Design hourly volume estimation at freeway nodes using floating car data

Magdalena Schilling*[1], Marvin V. Baumann[2], Jörg Sonnleitner[1], Markus Friedrich[1], and Peter Vortisch[2]

[1]Institute for Road and Transport Science, University of Stuttgart, Germany
[2]Institute for Transport Studies, Karlsruhe Institute of Technology, Germany

Short summary

The estimation of the Design Hourly Volume (DHV) is an essential step for a traffic assessment. At freeway nodes, not all ramps are detected by permanent traffic count (PTC) stations. Therefore the German HCM recommends additional short-term counts (STC) to determine the DHV. Since conducting STC is usually associated with high effort, the question arises whether the information obtained by STC can also be derived from Floating Car Data (FCD). We propose an approach for processing the FCD in order to apply it instead of the STC for the determination of the DHV at ramp junctions. The performance of the method is evaluated on five nodes, for which FCD from 2017 and a reference database covering all 8,760 hourly volumes of all ramps and main lanes of the road section are available. The result show, the usage of representative FCD days is possible.
Keywords: Design Hourly Volume Estimation, Floating Car Data, Highway Capacity Manual, Permanent Traffic Counts

1 Introduction

Design hourly volume (DHV) estimation is an essential step when it comes to estimating the Level-of-Service (LOS) of traffic facilities. In the Highway Capacity Manual (HCM, National Academies of Sciences (2016)) and the German HCM (FGSV (2015)), the DHV is determined based on a traffic volume estimation concept known as the n^{th} hour or respectively the hour of the year with the n^{th} highest traffic volume. To calculate this n^{th} hour precisely, a permanent traffic count (PTC) station is necessary at the corresponding traffic facility, since the traffic volume for all 8,760 hours of the year must be known. To determine the DHV at ramp junctions of nodes, supplementary short-term traffic counts (STC) are also necessary since, usually, not all ramps are recorded with PTC. The German HCM proposes a method for this process, which was validated and enhanced to a concept of the n^{th} highest saturated hour by Baumann et al. (2023).
The method proposed by the German HCM determines the representative turning flows at the node from STC on the ramps, which are then extrapolated to the DHV using the PTC available at the node. Conducting STC is generally associated with high effort, which is why the question arises if the required representative turning flows can be determined using Floating Car Data (FCD). Ceccato et al. (2022) demonstrate that the use of FCD is competitive compared to traditional data sources in terms of cost-effectiveness. Furthermore, Vogt et al. (2019) and Dabbas et al. (2020) show that the data fusion of FCD and PTC enables the estimation of origin-destination matrices for motorway networks. Travel times and route choice probabilities derived from FCD can moreover be used as input for Dynamic Traffic Assignment models to map OD matrices to link flows (Nigro et al. (2018); Tsanakas et al. (2022)). Nohekhan et al. (2021) use FCD, temporary volume counts (e.g., a week), and road characteristics to estimate hourly traffic volumes on off-ramps. FCD can also be used to determine travel time (Olszewski et al. (2018)), free flow speed (Diependaele et al. (2016)), or operating speed (Bruwer et al. (2021); Lobo et al. (2018)) on motorways.
The literature review demonstrates that traffic flow assessment using FCD is possible, but to the best of the authors' knowledge, there are no approaches in the literature that use FCD to determine the DHV in the context of the HCM, German HCM, or similar international guidelines. Therefore, this paper examines the potential of using FCD as a substitute for STC in estimating the DHV at nodes. We propose an approach for processing the FCD and evaluate the results afterward using FCD of route sections with a total length of 15 km and compared with STC results from Baumann et al. (2023).

2 Methodical approach

Concept and data availability scenarios for DHV estimation at freeway nodes

An example of how to combine PTC and STC using topological relationships of a freeway interchange is shown in the German HCM. It assumes that a cloverleaf interchange has eight PTC, one for each inflow or outflow, and at least two STC for each ramp junction. Each PTC defines a specific demand-situation, which needs to be analyzed. In the following, these demand-situations are referred to as PTC demand-situation. Each PTC demand-situation describes a temporary state with consistent traffic flows at the entire interchange, such that inflow equals outflow. In the German HCM these demand-situations are defined using the method '50^{th} hour of the PTC'. In this example it leads to eight demand situations, which may occur on different weekdays and times of day. STC are usually conducted at a different date. Therefore, the German HCM uses the day hour of the PTC demand-situation to derive a second demand-situation based on the STC (STC demand-situation). In the next step a matrix estimation method is applied using the PTC demand-situation as boundary conditions and the STC demand-situation as initial matrix to derive DHV for each count station. This procedure is repeated for all eight PTC demand-situations of the cloverleaf inter-change. After that, all eight demand-situations are evaluated. For each ramp junction, a separate saturation rate is estimated per demand-situation. The resulting saturation rate of a ramp junction is the worst-case saturation rate of all demand situations considered.
For this concept for DHV estimation at freeway nodes we compare the usage of traffic flows derived from FCD instead of STC. Furthermore, the number of PTC can be varied. To understand the impact of these data sources, the following data availability scenarios are defined and will be analyzed:

1. 'PTC: in-/outflow main lanes, STC: -': This data availability scenario again uses no STC data, but the numbers of PTC stations are reduced to one count station for each inflow or out-flow on the main lanes of the node. This leads to eight PTC stations at a four-leg-interchange and to four PTC stations at a freeway exit.
2. 'PTC: in-/outflow main lanes, STC: all': This data availability scenario adds STC information for all counting stations (Figure 4, third row).
3. 'PTC: in-/outflow main lanes, FCD: all': This data availability scenario equals the data availability scenario above but uses FCD information instead of STC.
4. 'PTC: in-/outflow main lanes, FCD: representative traffic days': This data availability scenario uses representative traffic days obtained from FCD instead of single days.

Data basis

Hourly FCD-hits are available for 164 working days in 2017 on route sections with a total length of 15 km. 'Hits' refer to the number of vehicles recorded. The number of hits is a subset of the total traffic volume. The route sections are part of 63 ramp junctions for whom PTC are available. Missing hits on some count stations are derived based on adjacent hits. The ramp junctions are part of five nodes: two interchanges and three freeway exits.

Preprocessing to generate consistent FCD for all count stations

An initial plausibility check reveals some inconsistencies:

- There are some negative values as results of balancing checks considering the topology.
- Conflicts appear if several FCD route sections can be assigned to the same count station, leading to an over determination of some count stations by FCD route sections.

Thus, a matrix estimation procedure is executed to get consistent FCD using the VFlowFuzzy algorithm implemented in PTV Visum (PTV AG (2022)). This implementation allows to define tolerances for each count station in case the hits are too inconsistent. For each hour the related hourly hits are considered and tolerances are increased successively until a solution of consistent hits is found. If no solution is found, it is analyzed which count station has implausible hits. This allows us to ignore these values or increase the tolerance for these count stations.

Generate representative traffic days

In order to get more robust daily hits, we use the amount of FCD days to generate representative traffic days. A representative traffic day has a daily distribution, which occurs as often as possible in this or a similar way.
A method suitable for matching the properties of daily distributions is cluster analysis.
We define three types of traffic days:

- Monday
- Tuesday, Wednesday and Thursday
- Friday

For the cluster analyses of a traffic day type we use network load curves containing daily hits of all count stations for the days belonging to the traffic day type. Depending on the size of the distance measure and the clustering algorithm, there will be a different number of clusters and consequently a different distribution of days per cluster.
We use the average linkage cluster algorithm and the GEH value serves as distance measure for the comparison of hourly traffic volumes. The cluster containing most days is defined as main cluster. The smaller the maximal allowed distance, which is given is input by the user, the more clusters there are. However, the days in a cluster are more similar with a lower distance measure, the main cluster is more characteristic and it is less influenced by smoothing due to averaging of more divergent days.
We choose GEH 6, resulting in main cluster that represent about 50% of the days belonging to the traffic day type (table 1).

Table 1: result of cluster analysis using GEH-value ≤ 6.

traffic day	**days total**	**number of clusters**	**days in main cluster**	**share of days in main cluster**
Mo	32	9	18	56%
TuWeThu	96	22	51	53%
Fri	34	9	16	47%

DHV estimation combining FCD and PTC on nodes

To adapt the DHV estimation at nodes as described above to representative FCD days, we calculate for each PTC demand-situation scenarios for all representative days. This leads to three scenarios per PTC. The day hour of the PTC demand-situation defines the hour of the representative day. Then the day hour of the representative day is used to get the traffic flow matrix. After that, the process using matrix estimation and determination of the worst-case saturation is the same.

3 Evaluation

Combining PTC and FCD enables the DHV estimation for each ramp junction of a node. In this study, the DHV is defined as the 50^{th} hour as it is the common standard in Germany. Since a consistent determination of the 50^{th} hour based on the traffic volume is not possible at ramp junctions due to several traffic flows, the 50^{th} highest saturated hour (calculated according to the methods of the German HCM) is used instead of the hour with the 50^{th} highest traffic volume (Baumann et al. (2023)).
For all nodes a reference database is available that provides all 8,760 hourly volumes of 2017 for all ramps and main lanes (Baumann et al. (2023)). The reference database enables the calculation of the saturation of each ramp junction for all hours. So the actual 50^{th} highest saturated hour can be determined for each ramp junction, which is referred to as the reference scenario in the following. In order to analyze the performance of the proposed method, the method is applied

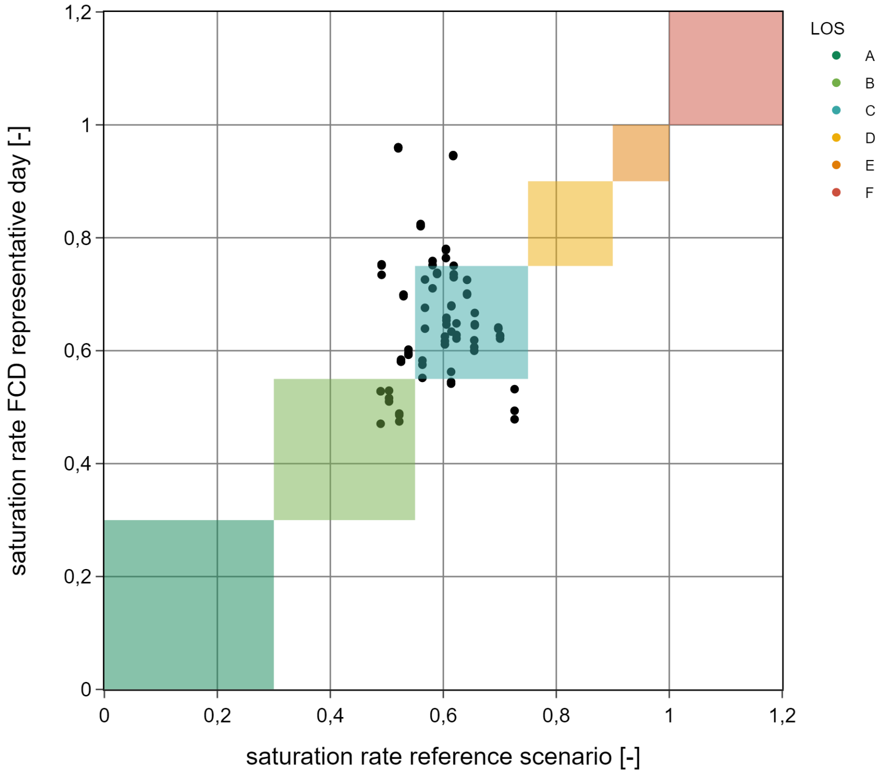

Figure 1: Exemplary results of a node for an estimation scenario using FCD representative days.

to all ramp junctions of the nodes considered, and for each ramp junction, the obtained results (referred to as estimation scenario in the following) are compared with the reference scenario. Figure 1 shows the correlation between reference and estimation scenario exemplary for all ramp junctions of a node. Each point represents the result of one ramp junction. The colored squares illustrate the corresponding LOS according to the German HCM. If a point is located in one of the squares, the saturation rate of the estimation scenario results in the same LOS as that of the reference scenario. Otherwise, the estimation scenario differs from the target LOS of the reference scenario. For a further aggregation of the results, we introduce the metric of the 'average LOS-accuracy'. Based on the results of the estimation scenario, this metric describes the relative share of the estimation scenario that achieves the target LOS of the reference scenario. Regarding the visualization in fig. 1, this corresponds to the proportion of points located within one of the colored LOS squares.

$$\textit{average LOS-accuracy} = \frac{n(LOS_{es} = LOS_{rs})}{n} \tag{1}$$

with

n	=	number of estimation scenarios
LOS_{es}	=	calculated LOS of estimation scenario
LOS_{rs}	=	LOS of corresponding reference scenario
$n\left(LOS_{es} = LOS_{rs}\right)$	=	number of estimation scenarios, which hit the LOS of their reference scenario

4 Results and discussion

The aggregated results for 63 ramp junctions on the several data availability scenarios are shown in table 2, fig. 2 allows a less aggregated and more detailed view. Figure 2 shows the range of saturation for the STC and FCD scenarios for some ramp junctions. Additionally the results of the reference scenario and for the data availability scenario considering clusters as representative traffic days are included in the figure.
The results lead to the following conclusions:

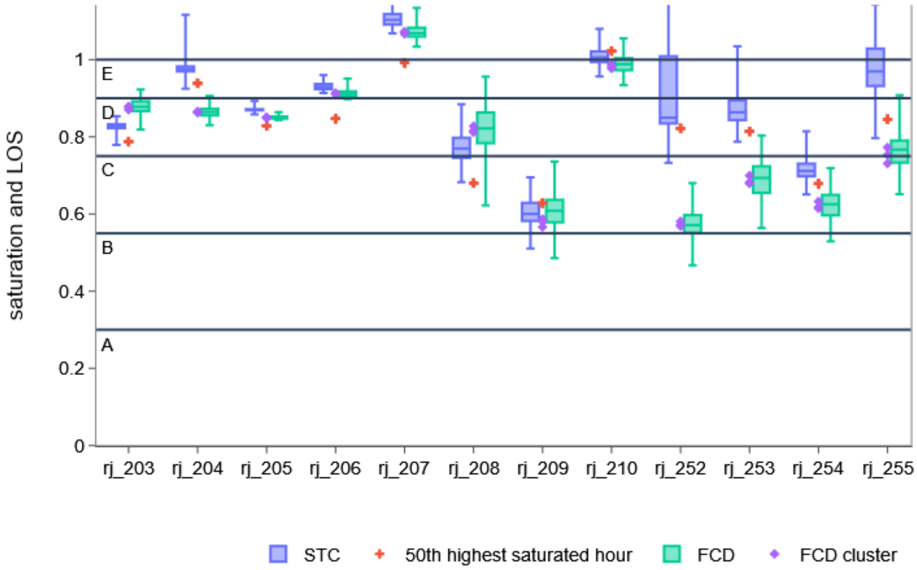

Figure 2: saturation using different data availability scenarios for some ramp junctions.

Table 2: LOS-accuracy for different data availability scenarios.

data availability scenario	**LOS-accuracy**	**LOS-under-estimation**	**LOS-over-estimation**	**number of ramp junctions**
'PTC: in-/outflow main lanes, STC: -'	33%	2%	65%	63
'PTC: in-/outflow main lanes, STC: all'	87%	3%	10%	63
'PTC: in-/outflow main lanes, FCD: all'	73%	10%	17%	63
'PTC: in-/outflow main lanes, FCD: representative traffic days'	78%	10%	12%	63

- STC or FCD on all ramps are crucial for an accurate estimation of the DHV.
- The method recommended by the German HCM - conducting STC at all ramps of a node with subsequent matrix correction or extrapolation at the nearest PTC stations - provides a data basis with which the hour with the 50^{th} highest saturation is well met.
- Nevertheless, the saturation (and LOS) varies, depending on the day on which the STC is conducted.
- In this context, it must be taken into account that for the STC considered, it is assumed for reasons of convenience that the results of STC and PTC derive from the same year, while for practical reasons the STC is often conducted a year after the year of the PTC.
- The proposed FCD method underestimates the traffic flow on ramps in some cases.
- Using representative days slightly improves the LOS-accuracy.

For further research it would be interesting to expand the approach of using FCD as introduced in this paper to additional nodes or data sources. Further evaluation of the robustness of clustering to get representative traffic days is necessary. Furthermore, it can be expected that the performance of the proposed method will increase in the future, as the availability of FCD will improve and will thus lead to more representative results.

Acknowledgements

This paper was based on research sponsored by the German Federal Ministry for Digital and Transport, represented by the Federal Highway Research Institute.

References

Baumann, M. V., Schilling, M., Friedrich, M., Reichert, S., Vortisch, P., & Wassmuth, V. (2023). Design hourly volume estimation at freeway nodes from short-term counts. *(unpublished manuscript, under review)*.

Bruwer, M. M., Andersen, S. J., & Merrick, W. (2021). Measuring the impact of roadworks on traffic progression using floating car data. Southern African Transport Conference 2021.

Ceccato, R., Gecchele, G., Rossi, R., & Gastaldi, M. (2022). Cost-effectiveness analysis of Origin-Destination matrices estimation using Floating Car Data. Experimental results from two real cases. *Transportation Research Procedia*, *62*, 541–548. (Publisher: Elsevier)

Dabbas, H., Fourati, W., & Friedrich, B. (2020). Floating car data for traffic demand estimation-field and simulation studies. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1–8). IEEE.

Diependaele, K., Riguelle, F., & Temmerman, P. (2016). Speed behavior indicators based on floating car data: results of a pilot study in Belgium. *Transportation research procedia*, *14*, 2074–2082. (Publisher: Elsevier)

FGSV. (2015). *Handbuch für die bemessung von straßenverkehrsanlagen: Hbs* (Vol. 299 B). FGSV-Verlag.

Lobo, A., Amorim, M., Rodrigues, C., & Couto, A. (2018). Modelling the operating speed in segments of two-lane highways from probe vehicle data: a stochastic frontier approach. *Journal of Advanced Transportation*, *2018*. (Publisher: Hindawi)

National Academies of Sciences. (2016). *Highway capacity manual: Hcm* (6th ed.). Washington, D.C..

Nigro, M., Cipriani, E., & del Giudice, A. (2018). Exploiting floating car data for time-dependent Origin–Destination matrices estimation. *Journal of Intelligent Transportation Systems*, *22*(2), 159–174. (Publisher: Taylor & Francis)

Nohekhan, A., Zahedian, S., & Haghani, A. (2021). A deep learning model for off-ramp hourly traffic volume estimation. *Transportation Research Record*, *2675*(7), 350–362. (Publisher: SAGE Publications Sage CA: Los Angeles, CA)

Olszewski, P., Dybicz, T., Jamroz, K., Kustra, W., & Romanowska, A. (2018). Assessing highway travel time reliability using Probe Vehicle Data. *Transportation Research Record*, *2672*(15), 118–130. (Publisher: SAGE Publications Sage CA: Los Angeles, CA)

PTV AG. (2022). *Ptv visum: Manual.* Karlsruhe.

Tsanakas, N., Gundlegård, D., & Rydergren, C. (2022). O–D matrix estimation based on data-driven network assignment. *Transportmetrica B: Transport Dynamics*, 1–32. (Publisher: Taylor & Francis)

Vogt, S., Fourati, W., Schendzielorz, T., & Friedrich, B. (2019). Estimation of origin-destination matrices by fusing detector data and Floating Car Data. *Transportation research procedia*, *37*, 473–480. (Publisher: Elsevier)