Evaluation of Floating Car Data quality for traffic congestion analysis on motorways

Bartosz Bursa[a], Wolfgang Blumthaler[a], Markus Mailer[a]

[a]*Unit for Intelligent Transport Systems, University of Innsbruck, Technikerstraße 13, 6020 Innsbruck, Austria*

Problem Statement

With the increasing number of data recording devices like mobile phones and on-board navigation systems, the availability of floating car data (FCD) is also on the rise. However, dynamic FCD is substantially different from classic stationary detector data (SDD). In addition, the data can be of mixed quality, which is subject to the penetration rate, type of recording device, data provider or purpose of data collection. The combination of these aspects restricts the usability of FCD for traffic state estimation and requires checks on data quality.

Research goals

In the current study, we exploit the FCD to analyze the traffic state on the A12 motorway in Tyrol, Austria. We identify congestion clusters within the dataset and classify these clusters into different congestion types. Previous results of a congestion cluster analysis based on SDD (Bursa et al., 2019) are used to validate the FCD-based results. By examining the differences in outcomes of both approaches, the aim is to identify criteria within the FCD set that can be used as indicators of its validity without the need for a second data source for cross-checking.

State-of-the-Art

Along with the increase of available FCD, the research on the use of this data source grows. In the study initiating this work, Transver GmbH (2010) validated the SDD-based macroscopic traffic classification using a synthetic FCD set generated specifically for this purpose. This type of data is of high and controlled quality, in contrast to the actual data usually available on the market.

For estimating FCD data quality, most studies concentrate on the penetration rate of probe vehicles within the traffic flow. Hong et al. (2007) claim that 2% level is sufficient for traffic state analysis, whereas Dai et al. (2003) argue for 3-5%.

A few researchers have already compared the FCD with other sources. Kessler et al. (2018) worked on a database relatively close to ours in their examination of jams on the German A9 near Munich. However, while their focus lies on the usability of different data sources for real-time traffic information and data delay, this work focuses on the robustness of the analysis depending on the FCD quality.

Methodological approach

The FCD set used in this work was provided by the German ADAC (General German Automobile Club) and consists of position points and velocities of an unspecified vehicle fleet registered in Tyrol in January 2016. From this set, the trips along the 85 km long section of the A12 motorway between Kufstein and Innsbruck were extracted. Around 10.000 trips are available in each driving direction, being equivalent to an average of two trips per 10-minutes-timeslot. In reality, most timeslots, especially at nighttime, contain no trips and the number per timeslot ranges from one to ten. The SDD dataset for the study was provided by the Austrian motorway administration ASFiNAG and comes from 53 detector stations.

Both the SDD- and FCD-based traffic state analysis utilizes the definitions developed by Transver GmbH (2010). The number of congestion types distinguished in this approach is limited to four:

- Short-term speed breakdown (type 1)
- Stop-and-go wave (type 2)
- Wide jam (type 3)
- Mega jam (type 4)

A detailed description of the classification algorithm can be found in Bursa et al. (2018).

The adaptive smoothing method (ASM) (Helbing and Treiber, 2002) was applied to the raw detector data and, based on macroscopic criteria, the congestion clusters were classified into the four aforementioned jam types.

The comparison between the two approaches consists in comparing positions of the congestion clusters. In effect, three types of results can be expected:

- Coinciding incidents, which indicate a traffic flow disturbance captured both by probe vehicles and by stationary detectors
- SDD only incidents, indicating a discrepancy between data sources or lacking FCD data
- FCD only incidents, indicating a discrepancy between data sources or more sensitive FCD data

In order to discern between these three types of outcomes, different attributes of the underlying FCD data were examined. The attributes with the highest correlation with the outcome were: congestion type, FCD quality and trip density.

The congestion type definitions used in our algorithm, based on duration and number of stops, are the main criterion differentiating between jam and non-jam trips. The distinction by jam duration leads to a correlation between congestion type and cluster extension. In return, the cluster composition gives an indication on the nature of the congestion.

The quality attributes of FCD are the average log interval and number of data points available. We found a distinct statistical relationship between congestion type and data quality. Congestions of type 2 on average had a higher point density and lower timestep size compared to type 3, with overall jam duration being similar. This effect is explained by the higher data density required to describe the stop-and-go pattern.

Penetration rate of probe vehicles is an often-used criterion for estimating validity of the FCD. However, full traffic flow at the time of FCD collection is needed to calculate this metric. Instead, we operate with an FCD-only parameter – density of probe vehicles.

Results

On the individual jam level, the congestion types 1 and 3 are vastly overrepresented and stand for over 90% of registered incidents. This indicates a higher prevalence of shorter congestion incidents. On the cluster level, the distribution of congestion types turns around. About 85% of congestion incidents cannot be clustered with other incidents. This mostly concerns type 1 jams, which occur sporadically and are not part of larger clusters. Another 10% are small clusters containing two to four incidents, and 5% are large clusters containing 5 or more incidents. These large clusters consist of 60% of type 3 jams and 70% of type 4 jams, therefore representing the big incidents well.

The comparison between FCD and SDD clusters provides a diverse picture. The large FCD clusters have a high overlap with the corresponding SDD clusters. All type 4 SDD clusters have an equivalent in the FCD set, as well as over 50% of the type 3 jams. The other way round, only 60% of the large FCD clusters have an SDD counterpart, and only 3% of small clusters. This indicates a higher sensitivity of FCD compared to SDD.

Examining the cluster attribute differences between matches and non-matches, the most notable ones are cluster size and composition. The number of jams within a cluster contribute to the match quality, with larger clusters having better SDD equivalents and all clusters with 13 or more jams having 100% match rate. In addition, clusters containing mega jams tally at the 80% level, while large clusters consisting of mostly type 1 jams have low matching rates.

In the available FCD set, penetration rates and probe vehicle densities are low. Over 70% of ten minutes timeslots contain no probe vehicles at all. The average penetration rate is close to 1%. The invalid jam clusters are situated in areas with probe vehicles density close to zero and penetration rates of 2‰. Most timeslots contain no FCD trips at all, while SDD data shows 50 and more vehicles passing. However, as soon as some probe vehicles enter the congestion zone, the analysis delivers satisfactory results, even at low penetration rates. Therefore, trip density instead of penetration rate can be used to assess whether clusters derived from FCD for a certain timeslot can be considered reliable.

To conclude, a mixed quality FCD set like ours is usable only for detecting major traffic incidents. Jams classified as types 2, 3 or 4 can be captured satisfactorily, whereas small incidents of type 1 are hard to validate since they can be caused by speed fluctuations of a single vehicle.

References

Bursa, B., Gajic, N. and Mailer, M. (2018). *Classification of traffic jams on alpine motorways*, Zenodo. 10.5281/ZENODO.1441016.

Bursa, B., Gajic, N. and Mailer, M. (2019) Insights into the congestion patterns on alpine motorways based on separate traffic lane analysis. *Transportation Research Procedia*, *37*, 441–448.

Dai, X., Ferman, M.A. and Roesser, R.P. (2003) A simulation evaluation of a real-time traffic information system using probe vehicles, *IEEE,* pp. 475–480.

Helbing, D. and Treiber, M. (2002) Reconstructing the Spatio-Temporal Traffic Dynamics from Stationary Detector Data. *Cooperative Transportation Dynamics*, *1*, 3.1–3.24.

Hong, J., Zhang, X., Wei, Z., Li, L. and Ren, Y. (2007) Spatial and Temporal Analysis of Probe Vehicle-based Sampling for Real-time Traffic Information System, *IEEE,* pp. 1234–1239.

Kessler, L., Huber, G., Kesting, A. and Bogenberger, K. (2018) Comparing Speed Data from Stationary Detectors Against Floating-Car Data. *IFAC-PapersOnLine*, *51*, 299–304.

Transver GmbH (2010). *Stauklassifizierung und Untersuchung des Zusammenhangs Verkehrsstärke und Verkehrszusammenbruch*. Munich.