Issue 20(4), 2020, pp. 22-37
https://doi.org/10.18757/ejtir.2020.20.4.5304

TU Delft OPEN

EJTIR

ISSN: 1567-7141
http://ejtir.tudelft.nl/

Influence of floating car data quality on congestion identification

Wolfgang Blumthaler[1]
Unit of Intelligent Transport Systems, University of Innsbruck, Austria.

Bartosz Bursa[2]
Unit of Intelligent Transport Systems, University of Innsbruck, Austria.

Markus Mailer[3]
Unit of Intelligent Transport Systems, University of Innsbruck, Austria.

This paper explores the usability of floating car data (FCD) of mixed quality in congestion analysis on motorways. The specific data quality aspects that we are investigating are the number and density of trajectories, the GPS interval, and the fleet representativeness. We use a dataset provided by the German Automobile Club ADAC covering the Tyrolean road network in 2016. From this dataset, trajectories along the A12 motorway were extracted for congestion analysis. These data are characterized by high GPS time interval, low number of trajectories, and are not representative for total traffic due to overrepresentation of trucks. The influence of these quality parameters on congestion identification is explored by analyzing the parameter distribution among different congestion types. In addition, we validate the results by comparing them with congestion incidents obtained from the stationary detector data (SDD) and examining the impact of quality parameters on the validation results. We find that the given data set does not allow short-term congestion patterns to be identified due to quality flaws. Especially the low number of trajectories proved problematic, whereas the influence of other parameters was less distinct. Despite these flaws, for large-scale congestion incidents, floating car data provide outcomes similar to those derived from stationary detectors.

Keywords: *classification of traffic jams, congestion analysis, floating car data, probe vehicle data, traffic data quality, traffic patterns.*

[1] A: Technikerstraße 13, 6020 Innsbruck, Austria T: +43 512 507-62405 F: +43 512 507-62498 E: wolfgang.blumthaler@student.uibk.ac.at
[2] A: Technikerstraße 13, 6020 Innsbruck, Austria T: +43 512 507-62405 F: +43 512 507-62498 E: bartosz.bursa@uibk.ac.at
[3] A: Technikerstraße 13, 6020 Innsbruck, Austria T: +43 512 507-62405 F: +43 512 507-62498 E: markus.mailer@uibk.ac.at

1. Introduction

Driving vehicles that can serve as sources of traffic data, called floating car data (FCD) or probe vehicle data, have become an important form of traffic information technology in the recent years. This development was possible thanks to an increasing number of vehicles equipped with navigation devices and a technological progress in data transmission. Future innovations are expected to further strengthen the availability and quality of FCD as a source of traffic information.

In contrast to classic data generation methods based on stationary detectors (SDD), floating car data are available without investments in roadside infrastructure. This advantage however is outweighed by the fact that FCD register only a sample of all vehicles, raising doubts about the representativeness of the information collected. In addition, floating car data can differ greatly in scope and level of detail. Furthermore, gaining access to FCD for practical or scientific applications comes at a cost, which can be considerable for high-quality datasets, whilst some providers may offer data of lower quality but at a more reasonable price.

An important aspect of FCD quality is that different applications require different data quality levels. An FCD set of lower quality that is not suitable for one application, might still provide valuable information for another one. Therefore, data quality should always be considered in the context of the specific purpose for which the data are needed.

To fill this gap and expand on the relation between FCD quality and the intended application, we raise the following questions in this paper:

1. How do the quality attributes restrict possible applications of FCD to traffic congestion analysis?
2. What information on congested states can be derived from FCD depending on FCD quality?

2. State of the art

With the increasing availability of position data from navigation devices and mobile phones, floating car data (FCD) have become an established source of traffic information used for many purposes. An overview of datasets reported in scientific studies shows the diversity of data sources and providers used nowadays.

Kessler et al. (2018) use pre-processed data provided by the Dutch company TomTom to compare congestion analysis results between FCD and SDD. Ebendt et al. (2010) use FCD from a taxi fleet in Berlin in combination with a truck fleet to optimize urban routing. Houbraken et al. (2018) use FCD provided by the Dutch road operator DWS to test dynamic traffic management, whereas Altinasi et al. (2016) use data provided by the company ISSD to analyze traffic patterns on roads in Ankara, Turkey. Transver (2010) employed a small number of probe vehicles to collect FCD to evaluate congestion patterns.

2.1 *Data quality*

A number of publications discuss the quality of movement data, in particular FCD, and their impact on different applications. Andrienko et al. (2016) discuss this topic in a broad sense of movement data and provide a categorization of data quality aspects by defining four sets of properties. Mover set properties cover all aspects of the movers observed, especially their population coverage and representativeness. Spatial properties include spatial resolution, exactness, and coverage. Temporal resolution, exactness, and coverage are grouped into temporal properties. Data collection properties cover all aspects concerning the data collection method.

The majority of the FCD quality parameters reported in the research belongs to one of these categories. The most commonly discussed are the penetration rate of probe vehicles in the total traffic and the temporal resolution (time interval) of the GPS data points. Both parameters are associated with the expenditure on data generation (Leduc, 2008; Vandenberghe et al., 2012). As far as the penetration rate is concerned, equipping a high number of vehicles with GPS devices, or gaining access to large datasets of different providers, involves considerable financial effort. Regarding the sampling rate, it is limited by the cost of data transmission, bandwidth capacity and storage capacity of the device (Zhou et al., 2016). To reduce these costs, the position data generated by probe vehicles are transmitted at intervals that vary depending on the provider and the intended use cases. Sohr et al. (2010) study the combined impact of these two parameters on the relative error between the measured speed and the real speed based on simulated data. They conclude that the penetration rate is more impactful than the GPS interval, and suggest using penetration rates above 0.5% and a GPS interval below 60 seconds for reliable results.

Vandenberghe et al. (2012) explore the combined influence of these two parameters on congestion identification in different simulated environments. They find that choice of the road environment has a high impact on the required parameters: for motorways, a minimum penetration rate of 1% and a GPS interval of 20 seconds are suggested, while for urban networks, higher penetration rates and a lower GPS interval are necessary to correctly identify congested states.

A number of other publications evaluate GPS interval and penetration rate parameters separately and from different perspectives, and thus obtain differing results. Ranacher et al. (2016) analyze the proper sampling rate to minimize the effects of measurement and interpolation errors for the (mostly) urban road network around Salzburg, Austria. For vehicle speed analysis, they claim that an optimal interval is 1-2 seconds, whereas for distance measurements it is 6-12 seconds.

As far as the penetration rate is concerned, the reported values range from 1% (Dion and Rakha, 2006) to 4-5% (Nanthawichit et al., 2003). Breitenberger et al. (2004) recommend a minimum of 2.4%. For urban road networks, higher penetration rates up to 10% are suggested.

The parameter of fleet representativeness is less discussed in literature. Cohn et al. (2012) point out that regarding the FCD obtained from TomTom navigation devices, the population of TomTom users differs from the general population. Women and seniors use navigation devices significantly less than other demographic groups. In addition, navigation devices are not used for all trip purposes equally. They are more often used for trips in unknown environments and less for routine trips.

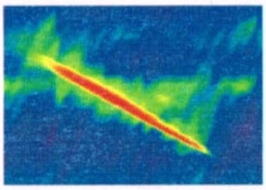
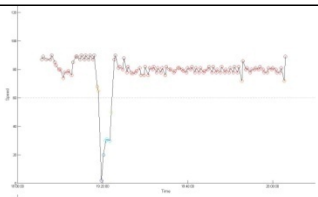
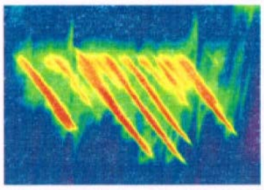
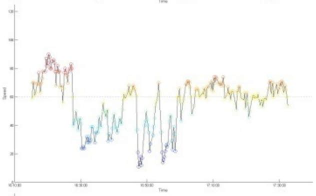
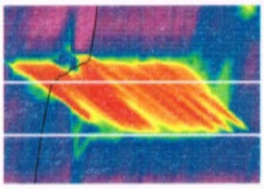
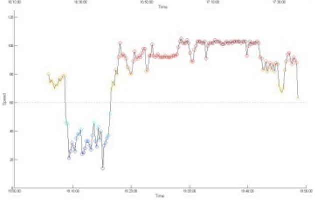
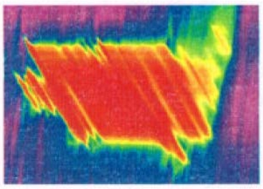
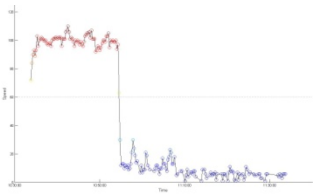
Keler (2017) discusses the influence of the data provider and compares the results based on FCD from a taxi fleet and from private vehicles. They state that taxis have different and less periodic movement patterns compared with the commute patterns of private vehicle users. Moreover, taxi drivers possessing extensive knowledge of the road network, often use less known routes to avoid congested areas.

2.2 *Congestion analysis based on FCD*

A number of studies use FCD to analyze traffic patterns and congested states. Most approaches utilize vehicle speed as a threshold parameter categorizing traffic as congested or not. Houbraken et al. (2018) use FCD for automated incident detection and dynamic traffic management based on velocities on road segments in the Netherlands, while Altinasi et al. (2016) use a speed-based level of service for road segments to identify urban traffic patterns in Ankara, Turkey. Bazzani et al. (2011) reconstruct empirical fundamental diagrams for roads in Rome, Italy based on FCD collected by 2% of the total number of vehicles. Kessler et al. (2018) compare how congested states can be identified in real time using FCD and SDD.

The consultancy Transver developed a congestion classification algorithm based on vehicle speed both for microscopic FCD and macroscopic SDD (Transver, 2010). Their classification aimed at combining congestion with the impact it has on energy management in electric vehicles and thus employed two jam characteristics: duration and number of stops, which are critical for efficient operation of electric vehicle engines (Bursa et al., 2018). The main criterion classifying the state of traffic as congested is vehicle speed below 60 km/h, which is maintained for an extended period. Further classification is based on congestion duration and driving patterns during the congested drive. Table 1 illustrates the distinguished congestion classes, their spatiotemporal visualization based on stationary detector data, an example FCD speed-time graph, and the defining characteristics. Transver (2010) used SDD from the German A9 motorway and FCD set with the GPS interval of one second produced by test vehicles used specifically to validate these definitions.

Table 1. Congestion classes

Type of congestion	Spatiotemporal SDD graphic example	Speed-time FCD graphic example	Defining characteristics
Type 1: Single congestion wave / short-speed-drop			Duration below 3 min
Type 2: Stop-and-go wave			Duration above 3 and below 30 min. More than 3 distinct breaking and acceleration patterns.
Type 3: Wide jam			Duration above 3 and below 30 min. 3 or fewer distinct breaking and acceleration patterns.
Type 4: Mega jam			Duration above 30 minutes

Based on the congestion definitions provided by Transver (2010), Bursa et al. (2018) analyzed congestion patterns on motorways in Tyrol, Austria using detector data from the Austrian motorway operator ASFiNAG. As long as they achieved plausible results in the identification of the congestion types short-speed-drop and mega jam along the A12 motorway, the differentiation between stop-and-go waves and wide jams proved challenging, as both classes have identical duration parameters and the identification of stop-and-go patterns was possible only visually. In addition, classification on the A13 motorway turned out to be problematic. The reason being that the high inclination brings the average velocities of trucks down to below 60 km/h even in non-congested traffic state. Bursa et al. (2018) suggest performing an FCD-based analysis for the same road segments to gain additional information on traffic patterns and solve the problems they encountered.

3. Methodology

3.1 Data set source and range

The FCD set evaluated in this study was provided by the German automobile club ADAC and consists of data generated over the year 2016. The dataset is a subsample of a larger dataset collected by ADAC and describes the movements of over 1.2 million vehicles travelling throughout the Tyrolean road network. The data consist of single data points collected by individual vehicles and include total of 24 parameters per data point. From among these parameters, the vehicle ID, vehicle class, timestamp, position (longitude and latitude), and vehicle speed were used.

The transformation of data points to trajectories was performed using vehicle IDs. This parameter assigned each vehicle a unique random identification number. At midnight, each vehicle was assigned a new ID, preventing tracking of individual vehicles across an extended period.

For the data analysis, only movements along the A12 motorway between the Austrian-German border near Kufstein (lat. 47.605 | lon. 12.193) and the Innsbruck motorway interchange (lat. 47.255 | lon. 11.420) in both driving directions were considered, which totals to 121,324 registered vehicle movements: 57,314 in westbound direction and 64,010 in eastbound direction. The geometry of the observed road segment simplified the map-matching process, as there was only one possible path connecting two data points along the motorway.

The attributes chosen for the evaluation of data quality were the density of trips (trajectories), GPS interval and fleet representativeness. The first two parameters were chosen since they are widely discussed by other researchers. Fleet representativeness was added as a parameter because of the specific nature of the given dataset as discussed below.

Trip density: the density of recorded FCD trajectories was used to describe the amount of information available at any given time and was calculated as the number of probe vehicles present per hour. This value is closely related to probe vehicle penetration rate, which is typically defined as percentage share of FCD trajectories in the total number of vehicles. However, calculation of a penetration rate requires knowledge of the total traffic, which must be collected from other sources (e.g. stationary detectors). Therefore, trip density was used as the preferred parameter.

GPS interval: the time difference between subsequent points of FCD trajectories was used to describe the temporal resolution of the individual trajectory data. Short GPS time intervals facilitate the identification of short disturbances in vehicle speed.

Fleet representativeness: to evaluate how closely the probe vehicle fleet represents the total traffic on the A12 motorway, the vehicle type composition and the spatial and temporal trip distribution were compared with the data from stationary detectors.

The overall quality of the present data set was assessed based on the suggested values for each parameter discussed in the literature and described in section 2.1. A dataset reaching the suggested thresholds for all parameters would be described as a set of high quality, whereas a dataset missing the thresholds for all parameters would be described as a low-quality dataset. A dataset meeting the thresholds for some parameters, while missing for others would be described as a mixed-quality dataset.

3.2 Congestion identification and classification

For the identification of congestion patterns, the framework developed by Transver (2010) presented in section 2.2 was chosen. The reasoning for this choice is that Transver created this framework and tested it with both SSD and FCD. This has the advantage that the results from the current FCD analysis can be compared with the results that Bursa et al. (2018) obtained using the same framework in the same survey area and period. The disadvantage of this approach is that the definitions of speed threshold or pattern duration could not be changed without losing comparability between the datasets.

The congestion identification process involved selecting the trajectories with segments where driving speed below 60 km/h was recorded. Afterwards, the duration of the low-speed segment and the acceleration pattern were used to classify the congested state according to the algorithm depicted in Figure 1.

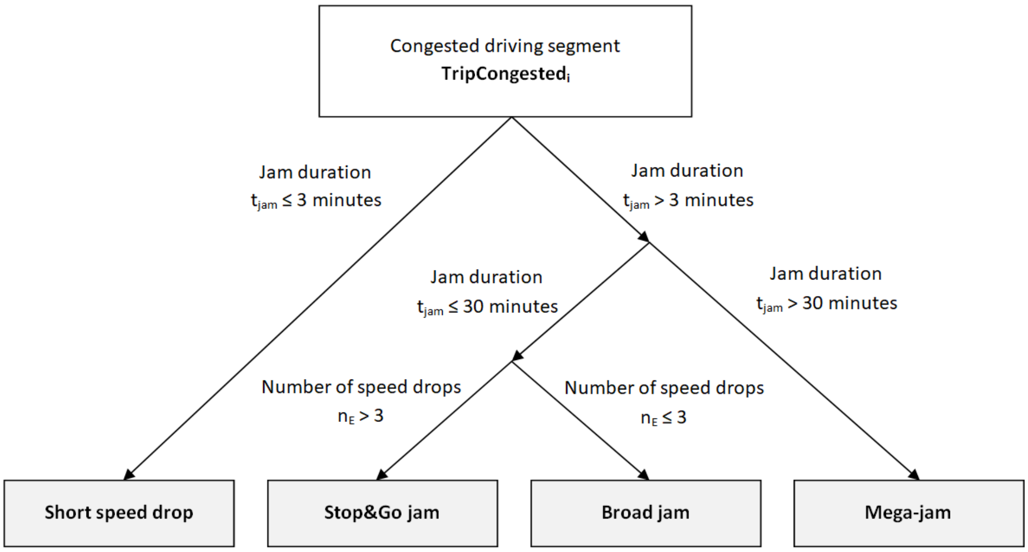


Figure 1. Classification of FCD trajectories based on definitions provided by Transver (2010).

As a primary assessment of the impact of data quality on congestion identification and classification, we grouped the FCD trajectories into those showing free-flowing traffic and those showing a congestion of some type. Then, we analyzed the relationship between the mean values of FCD quality parameters and the number and types of congested states identified in the FCD.

3.3 Validation using stationary detector data

To assess the validity of congested traffic states, we compared them with congestion clusters based on stationary detector data as reported by Bursa et al. (2018).

At the level of single congestion incidents, we calculated the number of FCD trajectories passing through the spatiotemporal areas of the SDD congestion clusters. At least one FCD trajectory classified as congested was necessary to confirm the SDD-based congestion. No FCD trajectories crossing the cluster or an uncongested state of these trajectories resulted in a cluster classified as unconfirmed. We then analyzed the interdependencies between the number of trajectories, congestion classes, and the confirmation rate between SDD and FCD.

In addition, we also compared the results from both datasets at the level of spatial and temporal distribution of congestion incidents over the entire A12 motorway. This comparison indicated how well both data sources describe the traffic situation along the A12 motorway, and whether using only FCD would provide acceptable results.

4. Results

4.1 Data set characteristics and representativeness

The dataset used in this study contained all observed FCD trips along the A12 motorway in 2016, which amounted to 121,324 individual trajectories. The mean values of the quality parameters are listed in Table 2.

Table 2. Parameter values of the FCD set (mean value and standard deviation)

	Number of trips	GPS interval [s]	Trip density [trips per hour]	Daytime trip density (04:00 – 22:00) [trips per hour]	Nighttime trip density (22:00 – 04:00) [trips per hour]
Total	121,324	15.106 ± 13.032	6.91 ± 5.36	7.75 ± 4.69	0.83 ± 0.82
Westbound direction	57,314	14.888 ± 12.928	6.53 ± 5.42	7.24 ± 5.90	0.99 ± 1.21
Eastbound direction	64,010	15.298 ± 13.119	7.29 ± 6.36	8.26 ± 5.76	0.67 ± 1.01

The number of trajectories in eastbound direction was 10% higher than the number of trajectories in westbound direction. GPS intervals were similar between driving directions, the high standard deviation of 13 seconds compared to the mean of 15 seconds indicates a wide dispersion of the GPS interval values.

The average trip density of seven trips per hour equaled one probe vehicle every nine minutes. The trip density parameter exhibits a high standard deviation too, indicating that we are dealing both with periods with very few trips and with periods with many trips. This effect becomes visible when splitting the day into a daytime slot between 04:00 and 22:00 and a nighttime slot between 22:00 and 04:00. During daytime, trip density averaged seven to eight trips per hour; during nighttime, only one trip per hour occurred on average. Employing this split also lowered standard deviation, implying that the high values observed without discerning between day and night were mostly based on the daily pattern. The trip density values were higher for eastbound trips in the daytime, which corresponded with the higher number of registered trips.

The average penetration rates of probe vehicles were calculated using total traffic values from stationary detectors. For many timeslots, no probe vehicles were registered, yielding a penetration rate equal to zero. During peak hours, total traffic reached 3,000 to 5,000 vehicles per hour, compared to 10 to 14 FCD trajectories per hour. This led to an average penetration rate of probe vehicles equal to 0.08% and reaching maximum values of 0.5% in selected timeslots.

Table 3. Vehicle classification distribution

	Passenger vehicles	Vans	Trucks	Unknown
FCD data set	9.4%	0.3%	50.3%	40%
SDD data set	83.6%	3.1%	13.3%	

To assess how well the given FCD set represented total traffic registered on the A12 motorway by stationary detectors, the vehicle type compositions were compared as shown in Table 3. The results from the two data sets differed considerably. 40% of FCD trajectories were classified as unknown. While trucks were overrepresented in the FCD set (50.3%) compared to the SDD set (13.3%), passenger cars were underrepresented (9.4%) compared to their share in the SDD (83.6%). The number of vans was very low in both data sets.

As an additional indicator of representativeness, we compared the temporal distribution of FCD trajectories with the temporal traffic distribution registered by stationary detectors. Figure 2 reveals that both data sources exhibit a distinct day-night pattern. Despite the differences in vehicle type composition, both data sources provided a similar temporal coverage of traffic, and the fluctuation of probe vehicles corresponded with the fluctuation of total traffic.

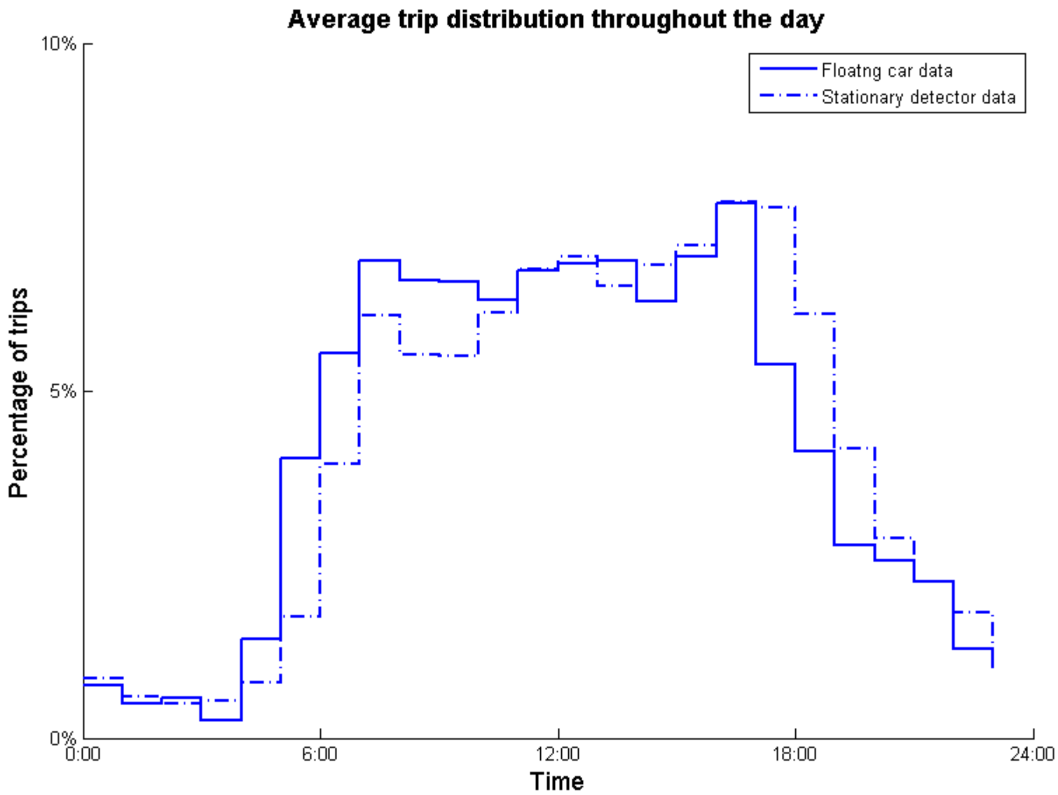


Figure 2. Daily distribution of FCD trips and SDD traffic

4.2 Parameter values for FCD-based congestion classes

The primary step in the data quality assessment was to evaluate quality parameter values for different congestion classes. Table 4 summarizes the GPS interval, trip density and vehicle type distribution for the entire FCD set as well as for each congestion class separately.

The jam duration values were consistent with the definition provided by Transver, but, despite the same duration thresholds, stop-and-go waves had longer average durations than broad jams.

The congestion types short-speed-drop and stop-and-go wave exhibited similar patterns in terms of GPS interval and vehicle classification. These two types had a lower GPS interval than broad and mega jams. Passenger cars were overrepresented with a share of about 20% compared to other congestion classes and to the entire FCD set.

In terms of trip density, stop-and-go waves and mega jams were characterized by values above the average for the complete dataset, whilst for short-speed-drops and broad jams, values below the average were reported.

Table 4. Average parameter values for different congestion classes

	Short-speed-drop	Stop-and-go wave	Broad jam	Mega jam	Total FCD set
Jam duration [min]	1.1	16.3	9	38.7	-
GPS interval [s]	19.1	17.4	24	24.9	15.1
Trip density [trips/hour]	4.8	6.9	5.4	7.8	6.5
Vehicle types [%] [passenger car | van | truck | unknown]	18|0|56|26	22|0|38|40	7|0|53|40	7|0|52|41	9|0|50|40

4.3 Influence of the parameters on validation results

Individual trip validation
Evaluation of the FCD validity by comparing SDD-based congestion clusters from Bursa et al. (2018) with congested FCD trajectories revealed differences between congestion classes as summarized in Table 5. An SDD congestion cluster was deemed confirmed if at least one FCD trajectory passing through the cluster was classified as congested. Across all classes, only 56% of SDD congestion clusters could be confirmed by FCD. The short-speed-drop congestion class had the lowest confirmation rate of only 8.1%, while the mega jams reached the highest rate of 83.2%. The confirmation rates of stop-and-go waves and broad jams were in-between, with higher values for stop-and-go waves.

Table 5. Congestion validation results

	Short-speed-drop	Stop-and-go wave	Broad jam	Mega jam
Number of SDD clusters	37	56	195	304
Number of SDD clusters confirmed by FCD trajectories	3 (8.1%)	28 (50%)	48 (24.6%)	253 (83.2%)
Average number of FCD trajectories per confirmed cluster	1	3.4	3.2	15.3
Number of SDD clusters not confirmed by FCD trajectories	34 (91.9%)	28 (50%)	147 (75.4%)	51 (16.8%)
Average number of FCD trajectories per unconfirmed cluster	0.03	0.75	1.3	4.4

These results can be associated with the number of FCD trajectories passing any individual SDD cluster. Table 5 lists the number of FCD trajectories passing through confirmed and unconfirmed SDD clusters. For all congestion classes, the number of trajectories within the confirmed clusters was higher than within the unconfirmed clusters. For example, for short-speed-drops, out of 37 congestion clusters identified by SDD, three were confirmed by FCD, while 34 could not be confirmed. All three confirmed clusters were passed through by exactly one FCD trajectory, resulting in an average of one trajectory per cluster. In contrast, out of the 34 unconfirmed clusters, only a single one was passed through by an FCD trajectory, resulting in an average of 0.03 trajectories per cluster. Congestion classes of longer duration show a higher average number of FCD trajectories passing through as well as higher confirmation rates.

We employed a logistic regression model to investigate how strongly the number of trajectories (i.e. FCD quality) within a cluster influenced whether the FCD could detect congestion in a cluster where it was detected by SDD. Assuming the logit link function (i.e. logistic distribution of the error terms) in a binary model, the probability P of the SDD congestion incident being detected ($Y = 1$) by FCD is given by:

$$P(Y = 1) = \frac{e^{x_i{}^T \beta}}{1+e^{x_i{}^T \beta}} \tag{1}$$

In our case:

$$x_i{}^T \beta = \beta_0 + \beta_1 \cdot x_1 \tag{2}$$

where β_0 is a constant (intercept), x_1 is the predictor variable represented by the number of trajectories and β_1 is the coefficient attached to that predictor.

The binary logit model was estimated using the *stats* package in the statistical software R (R Core Team, 2019). The general model shows a very significant positive correlation between the number of trajectories and congestion found by both SDD and FCD (Table 6).

Table 6. Estimation results of the general model

Predictor	Estimate	Standard Error	t-statistic	Significance
(intercept)	-0.9143	0.2013	-4.542	***
Number of trajectories	0.2839	0.0328	8.658	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1;
Log-Likelihood = -182.7
R-squared = 0.3135

The second model interacting the number of trajectories with different types of congestion in the SDD cluster confirmed this finding (Table 7).

Table 7. Estimation results of the model with interactions

Predictor	Estimate	Standard Error	t-statistic	Significance
(intercept)	1.0986	1.1547	0.951	
Type: stop-and-go	-1.4172	1.4921	-0.950	
Type: wide jam	-2.3389	1.2124	-1.929	.
Type: mega jam	-2.1602	1.2611	-1.713	.
Stop-and-go : number of trajectories	0.5250	0.3420	1.535	
Wide jam : number of trajectories	0.2698	0.1148	2.351	*
Mega jam : number of trajectories	0.2968	0.0538	5.514	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1;
Log-Likelihood = -173.3
R-squared = 0.3622

Due to a singularity issue, it was not possible to estimate the coefficient for the short-speed-drop class (only one trajectory in our FCD set detected congestion within an SDD cluster of that congestion class). For all other classes, one can see a positive correlation between the number of FCD trajectories within an SDD cluster and the odds of confirming a congested state in this SDD cluster by at least one FCD trajectory. The coefficients are significant for wide and mega jams and are not significant for stop-and-go jams. This is again in line with the general observation that the short and more dynamic congestion classes suffer from problems with vehicle composition, definitions of the congestion types and low share in the FCD set. The model outcomes in Table 7 informs us that if there was one additional trajectory, the log odds that a congestion in an SDD cluster will be confirmed by FCD would increase by a factor of 0.270 for a wide jam and 0.297 for a mega jam. That is, the odds that a congestion will be confirmed would increase by an exponent of the above, in this case $exp(0.270) = 1.31$ and $exp(0.297) = 1.35$ respectively.

Overall distribution validation

To assess the overall quality of congestion identification based on the FCD analysis, we compared the distribution of congestion incidents with the results obtained by Bursa et al. (2018) from stationary detectors.

The temporal distribution of congestion incidents over the entire year revealed similar patterns for both data sources, as depicted in Figure 3. Some months were characterized by a higher discrepancy between FCD and SDD, e.g. the month of April in the westbound direction. In both data sources, we found higher levels of congestion during summer; in the case of the eastbound direction, also during winter; with lower levels of congestion in spring and fall.

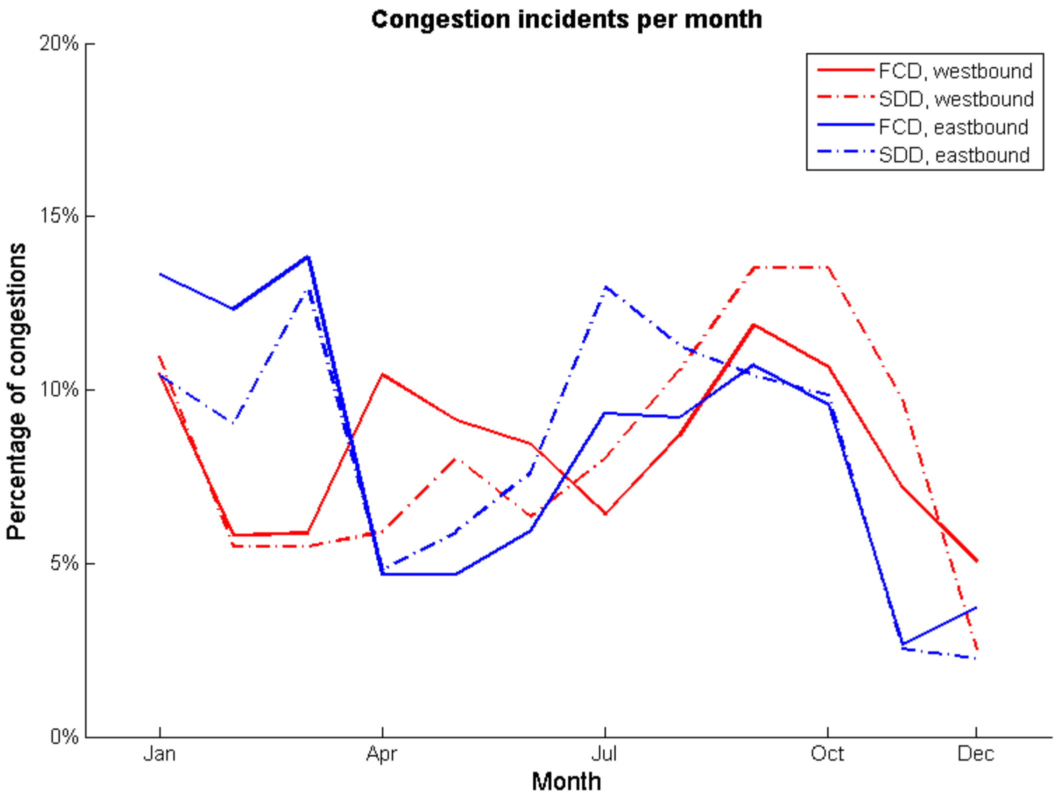


Figure 3. *Congestion distribution over the year*

The spatial congestion distribution correlated well between both data sources as shown in Figure 4, with high congestion on the motorway segments close to the German border in eastbound direction and on the segments close to Innsbruck in westbound direction. A significant difference could be noticed on the segments close to Innsbruck in eastbound direction.

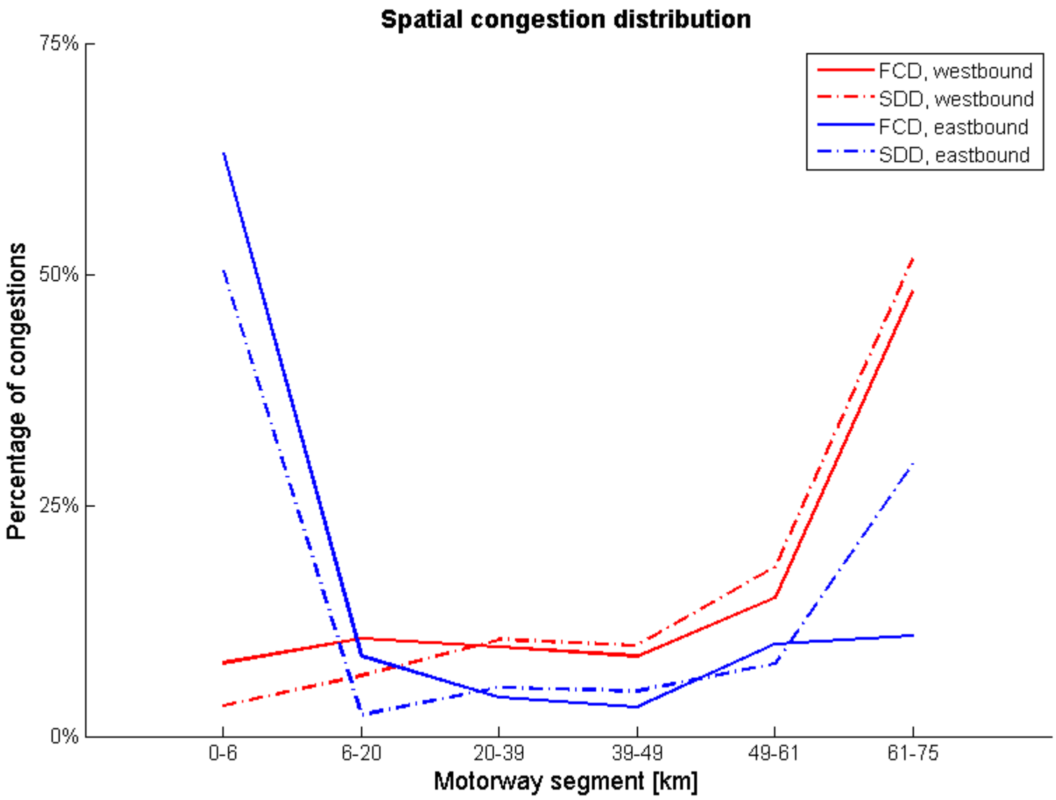


Figure 4. *Spatial congestion distribution*

5. Conclusion and discussion

The quality parameter analysis of the given FCD set exhibited a number of substantial imperfections and parameter values below those suggested in the literature. The values of trip density in particular, and hence also the probe vehicle penetration rate, were very low. Compared with the penetration rate values between 0.5% (Sohr et al., 2010) and 5% (Nanthawichit et al., 2003) suggested in the literature, the observed penetration rate reached 0.5% only in selected timeslots with an average value of 0.08%, which is considerably lower. In addition, the vehicle class composition of the dataset differed significantly from the vehicle class composition of total traffic. However, despite this unrepresentative vehicle fleet, the temporal distribution of FCD trajectories for a single day tallied well with the traffic distribution obtained from SDD. The GPS interval of 15 seconds was in accordance with the values of 20 seconds (Vandenberghe et al., 2012) or 60 seconds (Sohr et al., 2010) mentioned in the literature.

These different imperfections impose limitations on possible applications of the FCD to congestion analyses since the results might be skewed in different ways. For example, the congestion classes of short-speed-drop and stop-and-go wave are characterized by short-term changes in vehicle speeds. Such driving patterns are characteristic of passenger cars, capable of accelerating and decelerating more dynamically compared to trucks. The low overall percentage of passenger cars among probe vehicles in our FCD set (considerably lower than the average in total traffic) could therefore have led to an underestimation of short-speed-drops and stop-and-go waves.

In addition, traffic patterns characterized by short-term speed changes require low GPS intervals (high sampling rate) in FCD trajectories to be properly recognized. Average GPS interval values for the congestion classes short-speed-drop and stop-and-go wave were lower than for other congestion classes, but still not low enough to identify these congestion patterns reliably. Therefore, the expectation of Bursa et al. (2018) that using FCD could prove useful to discern stop-and-go waves from broad jams could not be fulfilled with the present data set. A much higher temporal resolution, closer to the value of one second used by Transver (Transver GmbH, 2010), would have been required to reconstruct the acceleration and deceleration patterns characteristic for these congestion classes.

The validity of congestion identification based on FCD was influenced heavily by trip density, i.e. the number of probe vehicles providing information. Confirming small-scale forms of congestion like short–speed drops registered in the SDD by means of the present FCD set was hardly possible because the chances of one of the few probe vehicles passing through a short und non-homogenous congestion cluster were very low.

In contrast to these limitations, the FCD was well suited to identify and describe large-scale congested states that have no abrupt speed changes. Since such incidents dominate the traffic situation along the entire A12 motorway, the present FCD set provided overall results similar to the SDD-based analysis by Bursa et al. (2018). This demonstrates the power of floating car data as a source of traffic information. In accordance with Vandenberghe et al. (2012), who suggested different minimum thresholds of quality attributes for different applications, it also confirms that the choice of application has a high influence on the required data quality. The present FCD set was not capable of identifying short-term congestion classes such as short-speed-drops reliably. Neither was it capable of providing insight into acceleration patterns inside the stop-and-go waves. Nevertheless, it was well suited to evaluate the overall traffic flow regime along the A12 motorway.

Outlook

As far as the methodological approach is concerned, the combined use of low-quality FCD and fixed congestion definitions (originally created for high-quality FCD) proved difficult. Redesigning the congestion classification algorithm, particularly for the stop-and-go class, could provide more valuable insights, even with the present dataset. Especially the 60 km/h speed threshold appears to be too high for alpine motorways characterized by a speed limit of 100 km/h and an average speed of trucks well below 80 km/h.

To estimate this effect we evaluated the impact of variations of the speed parameter on congestion identification based on the data from January 2016. Figure 5 illustrates how much, in percent, the number of detected congestion classes changes for different speed thresholds: 30 km/h, 40 km/h and 50 km/h. The reference speed is 60 km/h (set to 100% in the diagram). While the total number of identified congestion incidents decreases with a stricter speed threshold, the number of Stop-and-Go waves increases considerably. This indicates that the acceleration and deceleration patterns of this congestion type occur mostly below 60 km/h. The overall spatial and temporal incident distribution remains the same for all speed thresholds. Adjusting the speed threshold reveals more details on the Stop-and-Go waves but also deviates from the original classification framework, which makes the results less comparable with those of Bursa et al. (2018).

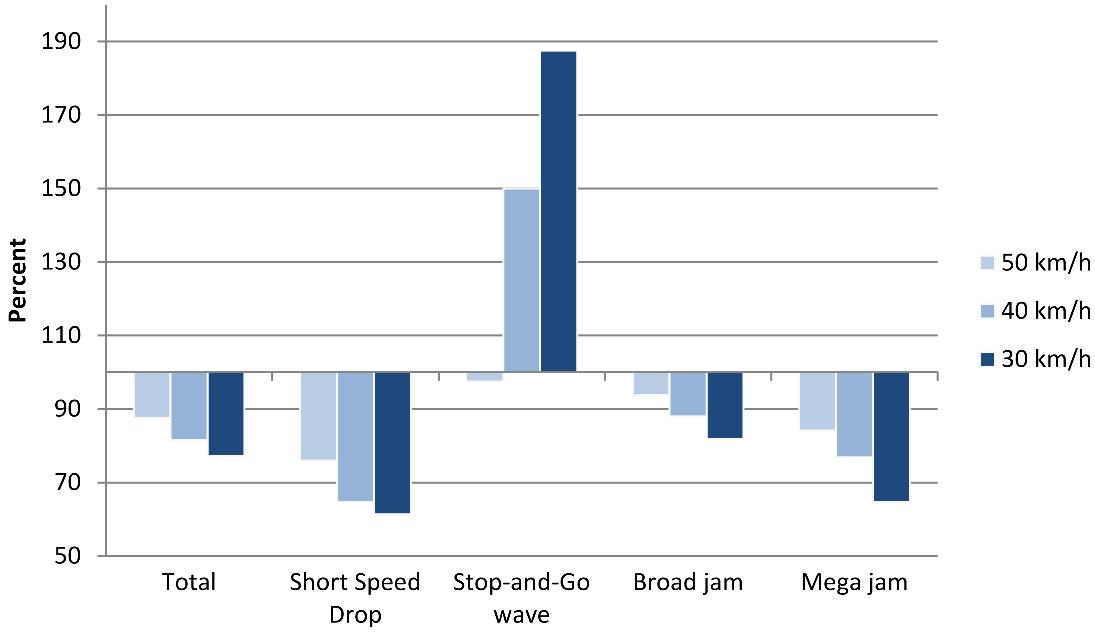


Figure 5. Percentage change in detected congestion classes depending on the speed threshold. (the reference threshold of 60 km/h equals 100%).

Another aspect worth revisiting is that the definitions of congestion classes used in the study assume the same duration thresholds (3 min, 30 min) for both microscopic and macroscopic approaches. This becomes problematic when the total duration of a congestion incident and the duration experienced by individual vehicles differ, leading to a different classification, e.g. the cluster is over 30 minutes, but individual vehicles pass the congested section in only 20 minutes, which results in different classification. This is particularly visible when attempting to confirm the SDD-based clusters using FCD trajectories. While the simple confirmation of a congested state presented in section 4.3 worked well, an attempted comparison of how this congested state was classified in both approaches resulted in high discrepancies, even for clusters with many trajectories. Therefore, direct matching of congestion classes between SDD and FCD proved unfeasible using the current definition framework.

Another issue with results validation using the preprocessed SDD clusters is that these clusters take a form of rectangles circumscribing the spatiotemporal extent of the congestion incident. These rectangles may include a lot of uncongested traffic (and uncongested FCD trajectories), further complicating the matching process.

Based on the limitations of the ADAC dataset described in this paper, a number of additional applications for the FCD set are possible. For instance, a similar analysis of congestion patterns with adjusted definitions could be used along the A13 motorway to avoid the problems with low-speed trucks that Bursa et al. (2019) encountered. Another potential research topic could be evaluation of the usefulness of low-quality FCD for traffic analysis on urban road networks.

Acknowledgements

We would like to thank the German Automobile Club ADAC for providing the floating car data set and the Austrian motorway operator ASFINAG for providing the stationary detector data for the A12 motorway.

References

Altintasi, O., Tuydes-Yaman, H. and Tuncay, K. (2017) Detection of urban traffic patterns from Floating Car Data (FCD). *Transportation Research Procedia*, 22, 382–391. https://doi.org/10.1016/J.TRPRO.2017.03.057.

Andrienko, G., Andrienko, N. and Fuchs, G. (2016) Understanding movement data quality. *Journal of Location Based Services*, 10(1), 31–46. https://doi.org/10.1080/17489725.2016.1169322.

Bazzani, A., Giorgini, B., Gallotti, R., Giovannini, L., Marchioni, M. and Rambaldi, S. (2011) Towards Congestion Detection in Transportation Networks Using GPS Data. Paper presented at *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing.* Boston, MA, pp. 1455–1459. https://doi.org/10.1109/PASSAT/SocialCom.2011.249.

Breitenberger, S., Grüber, B., Neuherz, M. and Kates, R. (2004) Traffic information potential and necessary penetration rates. *Traffic engineering & control*, 45(11), 396–401.

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. https://doi.org/10.1016/j.trpro.2018.12.220.

Bursa, B., Gajic, N. and Mailer, M. (2018) Classification of traffic jams on alpine motorways. In: *Proceedings of 7th Transport Research Arena TRA 2018,* Vienna, Austria. https://doi.org/10.5281/zenodo.1441017.

Cohn, N. and Bischoff, H. (2012) Floating car data for transportation planning: Explorative study to technique and applications and sample properties of GPS data. Paper presented at *National Travel Monitoring Exposition and Conference (NaTMEC).* Transportation Research Board, 2012. http://onlinepubs.trb.org/onlinepubs/conferences/2012/NATMEC/Cohn.pdf.

Dion, F. and Rakha, H. (2006) Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation research. Part B, Methodological.*, 40(9), 745–766. https://doi.org/10.1016/j.trb.2005.10.002.

Ebendt, R., Sohr, A., Tcheumadjeu, L. and Wagner, P. (2010) Utilizing historical and current travel times based on floating car data for management of an express truck fleet. In: "*Theoretical and Practical Issues in Transport*"*: Proceedings of the 5th international scientific conference,* 11-12 February, 2010. Univerzita Pardubice: Pardubice. http://hdl.handle.net/10195/37753.

Houbraken, M., Logghe, S., Audenaert, P., Colle, D. and Pickavet, M. (2018) Examining the potential of floating car data for dynamic traffic management. *IET Intelligent Transport Systems*, 12(5), 335–344. https://doi.org/10.1049/iet-its.2016.0230.

Keler, A. (2017) *Traffic Pattern Analysis Framework with Emphasis on Floating Car Data (FCD)*. Doctoral thesis. Augsburg University.

Kessler, L., Huber, G., Kesting, A. and Bogenberger, K. (2018) Comparing Speed Data from Stationary Detectors Against Floating-Car Data. *IFAC-PapersOnLine*, 51(9), 299–304. https://doi.org/10.1016/j.ifacol.2018.07.049.

Leduc, G. (2008) *Road traffic data: Collection methods and applications*. Technical report JRC 47967. Working Papers on Energy, Transport and Climate Change. ftp://ftp.jrc.es/pub/EURdoc/JRC47967.TN.pdf.

Nanthawichit, C., Nakatsuji, T. and Suzuki, H. (2003) Dynamic estimation of traffic states on a freeway using probe vehicle data. *Journal of Infrastructure Planning and Management*, (730), 43–54. https://doi.org/10.2208/jscej.2003.730_43.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*, Vienna, Austria. https://www.r-project.org.

Ranacher, P., Brunauer, R., van der Spek, S. and Reich, S. (2016) What is an Appropriate Temporal Sampling Rate to Record Floating Car Data with a GPS? *ISPRS International Journal of Geo-Information*, 5(1), 1. https://doi.org/10.3390/ijgi5010001.

Sohr, A., Brockfeld, E. and Krieg, S. (2010) Quality of Floating Car Data. Paper presented at *12th World Conference on Transport Research (WCTR)*, Lisbon, Portugal. https://elib.dlr.de/65978/.

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

Vandenberghe, W., Vanhauwaert, E., Verbrugge, S., Moerman, I. and Demeester, P. (2012) Feasibility of expanding traffic monitoring systems with floating car data technology. *IET Intelligent Transport Systems*, 6(4), 347–354. https://doi.org/10.1049/iet-its.2011.0221.

Zhou, X., Yang, Z., Zhang, W., Tian, X. and Bing, Q. (2016) Urban Link Travel Time Estimation Based on Low Frequency Probe Vehicle Data. *Discrete Dynamics in Nature and Society*, 2016, 1–13. https://doi.org/10.1155/2016/7348705.