Estimating Fundamental Diagrams of Signalized Links from Aggregated Trajectories

Walid Fourati, Aleksandar Trifunović, and Bernhard Friedrich

Institute of Transportation and Urban Engineering,
Technical University of Braunschweig, Germany

November 2020

1 Introduction

Space-time continuity and availability offered by mobile sensors make them suitable to study the fundamental diagrams (FD) of existing traffic facilities individually. Seo et al. [5] presented a method to infer a triangular fundamental diagram of a freeway section using only probe trajectories and jam density k_{jam}. The method requires a minimum of three steady traffic states between each pair of consecutive trajectories. For signalized links, this requirement cannot be respected due to regular and systematic traffic interruption. The method is therefore not applicable. At the same time, signal control is rather an enabling factor to infer traffic states from sample trajectories thanks to the resulting recurrence of clear and stationary traffic states in a cyclical fashion. When trajectories are aggregated over time based on the respective cycle crossing time, traffic states can be inferred from each elementary region of a space-time grid using an extension of Edie's traffic stream measurements definition [1]. Time based aggregation allows inference on temporary sparse trajectories and reflects long-term characterization of fundamental diagrams.

First, a new clustering approach is presented following improvement on a previous method used in [2] and [3]. Next, Edie's traffic states measurement from disaggregated full-coverage space-time diagrams are extended to aggregated sampled space-time diagrams to produce scatter plots that can be fitted to a fundamental diagram. Additional considerations for dimensional homogeneity and scale proportionality are also discussed. The method is tested with a six months data set of real trajectories and the obtained capacity of the fundamental diagram is validated against a capacity obtained from a few hours of automated video measurement.

2 Methodology

Even though trajectories are first clustered then accumulated, we present the accumulation first for the sake of clarity.

2.1 Trajectories accumulation

Figure 1a illustrates the principle of trajectories accumulation[1] belonging to a same signal program (cluster) and Figure 1b illustrates the accumulation of accumulation, i.e. accumulation of clusters of trajectories. The clustering identifies the trajectories that occurred during the same signal program. It produces as many clusters as there were fixed signal timing patterns during a typical day. A cluster i is therefore identified by a time interval of the day, a cycle length C_i and a green window defined by its start and end seconds within the cycle $[gs_i, ge_i]$. When trajectories belonging to a same cluster i (i.e. signal program) are isolated, they can be shifted in time by the multiples of C_i (a modulo operation) to fit into the same artificial cycle. To complete the accumulation of all trajectories from all clusters in one space-time plot, every cycle accumulation is shifted with the respective gs_i so that the green start

[1] We opt for the term "accumulation" over "aggregation", as we think that aggregation supposes the loss of some information, whereas in our case all trajectories information is technically still present after grouping them in one plot

becomes the time origin (Figure 1b).

The obtained plot contains all trajectories that belong to a cluster. Space origin is the stop line and time origin is the relative green start of each trajectory. In particular, flowing at capacity downstream after the green start (in the positive space and time domains) and complete congestion before the stop line and before the green start (in the negative pace and time domains) are the two mostly visible traffic states in the accumulated trajectories plot.

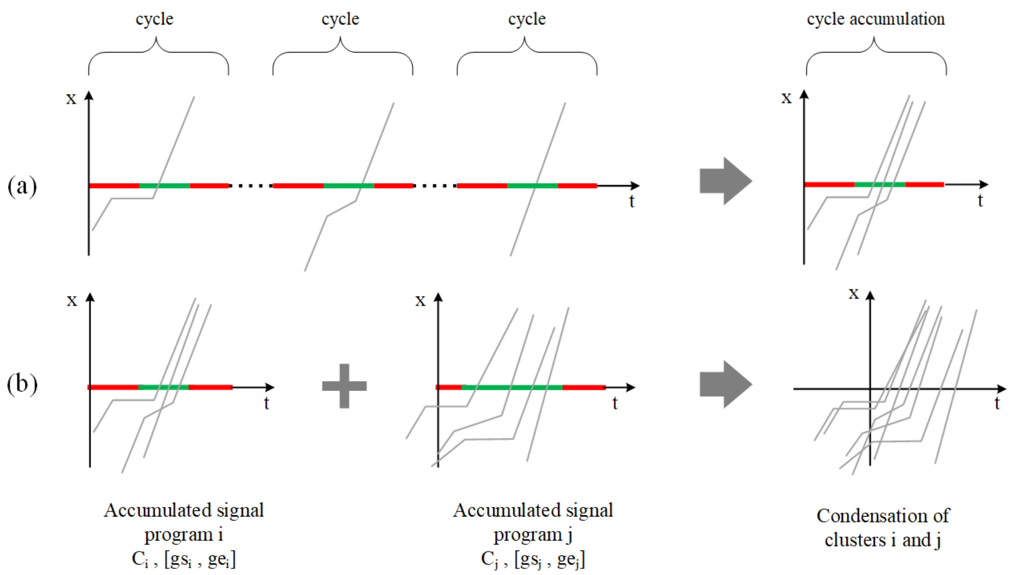


Figure 1: (a) Accumulation of trajectories belonging to a same signal program and (b) Accumulation of different clusters in a green-start referenced time scale

2.2 Trajectories' clustering

The clustering of trajectories is based solely on their respective stop line crossing time, as that instant is a direct indicator of a green signal. Every trajectory can cross the stop line during a cycle of unknown length, however, possible cycle lengths lay within a certain known range. Therefore, a series of supposed observations is created for every trajectory, each one corresponding to a cycle length candidate (CLC). For each candidate, it's possible then to obtain the crossing time inside the cycle (CCS: Cycle Crossing Second) for the assigned CLC as the rest of integer division of the absolute time by the cycle length. In order to improve the clustering reliability, a third feature is added to the dataset for clustering in addition to CLC and CCS, which is the time of the day (TOD). It is assumed that traffic signaling follows a schedule of programs where a program is served continuously during a certain time before switching to another continuous program. This implies that trajectories belonging to a same program should have occurred during a same time window of the day. It's likely that the clustering assigns more than one supposed observation (different CLCs) to a same trajectory. Ideally, the closest candidate to its respective cluster should be kept. In this research, we keep the candidate belonging to the biggest cluster in size.

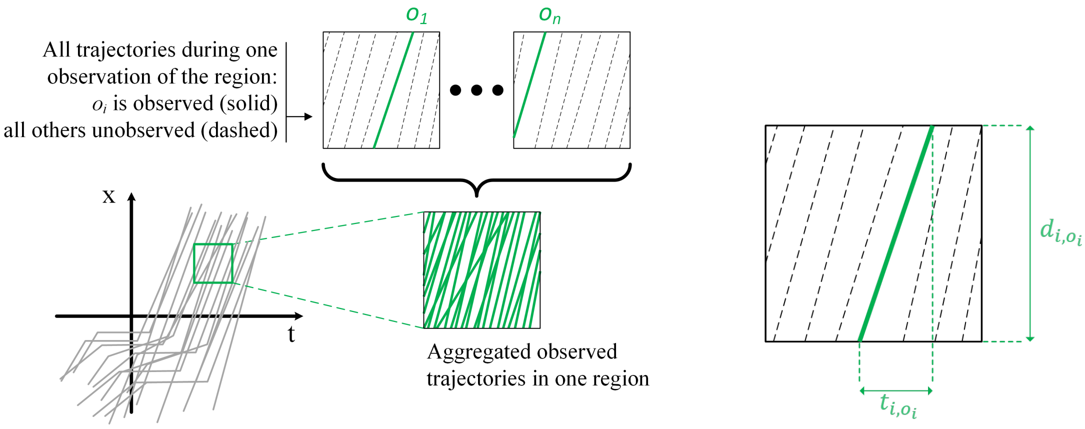
The proposed concept is independent from the clustering method, as several clustering algorithms and approaches in the literature and practice produce the desired result. In this research, the single linkage clustering [4] is adopted. The three features (CLC, CCS, TOD) are derived from the stop line crossing timestamp. The features are then normalized with the right scale and range. With a calibrated hierarchy threshold and a calibrated minimum of observations per cluster, the algorithm creates a first set of clusters. A dataset with only one cluster per trajectory is then extracted to make sure to select the best CLC for each trajectory.

2.3 Adaptation of Edie's definition

The Edie's generalized definitions of traffic states from space-time plane regions can be adapted to our problem. For a given region A [2] that is observed n times and crossed at each observation i by m_i vehicles, we can define the average flow, density and speed as follows:

$$\begin{aligned}
\overline{q} &= \frac{\sum\limits_{i\leq n} q_i}{n} = \frac{\sum\limits_{i\leq n} m_i\,\widetilde{d_i}}{n\,|A|} \text{ where } \widetilde{d_i} = \frac{\sum\limits_{j\leq m_i} d_{i,j}}{m_i}\\
\overline{k} &= \frac{\sum\limits_{i\leq n} k_i}{n} = \frac{\sum\limits_{i\leq n} m_i\,\widetilde{t_i}}{n\,|A|} \text{ where } \widetilde{t_i} = \frac{\sum\limits_{j\leq m_i} t_{i,j}}{m_i}\\
\overline{v} &= \frac{\sum\limits_{i\leq n} v_i}{n} = \frac{\sum\limits_{i\leq n} \frac{\widetilde{d_i}}{\widetilde{t_i}}}{n}
\end{aligned} \tag{1}$$

where $|A|$ is the surface of the area, $d_{i,j}$ and $t_{i,j}$ are respectively the distance and time of travel of the trajectory j inside the region A during the observation i.



(a) (b)

Figure 2: An observed region in the accumulated plot is a superposition of single observations among all actual trajectories

The accumulated space-time plot resulting from the clustering in the previous step (bottom right plot in Figure 1) is a superposition of single observations among all actual trajectories as illustrated in Figure 2a. We aim to perform an equivalent measurement of aggregated flow q^{Agg} and density k^{Agg} in each region A. Let $o_i \in [1, m_i]$ be the value of the j index of the observed vehicle in the observation i. The aggregated quantities are defined as follows:

$$q^{Agg} = \frac{\sum\limits_{i\leq n} d_{i,o_i}}{|A|} \quad ; \quad k^{Agg} = \frac{\sum\limits_{i\leq n} t_{i,o_i}}{|A|} \quad ; \quad v^{Agg} = \frac{\sum\limits_{i\leq n_A} v_{i,o_i}}{n_A} \tag{2}$$

The displacement d_{i,o_i}, the travel time t_{i,o_i} and the speed v_{i,o_i} of the vehicle o_i in the observation i (Figure 2b) are directly measurable, which makes the defined quantities q^{Agg}, k^{Agg} and v^{Agg} also directly measurable. The dimensions of q^{Agg} and k^{Agg} are [observations per unit of time] and [observations per unit of space] respectively. They are dimensionally not homogeneous to flow and density, the relationship between them remains nevertheless comparable to the relationship between flow and density under certain conditions as discussed below. We define the following ratios between full-scale average quantities and measured aggregated quantities:

$$\rho^q = \frac{\overline{q}}{q^{Agg}} = \frac{\sum\limits_{i\leq n} m_i\,\widetilde{d_i}}{n \sum\limits_{i\leq n} d_{i,o_i}} \quad ; \quad \rho^k = \frac{\overline{k}}{k^{Agg}} = \frac{\sum\limits_{i\leq n} m_i\,\widetilde{t_i}}{n \sum\limits_{i\leq n} t_{i,o_i}} \tag{3}$$

[2] All quantities and equations are for a single region, all the regions are the same size, we therefore omit the A index for the sake of simplicity

ρ^q and ρ^k have the same physical dimension: [number of vehicles per observation]. In a certain way, they indicate the mathematical inverse of the observation rate i.e. the penetration rate of the used sampled trajectories dataset when observing the region A. The exact equivalence between the $(\overline{k}, \overline{q})$ relationship and the (k^{Agg}, q^{Agg}) relatipship is an immediate result of the equality between ρ^q and ρ^k in all the regions of the plot. In order to establish that equality, we first make the assumption that o_i represents an average vehicle - in terms of travel time and distance - of the total group of vehicles in the observation i in the region A. This allows to establish the following approximations:

$$\widetilde{d_i} \approx d_{i,o_i} \quad ; \quad \widetilde{t_i} \approx t_{i,o_i} \quad ; \quad \frac{\widetilde{d_i}}{\widetilde{t_i}} \approx \frac{d_{i,o_i}}{t_{i,o_i}} = v_{i,o_i} \tag{4}$$

We can then simplify ρ^q and ρ^k:

$$\rho^q = \frac{\sum\limits_{i \leq n} m_i \, d_{i,o_i}}{n \sum\limits_{i \leq n} d_{i,o_i}} = \frac{\overline{m}^d}{n} \quad ; \quad \rho^k = \frac{\sum\limits_{i \leq n} m_i \, t_{i,o_i}}{n \sum\limits_{i \leq n} t_{i,o_i}} = \frac{\overline{m}^t}{n} \tag{5}$$

where $\overline{m}^d$ and $\overline{m}^t$ are respectively the distance-weighted and the time-weighted average values of m. To illustrate these averages, we consider a region with a certain number of crossing vehicles m. If we observe more slower than faster vehicles, there would be more region crossing times t_{i,o_i} that are high, and less displacements d_{i,o_i} that are high, thus $\overline{m}^t$ would be bigger than $\overline{m}^d$, and inversely. This is interpreted by a relationship between the dispersion of speed of observed vehicles v_{i,o_i} and the equality between $\overline{m}^d$ and $\overline{m}^t$, consequently between ρ^q and ρ^k. We should therefore take more into consideration the regions with less disperse speeds to get closer to the equality between $\frac{\overline{q}}{q^{Agg}}$ and $\frac{\overline{k}}{k^{Agg}}$. This result is formalized by a weight used when fitting the scatter of regions to a fundamental diagram model.

3 Results

The data set of sampled trajectories used for testing corresponds to the first half of 2019 at the intersection Amalienplatz in Braunschweig, Germany on the direct southbound movement. The obtained number of observations (accumulated cycles) is n=1425. Automated computer-vision detection and tracking used in a previous case study at the same intersection [3] allowed to measure an average jam density during 4 hours of video recording on a day that falls within the collection period of the trajectories. The obtained value k_{jam} is used to adjust the scale of the probe based fundamental diagram as in [5].

In Figure 3-a and 3-b accumulated trajectories are augmented respectively with the regions' sample sizes and the inverse of the standard error of speed (which is used as fitting weight). Regions with sample sizes below the average are eliminated. Figure 3-c and 3-d show two tested fittings on the obtained scatter, once using an exponential v-k model and once using a triangular q-k model. The upscaling of the q-k triangular FD using the ratio between k_{jam} and the measured k_{jam} results in a capacity of 1499.069 [veh/h] which is close to the average observed value of 1551.052 [veh/h] from a single-day automated video measurement [3].

4 Conclusion

The presented method allows to construct an empirical FD from sampled trajectories at signalized intersections through clustering and accumulation. The main contribution is the purely empirical measurement of intersections' capacities over a long period, which provides a reliable figure as important input in signal control planning, traffic modeling or simulation of existing networks. The few assumptions made to reach the full scale estimation shall be validated with an in-depth empirical and simulation study. The sensitivity of the result to the resolution of regions grid should also be studied for a better fine-tuning.

5 Acknowledgement

This research is supported by the German Research Society DFG under the grant FR 1670/7-1 *Methoden für die kontinuierliche raum-zeitliche Analyse der Verkehrsqualität in Straßennetzen*

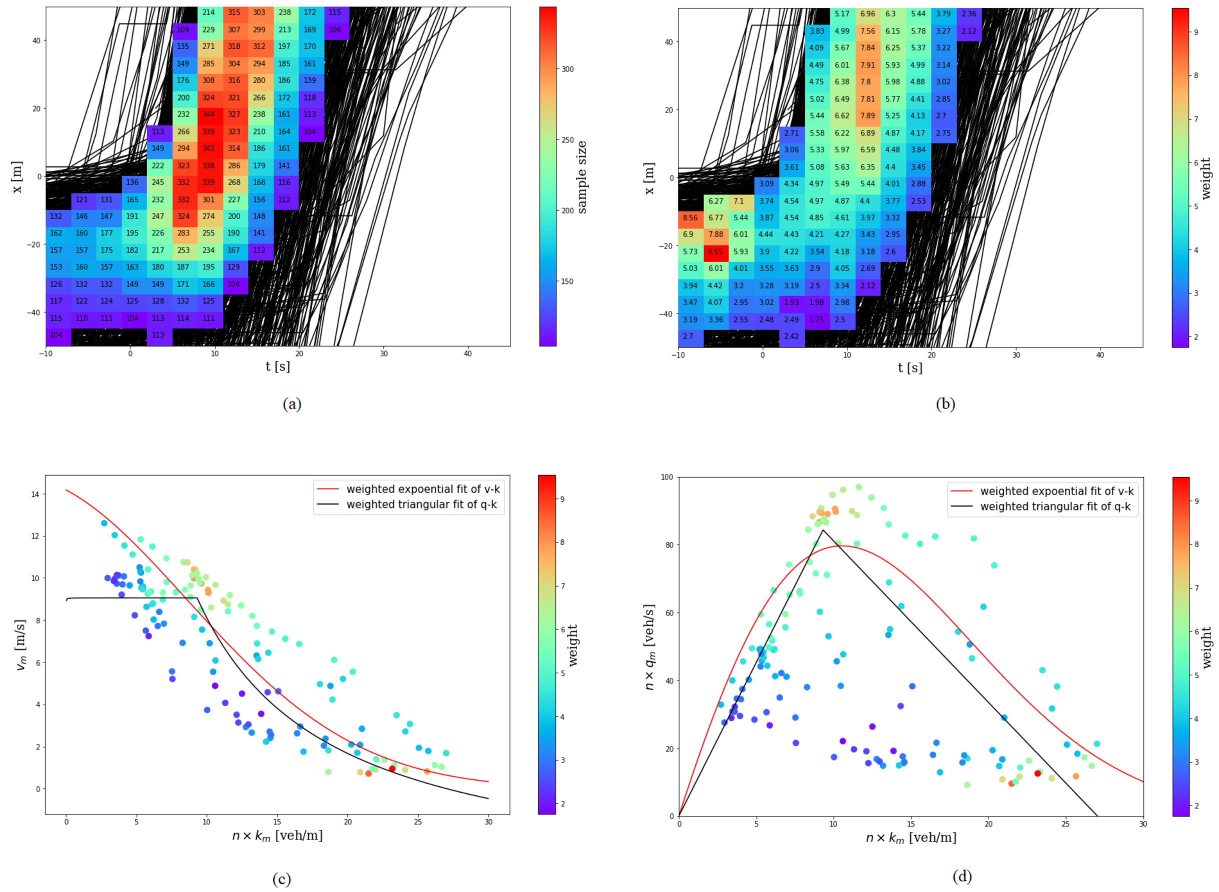


Figure 3: Results with a real data set: condensed trajectories augmented with region sample sizes (a) and regions homogeneity indicator (b), obtained v-k (c) and q-k (d) FD scatter with two fittings

References

[1] Leslie C. Edie. Discussion of traffic stream measurements and definitions. In *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, pages 139–154, 1965.

[2] Walid Fourati and Bernhard Friedrich. Trajectory-based measurement of signalized intersection capacity. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(10):370–380, 2019.

[3] Walid Fourati, Aleksandar Trifunovic, Morten Flesser, and Bernhard Friedrich. Trajectory based Monitoring of Saturation Flow Rate with Video Calibration: a Case Study. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4287–4294, Auckland, New Zealand, October 2019. IEEE.

[4] J. C. Gower and G. J. S. Ross. Minimum Spanning Trees and Single Linkage Cluster Analysis. *Applied Statistics*, 18(1):54, 1969.

[5] Toru Seo, Yutaka Kawasaki, Takahiko Kusakabe, and Yasuo Asakura. Fundamental diagram estimation by using trajectories of probe vehicles. *Transportation Research Part B: Methodological*, 122:40–56, April 2019.