QUALITY OF FLOATING CAR DATA

Alexander Sohr, alexander.sohr@dlr.de

Elmar Brockfeld, elmar.brockfeld@dlr.de

Sascha Krieg, sascha.krieg@onlinehome.de

German Aerospace Center (DLR) Institute of Transportation Systems, Rutherfordstr. 2, 12489 Berlin, Germany

ABSTRACT

Meanwhile, Floating Car Data (FCD) is a widely available and affordable data source for traffic surveillance. The German Aerospace Center, Institute of Transportation Systems (DLR-TS) receives FCD from taxi-fleets ranging from 300 to 4000 vehicles from various German and European cities since the year 2002.

To extract common traffic variables like travel times or travel speeds and generate traffic information out of it, the raw GPS data of the vehicles are matched to a digital road network. However, compared to data from inductive loops, where each vehicle passing a loop generates data and the whole traffic flow is covered, the FCD represent only singular measurements. Thus, they are very noisy (especially when passing signalized intersections) and typically need to be aggregated to yield useful information. This means, a single delivered value is not very reliable and does not necessarily represent the state of the traffic flow on the according street segment. Data from some vehicles have to be aggregated and smoothed for some time interval and/or space to obtain reliable traffic information.

Keywords: Floating Car Data, Traffic Management, Fleet-Size, Sample-Interval, Data-Quality, Aggregation, Quality-Definition for FCD

EXPLORATIVE DATA ANALYSIS

To determine the quality of the derived traffic variables from FCD, it is necessary to know the ‘ground-truth’. One approach is to compare FCD against loop-detector data, because loops continuously collect traffic information. In fact with FCD we generate area wide traffic condition information, which can only selectively be evaluated with loop-data. Other approaches are dedicated travel time measurement campaigns, or simulation studies, which yield the most comprehensive results.

During the project Orinoko (Brockfeld, E. (2008)) DLR-TS has set up a simulation of the Nuremberg VLS-area (fig. 1). This area is situated south-east from the Nuremberg city center, surrounded by federal streets in the east and west and by a highway in the south. Along the surrounding federal streets in direction to the city center, the VLS-area has a dimension of 6.5km and transverse direction to that 3.5km.

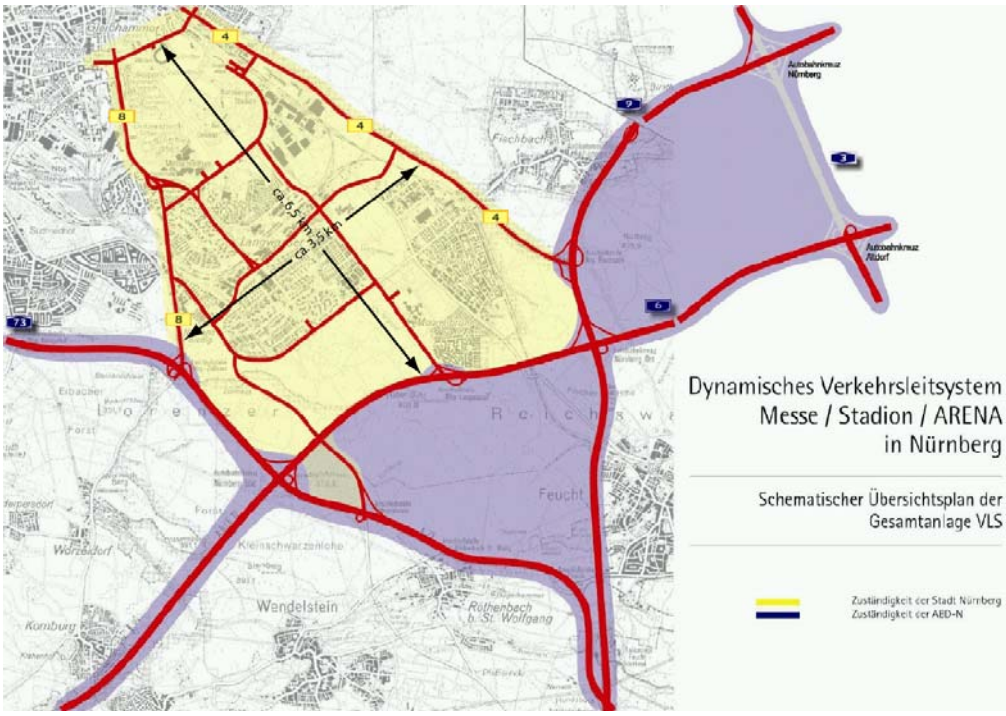

Figure 1 – VLS area in the south east of Nuremberg

The whole VLS-area is equipped with 54 loop measuring points, so that every incoming and outgoing vehicle is detected. The simulation was set up for the 18th july 2007, because this is a normal traffic day. Figures 2 – 3 show that the simulated traffic flow and the measured traffic flow for that day are nearly the same. Furthermore the VLS-Simulation has included the complete signal control plan. (Brockfeld, E. (2008))

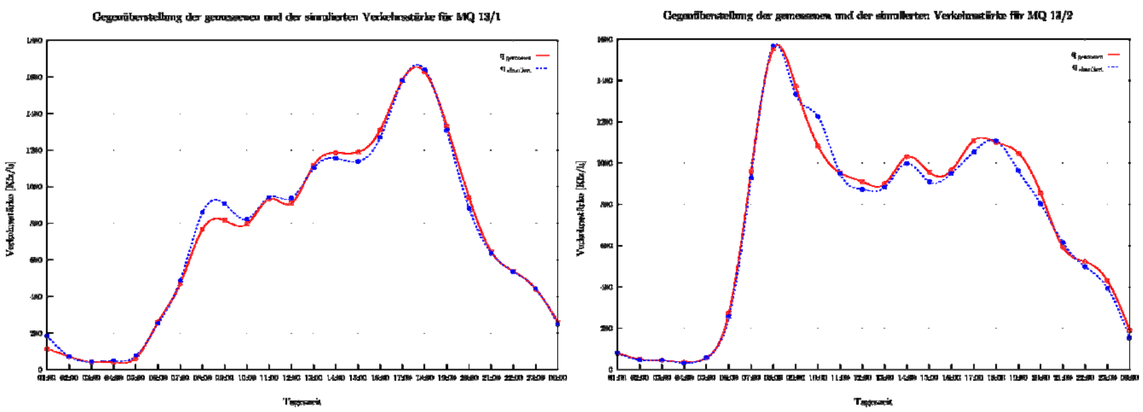

Figure 2 – measured vs. simulated traffic flow on Regensburger Straße (B4) left: MQ 13/1 direction highway right: MQ 13/2 direction city center

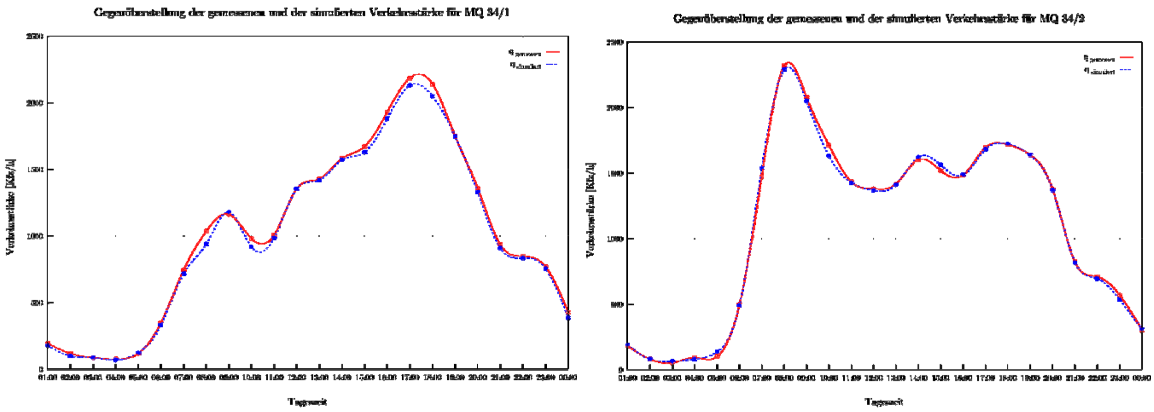

Figure 3 – measured vs. simulated traffic flow on Münchener Straße (B8) left: MQ 34/1 direction highway right: MQ 34/2 direction city center

Hence, the simulation represents the normal traffic on that day.

DATA ANALYSIS

To show the theoretical possible coverage of the road network it is assumed that every vehicle in the net can be used as a probe vehicle. The simulated vehicles deliver their position every second. So with a simple random filter different equipment rates η and delivering intervals T – the average time between two GPS-measurements – can be tested. Then the number of hit edges in the road network can be compared. The result is as one would suggest it: The better the equipment rate and the shorter the delivering interval, the higher is the number of hit roads (edges) in the network. (see figure 4)

$$f_N(\eta,T) = 100\frac{\eta(\eta,T) - N}{N}$$

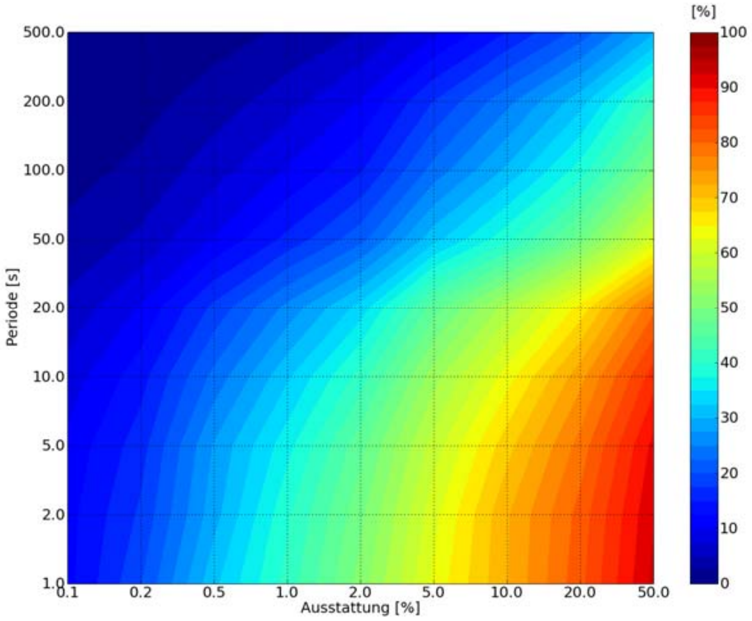

Figure 4 – percentage of 'hitted' edges as function of equipment rate and delivering interval

In addition to the coverage above, another aspect is the comparison of the measured and the simulated speeds. As quality measure the relative error (in %) f_V between the real (simulated) Speed V_{SIM} and the probe vehicle speed V_{FCD} is used.
Figure 5 shows this relative error as a function of equipment rate and delivering interval. It is obvious, that the equipment rate has the greatest impact on the quality. From equipment rates of 0.5% the relative error f_V is below 5% if the delivering interval is not too big. Even for very short intervals the error is bigger, that is because the small sample intervals <10s cause the delivering of redundant information.

$$f_V(\eta,T) = 100\frac{V_{FCD}(\eta,T) - V_{SIM}(\eta,T)}{V_{SIM}(\eta,T)}$$

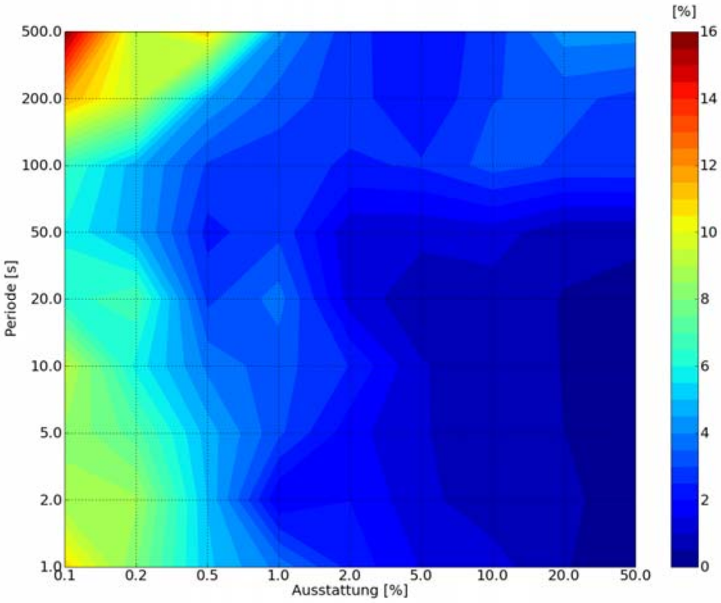

Figure 5 – relative error as a function of equipment rate and delivering interval

COMPARISON TO PROCESSED FCD

So far only the “raw”- FCD Positions were compared with the simulation, now the processed FCD are used. FCD processing means that the raw positions are map matched and a reconstruction of the driven speeds is done. Therefore the filtered raw positions for the different delivering intervals and equipment rates are treated like normal positions delivered by FCD fleets. When these data were processed, the processing had some limitations, the delivering rate has to be between 20 and 210s and there were some problems with higher equipment rates.
First the number of hit edges was compared. The processed FCD hit more edges than the raw data, because the processing reconstructs the driven trajectory, so for all edges between two positions is a speed calculated. The result in figure 6 shows that the delivering rate has not that strong impact as on the raw positions.

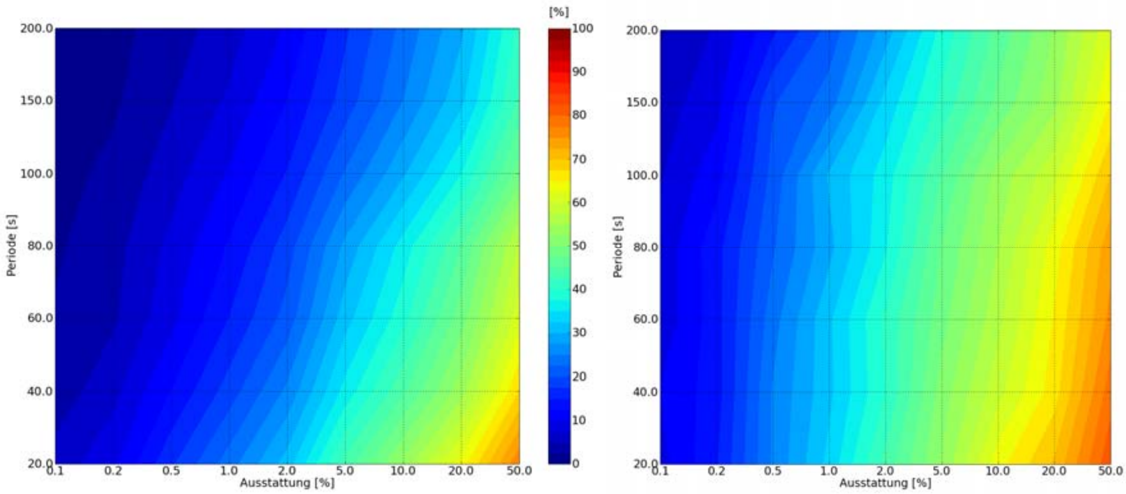

Figure 6 – percent of hit edges as function of equipment rate and delivering interval for raw Positions (left) and processed FCD (right)

The comparison of the generated Speeds in figure 7 show that the error for processed FCD for equipment rates below 0.5% is much smaller than for the raw Positions. As mentioned above equipment rates above 1% are not focus of the used processing method.

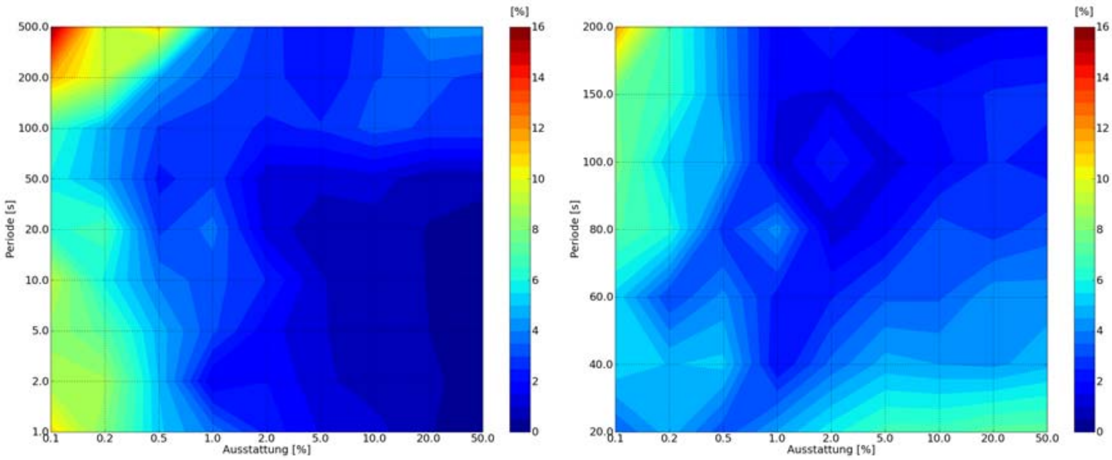

Figure 7 – relative error as a function of equipment rate and delivering interval for raw Positions (left) and processed FCD (right)

Since the equipment rate is often not the parameter which can be influenced, good results can be achieved if the delivering frequency does not exceed 60s (Krieg, S. (2008)).

AREA SIZE

Until now the investigations were made for a (small) closed area, because it was necessary to know the whole traffic to compare it. But another influencing factor on the quality of the FCD is the area size, for that the data is related. To analyse the influence of the area size the FCD from the city of Berlin was taken, because DLR-TS has data of the biggest Taxi-fleet 4300 vehicles here. Over the day normally about 2500 vehicles are simultaneously on the road. They have a delivering interval of 30s.

Starting from the most central point in the city – the "Brandenburger Tor" – nine rectangle net parts from 3.5km to ~26km edge length were defined and the FCD in the different net parts compared. It is clear, that the taxis are not uniformly distributed around the city centre – the airports are the most important destinations for the taxis – Figure 8 shows some samples of the net parts.

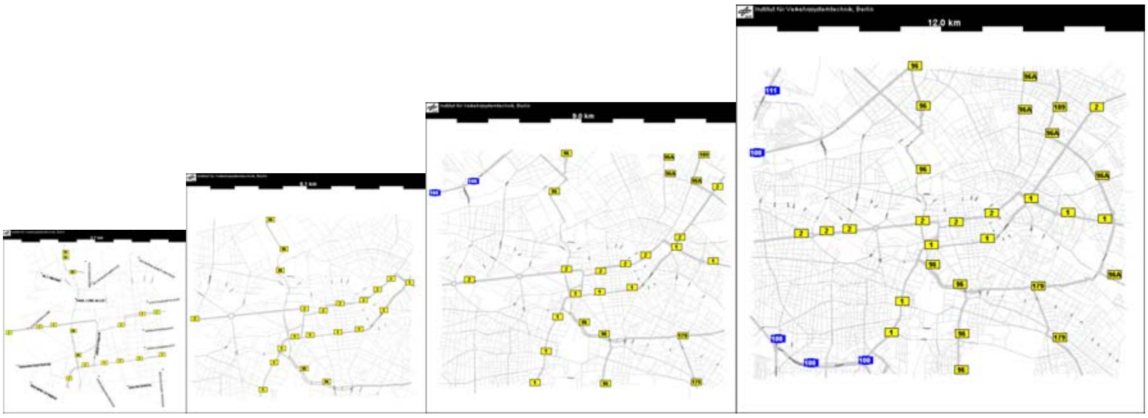

Figure 8: Different parts of the Berlin net (1-4)

For each net part the FCD for the 13th April 2010 was processed and the driven kilometre calculated for each Navtech® street-type. Figure 9 shows the relation of the driven km to the net length. This means that is approximately the mean vehicle count for the day.
Earlier investigations (Sohr, A. and Wagner, P. (2008)) showed that stable daily courses can be reconstructed from 100 data points. In the investigated sample case this would be possible for street type one and two for all net-sizes and for street-type 3 up to net-size 5. The next step will be the aggregation of the data for of month to improve the stability of the graphs.
Finally, to do some simple form of (online) traffic control the FCD data density have to be bigger still, at least every 10 minutes a new data-point is needed on the relevant links of the network (>360 data points per day). This would only be possible on the street-types up to size 3 for the road network.

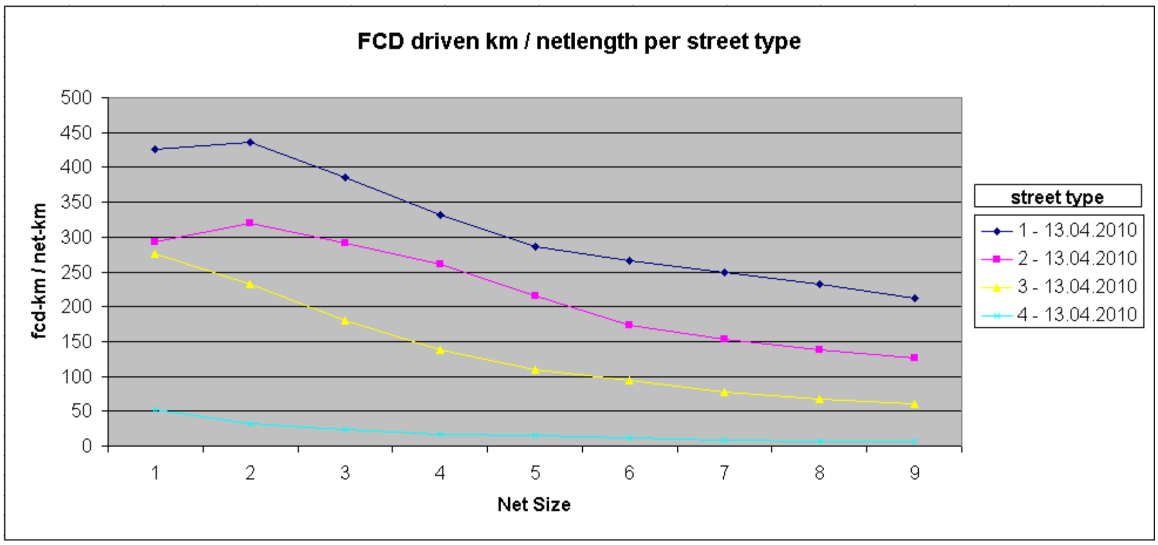

Figure 9 – approximated vehicle count for net-size and street-type

SUMMARY

There are many traffic related applications which use the derived traffic parameters. These applications have different complexity and need different data quality. For simplicity reasons they are grouped in four categories:

1. Travel Time Information on higher class road network
2. Offline signal control (assisted with historic data)(100)
3. travel time information for routes(24-7)(100+)
4. Online local signal control (360+)

As shown in the last chapter a delivering interval shorter than 60 seconds should be used. Considering the number of equipped vehicles it is necessary not only to look at the absolute number of vehicles, but also to take into account the daily mileage of each of the vehicles of fleet. If the FCD – fleets is a taxi-fleets, and since a taxi has more than one driver, so that every taxi is 24 hours of the day and 7 days of the week on the road and delivers data.

REFERENCES

Brockfeld, E. et al. (2008). Project ORINOKO (Operative Regionale Integrierte und Optimierte Korridorsteuerung) final report, p.117-173

Krieg, S. (2008). Simulation und Verkehrserfassung: Evaluierung von Floating Car Daten, Diploma Thesis, Brandenburg University of Applied Sciences (in German)

Sohr, A. and Wagner, P. (2008). Short Term Travel Time Prediction Using Floating Car Data Based On Cluster Analysis, 15th World Congress on ITS 2008 , New York (USA)