivity, provides network data of Lyon and surrounding towns. From this network, we extract a sub-network with lots of loop detectors and high traffic volume in Figure 1. Furthermore, 70 links which loop detectors are equipped and prove vehicles pass through are chosen from the selected area as a network of this study in Figure 1. The total link length of network in this study is approximately 10 km.

In the selected area, loop detectors are equipped on 196 links and link length penetration of equipped links against all links is 5.4 %. The data is average flow and occupancy at each loop in each 6 minutes. In this study, we use data of 6:00 to 18:00 of all days on April in 2018 as monthly data, of April 3rd as one day data, i.e. sunny and no event weekday.

Probe vehicle data is provided by *Mediamobile*, a real-time traffic information provider in many European countries. This probe vehicle data is related to their own network data. Therefore, we fuse both network definitions by the following roles, i.e. the links are on the common name street, the links have common direction and loop detectors are located on the links of probe network. Coverage of probe vehicle data, i.e. the ratio of link length according to probe trajectories by each 6 minutes and total link length of all links is around 30-40 % in the selected area and around 50-70 % in network of this study.

Figure1. Lyon network of this study and location of loop detectors (Black line: all links of selected area, Red line: chosen links, i.e. network in this study, Blue dots: location of loop detectors)

3. Methodology for Link Accumulation Estimation

3.1. Situation 1: All trajectories at are available at the link level

In case that all trajectories of individual vehicles on a link at every time periods are available, traffic accumulation $n_i(t)$ on link i at time t can be estimated by Edie's formula [6] as shown in Eq. (1).

$$n_i(t) = \frac{ttt_i(t)}{\Delta t} \tag{1}$$

$ttt_i(t)$ is the total travel time for all vehicles can be calculated by sum of travel time of each vehicle on road link i during time period Δt.

3.2. Situation 2: Probe trajectories and loop detector data are partially available

By using real-time traffic data, it is difficult to get all vehicle trajectories on each link at every time. When only a few trajectories, e.g. probe vehicle data, is available, temporal scaling factor, e.g. penetration of probe vehicles, should be determined first to accurately estimate traffic accumulation. Calculation of probe penetration requires total accumulation of all vehicles at a link at every time, which is impossible to obtain from loop detector data and probe vehicle data. Instead of using total vehicle accumulation, Geroliminis and Daganzo[7] propose to estimate the accumulation ratio between probe and all vehicles at a link level by the flow ratio observed at a loop location. This method has been named *Fishing* as probe vehicles are caught among all vehicles over the loop and the flow ratio is called Fishing Rate. Fishing Rate $\gamma_i(t)$ at link i at time t can be expressed as Eq. (2).

$$\gamma_i(t) = \frac{\frac{ttd_i^P(t)}{l_i * \Delta t}}{q_i^L(t)} \tag{2}$$

where total travel distance $ttd_i^P(t)$ is sum of travel distance of each vehicle and l_i is link length of a link i. Traffic flow directly measure from loop detector data is shown as q_i^L and $ttd_i^P/(l_i * \Delta t)$ equal to traffic flow derived from probe vehicle data. By using Fishing Rate, the traffic accumulation can be estimated as Eq. (3).

$$n_i(t) = \frac{ttt_i^P(t)}{\Delta t} * \frac{1}{\gamma_i(t)} \tag{3}$$

Furthermore, when loop detector data has some spatial or temporal lack, we need to complement Fishing Rate $\gamma_i(t)$ by using aggregate values, e.g. average Fishing Rate of many days, many places or long time period. However, the fishing method can have big variance in the estimation for small penetration rates.

3.3. Situation 3: Trajectories at a link are not available

Probe vehicle data is sparse in time, therefore sometimes no trajectory is observed in some links at several time periods. In this case, we need to estimate link accumulation from loop detector data. However, loop detector data provides occupancy instead of density, therefore two levels of transformation are necessary, i.e. occupancy to local level density and local level density to link level density. For the first transformation, it is known that local level accumulation can be calculated by dividing occupancy by Effective Vehicle Length (EVL). EVL can be calculated by sum of physical vehicle length and detection size of loops, however it is difficult to obtain detection size of loops from data. On the other hand, Coifman[8] explains that traffic velocity can be derived from EVL and occupancy. Therefore, under the assumption that traffic velocity from probe vehicle data is same as one from loop detector data, EVL $\alpha_i(t)$ on link i at each time period t can be calculated by using velocity derived from probe vehicle data as Eq. (4)

$$\alpha_i(t) = \frac{v_i^p(t) * o_i^L(t)}{q_i^L(t)} \tag{4}$$

where $v_i^p(t) = ttd_i^P(t)/ttt_i^P(t)$ is velocity derived from probe vehicle data, $o_i^L(t)$ and $q_i^L(t)$ are occupancy and traffic flow that directly calculated from loop detector data. Obviously, this requires that at some time some probe vehicle data are available to calculate EVL for a particular loop. By using velocity derived from probe vehicle data, scaling issue between local level and link level is already considered. Therefore, in this study, we use this definition of EVL to estimate link level accumulation from loop detector data as shown in Eq. (5).

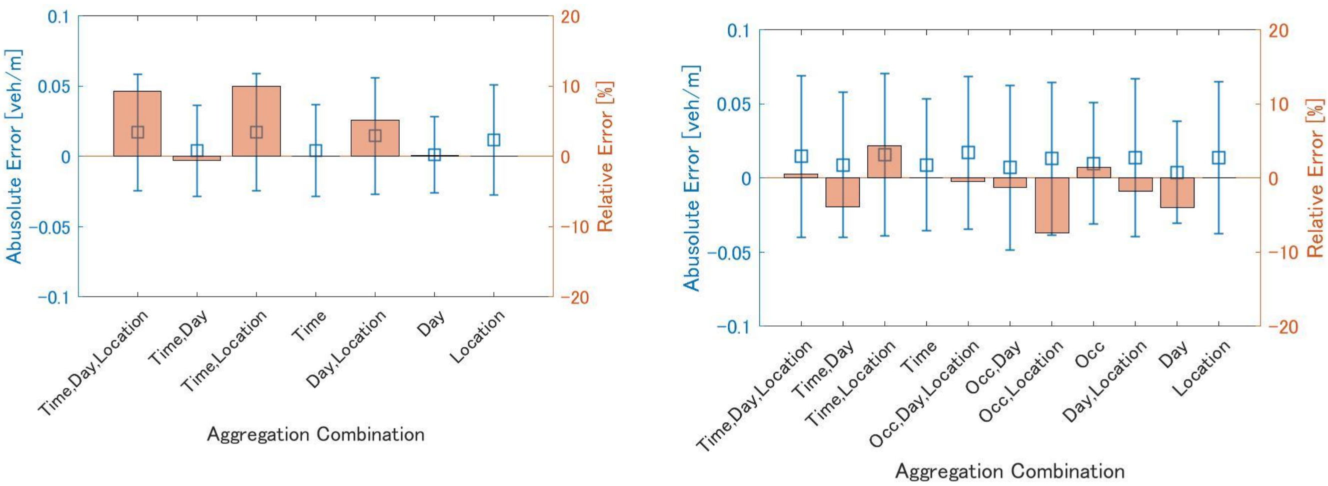
$$n_i(t) = \frac{o_i^L(t)}{\alpha_i(t)} * l_i \tag{5}$$

When trajectories from probe vehicle data is available, traffic accumulation estimation according to Eq. (5) is same as the one according to Eq. (3). However, in case that any trajectories cannot be obtained at some time periods, EVL of each link at each time cannot be calculated. At that time, it is necessary to replace EVL $\alpha_i(t)$ by using aggregate values, e.g. average EVL α' of many days, many places or long time period.

4. Accuracy Assessment of Link Accumulation Estimation

Without all vehicle trajectories, traffic accumulation by using Fishing rate $\gamma_i(t)$ can be a more accurate method. Therefore, we consider traffic accumulation estimation in case that probe vehicle data and loop detector data are available at a link at every time as the reference case.

When there are several lacks of loop detector data or probe vehicle data at some moment, we cannot calculate Fishing Rate $\gamma_i(t)$ or EVL $\alpha_i(t)$, therefore we need to replace by aggregate values. At this time, all candidates of aggregation combination, i.e. by days, loops or time periods, are assessed to clarify the best balance of data quantity. Absolute and Relative error between estimated traffic accumulation of each case and reference case are calculated at each link at every time periods as shown in Figure 2.



(1) Aggregated Fishing Rate (2) Aggregated ELV

Figure 2. Mean and STD of Absolut Error (blue) and Mean of Relative Error (red) between Reference Case and Estimated Density

According to Figure 2(1), in case of using Fishing Rate, absolute error shows that location aggregation leads to larger error than other aggregation. On the other hands, in case of time or/and day aggregation, the absolute error is quite close to zero and the relative error is quite small. Therefore, it is clear that Fishing Rate is different on each location but not different lot in time period or every day. On the other hand, in case that using EVL, STD of absolute error is larger than Fishing Rate. From Figure 2(2), only day aggregation has as small error as time or/and day aggregation of Fishing Rate according to mean and STD of absolute error. Thus, using ELV leads larger error than using Fishing Rate. When data is aggregated on all days, using the relation between occupancy and EVL leads better result than using constant value on time. However, this relation leads big error in case of location aggregation.

In conclusion, in case that probe vehicle data is available, time and/or day aggregation is efficient to calculate Fishing Rate. Also, in case that probe vehicle data is not available at the moment, day aggregation based on time dynamics or the relation between occupancy and ELV is useful to calculate ELV. Even in the worst result of both methods, the error from reference case is small enough. Therefore, using Fishing rate and using EVL can be validated for link accumulation estimation.

5. Conclusion

In this study, we show the confidence interval of each traffic accumulation estimation methods with cases that there are some data lack. As result, accumulation estimation by using Fishing Rate and EVL are effective method under several aggregations. The results of this study can be useful for considering on-line parameters or validate method for real-time estimation. For the next step of this research, to assess network level estimation with considering penetration of loop detector equipped links is important.

Reference

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

[2] Zefreh, M. M., Török, Á., & Mészáros, F. (2017). Average vehicles length in two-lane urban roads: A case study in budapest. *Periodica Polytechnica Transportation Engineering, 45*(4), 218–222.

[3] Leclercq, L., Chiabaut, N., & Trinquier, B. (2014). Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. *Transportation Research Part B*, *62*, 1–12.

[4] Kockelman, K. M. (1998). Changes in flow-density relationship due to environmental, vehicle, and driver characteristics. *Transportation Research Record*, (1644), 47–56.
[5] Leclercq, L. (2005). Calibration of flow-density relationships on Urban streets. *Transportation Research Record*, (1934), 226–234.
[6] Edie, L. C. (1963). Discussion of traffic stream measurements and definitions. *Proceedings of the Second International Symposium on the Theory of Traffic Flow*, 139–154.
[7] Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams : Some experimental findings. *Transportation Research Part B, 42*, 759–770.
[8] Coifman, B., Dhoorjaty, S., & Lee, Z. H. (2003). Estimating median velocity instead of mean velocity at single loop detectors. *Transportation Research Part C: Emerging Technologies, 11*(3–4), 211–222.