European Transport \ Trasporti Europei (2023) Issue 93, Paper n° 5, ISSN 1825-3997

https://doi.org/10.48295/ET.2023.93.5

Basic characteristics of floating car data from the perspective of traffic loss during the COVID-19 pandemic

Zuzana Purkrábková[1]*, Pavel Hrubeš[1]

[1]*Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Transport Telematics*

Abstract

Data from floating vehicles is a modern technology and can be another source of data. There is a free data source available in the Czech Republic, which is relatively new. The addressed source of data from floating vehicles covers the whole Czech Republic, which is a promising source for future use e.g. in transport planning in logistics, estimation of travel times and other related issues. For this reason, it is appropriate to examine the qualitative parameters of the data to see if they characterize the traffic stream.

The present paper deals with the size of the processed data. Furthermore, the paper compares the data quality and coverage. January data for four subsequent years was used. The period of the COVID19 pandemic, when traffic declined, was included. Finally, data from selected highways are compared and the period covered is evaluated.

Keywords: Floating car data, data processing, data model, car penetration, basic statistics.

1. Introduction

The increase in car traffic over the last decades is clear. Several requirements and needs are linked to this increase. An important element of the transport system is traffic data. In recent years, great demands have been placed on the quality and reliability of this component. To be able to use the data for mobility planning we need to know that it is good quality and truthful

Besides traditional data collection options, floating vehicles have been discussed in recent years. No additional installation in the vicinity of the infrastructure is required in technology of floating car. Vehicles are used as floating sensors. GPS units are installed in the vehicles, and they regularly transmit data. When connecting data with a spatial stamp, the data acquire a spatial meaning, and it is possible to examine the data in space. The data transmit predefined attributes that are stored regularly. Thus, it is possible to examine historical data and, in the future, data could be also used in real-time.

* Corresponding author: Zuzana Purkrábková (purkrzuz@cvut.cz)

In the Czech Republic, the data source is currently accessible in pilot operation. It is possible to download the data based on a contract signed with the provider, the Directorate of Roads and Motorways. The data is freely accessible, and state-guaranteed and the operation are paid for from public funds. The data have been available since 2019.

This article describes a model of floating car data in the Czech Republic in terms of conversion, data processing and data utilization. In the first part, a literature review is introduced. The authors present articles dealing with floating vehicles, but also a case studies from specific cities. The second part deals with the description of the data model in the Czech Republic and basic information about it. The third part analyses the inputs to the model more specifically. Attention is paid to data size and outputs. Subsequently, the authors then analyse the data source in terms of coverage at the time of the COVID-19 pandemic. The last part of the article deals with the final summary and promises for future work.

2. Literature review

Today, traffic data is a key input into intelligent transport systems. Thus, there are several authors who deal with various data sources or their comparisons. In addition to traditional data analyses such as (Khoury et al., 2003), today there is a lot of research dealing with floating vehicles.

The authors (Mehralian et al., 2020) suggest the use of traffic data to predict congestion, for which it would be possible to use a solved data source in the future. There are also comparisons of floating car data with conventional detectors (Mazaré et al., 2012). There are also several more sophisticated approaches to data.

The authors (Kessler et al., 2018) compare data from induction loops and floating vehicles obtained from TomTom company from the German A9 motorway. Both datasets have data aggregation in 1 minute. The paper shows that in densely populated areas, the loops are more accurate with column detection about 2 minutes earlier than floating car data. Today, the commercial sector invests considerable funds in the development of systems working with floating car data.

Paper (Gunawan et al., 2015) describes the transfer of data from vehicles to the server. It examines how the wireless network will affect the timeliness of traffic data. Another key feature of a data source must be compliance with compatibility standards. The authors (Brož et al., 2022) describe this, for example, on various use cases of transmitted data within cooperative systems. This also needs to be addressed on the solved data source. Cooperative systems could contribute to the expansion of the floating vehicle fleet in the future, so it is appropriate to monitor their development.

Data from floating vehicles can be used not only as an additional data source of speed or intensity but can also be used for further research. The research (Sun et al., 2009) used data from taxi vehicles to investigate traffic in the urban area. The authors (Teixeira et al., 2017) used the data to investigate how it would be possible to read and predict traffic congestion and emissions level. Article (Cowan and Gates, 2002) describes a floating vehicle data system developed in collaboration with a private service and a government agency in the United Kingdom.

Many authors and institutions consider data from floating vehicles as a promising source for the future. As these data are still not defined by any regulation, their parameters and quality are variable and each source needs to be investigated.

3. Available floating car data from Czech Republic

As mentioned earlier, data from floating vehicles have been available in the Czech Republic since 2019. These data are freely available, state-guaranteed traffic data that can be used for further research.

Compared to other countries, it is quite rare that data is published for free and available to all only on request. In general, it is rather the opposite phenomenon. Similarly complex data are charged for, and the owners sensitively select the institutions that will have access to the data. Examples could be large companies such as TomTom, PTV Group, BMW Group, Google and others. These companies usually have data from their systems, but it is not freely available.

Nowadays, each group of car companies develops its own navigation application. These apps usually work on collecting data from their vehicles. The dataset is also linked to official traffic data sources that can be freely used (typically weather data, road closures, actual accidents reported to the national traffic centre, etc.). Here, as well, there could be more participation of floating vehicles in the future as an additional supplementary source of traffic data.

It is essential to examine the quality of data if we want to use the data in telematics applications in the future. This is also associated with an understanding of the model itself and understanding of the distributed attributes and parameters of the data.

It is important to note that this is data collected by floating vehicle technology. This means vehicles that have a GNSS sensor and regularly send the required data. These vehicles usually include commercial fleets, public transport vehicles, trucks, emergency vehicles and others. However, this is not data from all vehicles on the network.

This implies that this is not data collected through all vehicles in the network. These are mostly vehicles driven by a professional driver, which implies some specific parameters. This should be kept in mind for the future. The larger group of these users greatly influences the characteristics of this data. The authors have previously described this issue using the example of speeding on an urban road. It is clear from the available sources that drivers of floating vehicles are more consistent with the prescribed speed than other road users. For this task, data from floating vehicles and induction loops in the selected section were compared.

The available data should cover the whole territory of the Czech Republic. Due to the characteristics of the floating fleet, the coverage of lower-class roads is currently very low. Floating vehicles by their characteristics mostly cover the main traffic routes, such as motorways and first-class roads in the case of the Czech Republic. Floating vehicles also have good coverage in urban areas where additional fleets of vehicles such as taxis are often added.

4. Model background

For a deeper understanding of the resulting data, it would be useful to further investigate the input data to the official floating vehicle model, i.e., the Roads and Highways Directorate interface model. It is also appropriate to further investigate the individual parameters provided by this model.

The model is managed by the Roads and Motorways Directorate (hereinafter referred as RSD), but data collection is provided by VARS BRNO a.s. This is not open data in the true sense of the word, it is necessary to sign a contract with RSD. Based on this agreement with RSD, VARS Company will set up and administer data subscriptions.

4.1 Input into the model from the Directorate of Roads and Motorways

The model was created under the heading VARS BRNO a.s. (VARS, 2020). This company uses cooperation with INRIX and their PTV platform Optima. The result is a modern traffic management platform that can be used to build proactive predictive traffic modelling. Information on the current traffic situation will be provided in real-time to the National Traffic Information Centre. By combining this data with traffic detector data, toll data, meteor data and other data, the National Traffic Information Centre can form a traffic situation model.

In addition to the National Traffic Information Centre, other entities may use the data. A schematic representation of the data flow can be seen in the figure below (Figure 1). Evidently, there are broader mathematical algorithms dealt with in the VARS model and the PTV platform. These algorithms result in 15 attributes that are the output of the Roads and Motorways Directorate model.

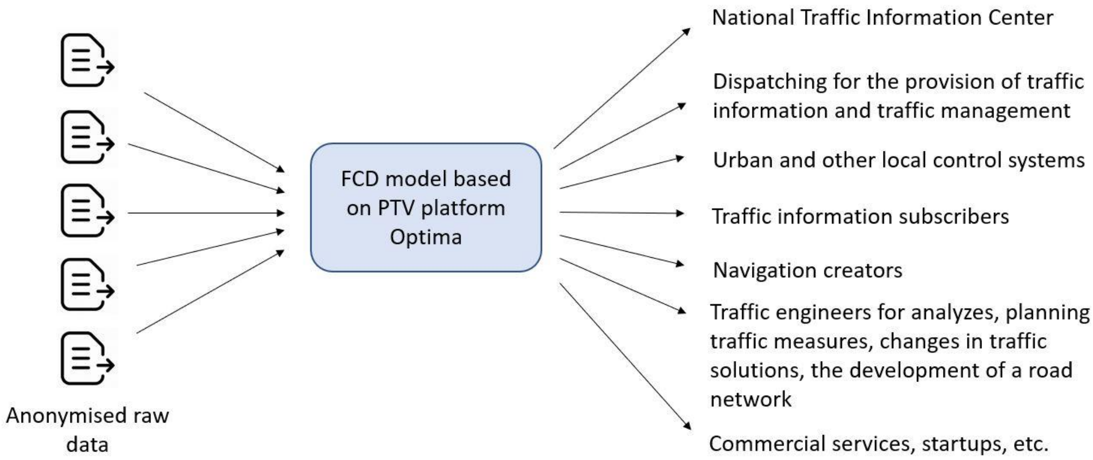

Figure 1: Schematic representation of the data flow in the model of the Directorate of Roads and Motorways (Source: Authors)

The data sent is anonymized. The vehicle ID is changed to a unique identifier. This identifier is then changed while driving to prevent tracking of the specific vehicle. The GPS position of the first segments is removed to comply with GDPR law and guarantee the protection of sensitive data. This occurs even after the identifier is changed while driving.

We can assume data flows so that we can estimate the total volume of data. It is not known what protocol sending messages from vehicles works on. Thanks to the knowledge of the principle of sending from the vehicle to the server, we can consider another communication known to the same with the same server-client communication model, for example, communication when charging electric vehicles. The documentation for the latest communication protocol (Open Charge Alliance, 2021) states the maximum message size is up to one hundred bytes, it is assumed that the FCD message will be slightly smaller. We will consider the size of the message sent by the vehicle at about 50 b and at the same time we assume that the message is sent at least once a minute.

The authors had previously examined the quality of the model output data in the initial phase of the pilot project when they debugged the model errors. Last report is from 2020 (Hrubeš et al., 2020).

The size of the floating fleet is known based on the report (Frydrišek, 2020). It is possible to estimate the size of the data sent into the system in one day as follows:

$$S_d = t * p * s_m$$

where S_d is the size of the received data, t represents the time, p represents the number of vehicles (FCD fleet size) and s_m is the size of one message.

The data is sent every minute, so we will continue to count the time for one day (60 * 24 for the number of files within 24 hours). When calculating the estimated size of the input data for one day, after substituting into equation (1), the calculation is as follows:

$$Sd = 60 * 24 * 150000 * 50 = 10{,}8\ GB$$

The calculated data size per day is not negligible. Any storage of such a large amount of data is already quite financially demanding. It is therefore important to consider how the data can be used and whether it is sufficiently high quality for these needs, i.e., whether data is usable.

4.2 Output from the model from the Directorate of Roads and Motorways

The received data are processed and adapted in the model before being made freely available, as described above. Unfortunately, more detailed information about the calculation is not publicly available.

The outputs of the model are statistical analyses and output data from the calculation module. These are fifteen parameters of already prepared data. Specifically, the TMC vehicle segment (location), timestamp (day and time), count of floating vehicles, two types of speed (average and free flow), travel time (average and free flow), delay, reliability, reaction time, traffic level and three are sent parameters connected with congestion (binary if the congestion exists, location of the beginning of the congestion, length of the congestion).

Unfortunately, it is not public how the individual parameters are calculated, this is VARS's internal information. The output data sent from the model has an average size of 300 MB per day. Thus, there is a significant reduction in data size in the model.

5. Analysis above available data

It is clear from the previous chapters that this is a large amount of traffic data that could be used for transport systems. To do this, it is necessary to examine the quality of this data. Quality is the key issue when using this data. If the data does not provide sufficient information about the current traffic flow, it is not possible to use them for optimization of management of the traffic flow. One of the first parameters of data quality is data penetration, which was used for counting the average speed.

Official data from 2020 show that there are over 150 000 floating vehicles on the network in the Czech Republic (Frydrišek, 2020). The expected increase may have been significantly affected by the COVID-19 pandemic, as well as the general use of vehicles for work purposes. Moreover, due to the situation in the world, business trips were limited and moved to online.

Therefore, the following analysis was performed comparing data from January 2019, 2020, 2021 and 2022. The authors have data available from 2019, the autumn and spring months are too affected by closures due to the COVID-19 pandemic. During these months there were also restrictions on the movement of people in the Czech Republic and the data would be too distorted. The summer months traditionally do not completely describe the average traffic and these months are affected by holiday traffic flow.

Concerning the previous sections, the basic coverage using TMC (Traffic Message Channel) segments with respect to the number of segments but also the directionality of the segments will be analysed, as well as basic statistical characteristics indicative of each month. TMC segments are segments defined based on location tables encoded on the linear road network.

To explain the next part, it is necessary to define what the TMC segment is. The TMC segment is the virtual line above infrastructure created above the Localization Tables (official data managed by company CEDA). It is assigned a unique ID and the data sent from the floating vehicles have these TMC segment IDs for locating on the network. This layer has almost 30,000 elements. It is worth pointing out that these TMC segments do not perfectly replicate road sections.

Comparing TMC segment coverage from January data, by year 2021, 99% of segments had been covered. The exact values are readable in Table 1 in the second column. This is the number of segments that were covered in the monitored month. Data for 2022 are not comparable in this context. In April 2021, the underlying layer of TMC segments changed as the layer gained additional segments. For this reason, this value for 2022 cannot be compared.

An analysis of the use of individual TMC segments in the months was also performed. Not surprisingly, the most frequently monitored segments were related to the motorway network and to urban ring roads.

An analysis of the most occupied segments was done. For a better comparison, overall statistics were also performed, which can be seen in the Table *1***Errore. L'origine riferimento non è stata trovata.**. The table describes the frequency of occupancy of TMC segments. That means, how many times each TMC segment was recorded in a month. It is interesting to note that the statistics are comparable for 2019 and 2020, and likewise for 2021 and 2022. There is a noticeable decrease in floating vehicles on the network in 2021 and 2022.

Table 1: Frequency of occupancy of TMC segments – Average and Median of total detected TMC segment

Year	Count of TMC segments	Average [detected segment/month]	Median [detected segment/month]
2019	29 163	10502	8216
2020	29 249	15004	12944
2021	29 276	6390	4387
2022	(34 781)	6330	4168

Source: Authors

Furthermore, the 5 most frequently occupied segments were selected for each month and these were compared. The initial assumption was that there would be highway segments close to the capital city Prague. This was refuted. The resulting segments are shown in the table below (see Table *2*).

Table 2: Most frequently occupied segments

Segment ID	Location	Frequency of detected segments [detected segment/month]			
		2019	2020	2021	2022
TS01385T01383	City Ring Prague: Bohdalecká → V Korytech	40648	43784	36711	37433
TS01383T01385	City Ring Prague: V korytech → Bohdalecká	40466	43882	36451	37237
TS01385T01387	City Ring Prague: Švehlova → V korytech	40259	43894	36472	37287
TS01387T01385	City Ring Prague: V Korytech → Švehlova	40474	43772	36503	37229
TS23085T01387	City Ring Prague: Švehlova → Lanový most	40329	43610	36151	36692
TS01317T01315	Highway D1: Brno centrum → Brno South	40578	43720	33169	35293
TS01315T01316	Highway D1 Intersection → Brno centrum	40301	43881	32285	33908

Source: Authors

Interestingly, in all four compared periods, the highest ranked segments are specific continuous segments on the city ring road of Prague, connecting large urban areas and residential districts. These segments are shown schematically in the figure below (see Figure *2***Errore. L'origine riferimento non è stata trovata.**). The red segments are the segments that were captured as the busiest in all monitored periods, the purple segments were captured as the busiest in 75% of the periods. Furthermore, sections of the main D1 highway near Brno, where the highway serves as an outer city ring road, were very frequently occupied.

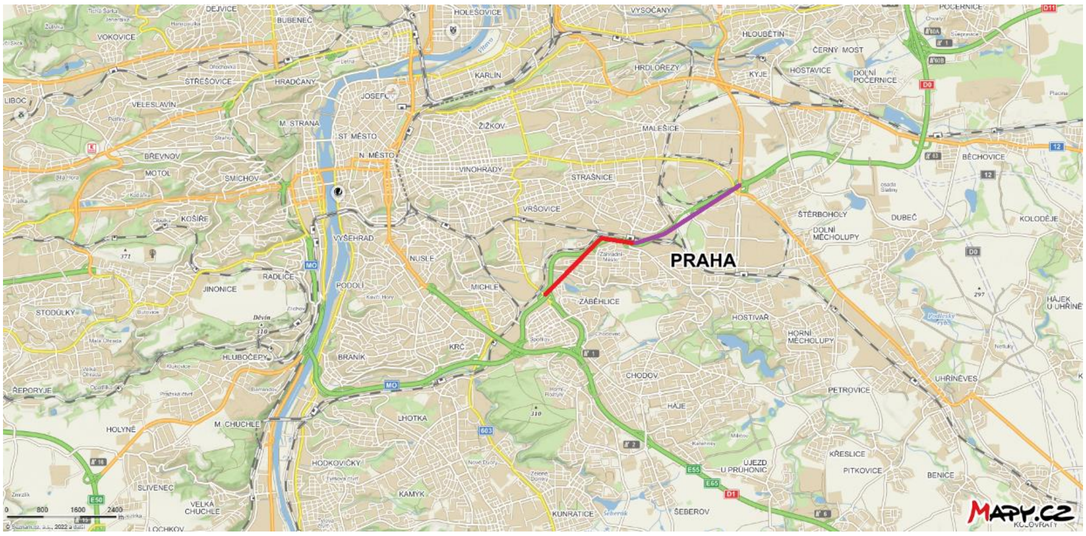

Figure 2: Schematic map of most occupied segment (Source: The authors with the base map from mapy.cz)

Similarly, the segments with the lowest number of records were compared, where again there is a noticeable decrease between 2020 and 2021. The number of segments recorded less than 10 times, less than 20 times and less than 100 times were compared (see

Table *3*). These segments were also compared in terms of location, but there is no visible spatial correlation compared to the busiest segments.

Table 3: Numbers of segments with the minimum number of records

	Number of segments that were recorded less than		
	10 times	20 times	100 times
2019	12	20	165
2020	8	22	120
2021	84	170	710
2022	222	397	1423

Source: Authors

In Figure 3 it is possible to monitor the most used segments during the studied months. For clarity, the top 10% of the most used segments are shown in the map. The most used segments that were used in all examined months are shown in bold grey. The segments that were occupied only in a specific year are then distinguished by colour. More specifically, the segments for January 2019 are shown in blue, data for January 2020 in orange, data for January 2021 in green and finally data for January 2022 in magenta.

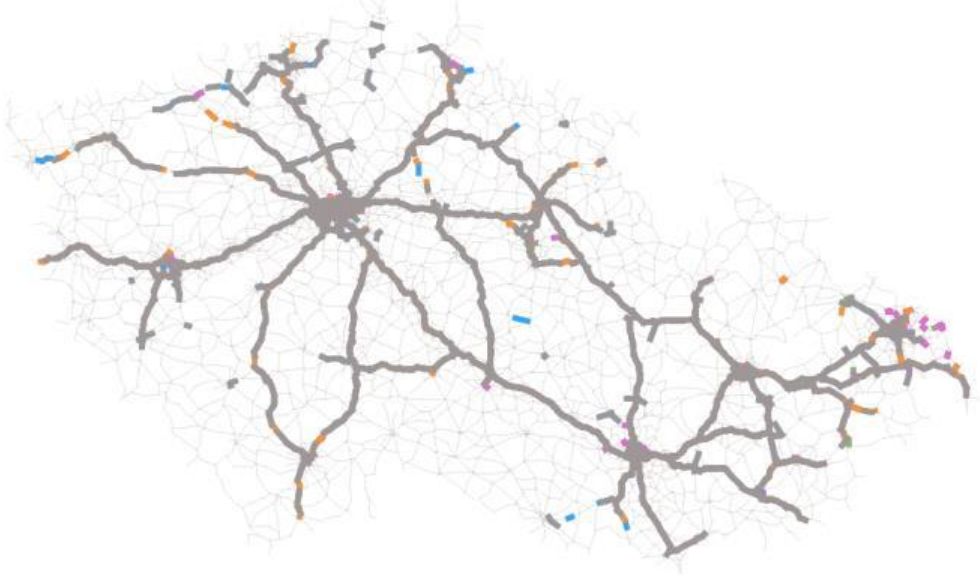

Figure 3: The busiest segments in the monitored period (Source: Authors)

5.2 Analysis of the speed

In the selected months, the speed was also analysed in a corresponding way. Given the previous findings, the question is how the decrease in the number of vehicles between 2020 and 2021 will be projected into the speed.

Table 4: Summary table of speeds for each period.

	AVG	Median	75_PCTL	90_PCTL	95_PCTL	MAX	10_PCTL
2019	59,25	54	77	99	113	240	28
2020	59,23	54	75	96	109	182	31
2021	60,09	53	78	101	115	162	30
2022	59,41	52	77	101	116	162	29

Source: Authors Note: All values are given in [km/h]

When comparing the basic statistical characteristics for speed, it is evident that the observed decrease in the number of floating vehicles is not reflected in this parameter. The average and selected percentile values were analysed. Percentiles were considered because of their greater telling value (compared to minimum and maximum values). The individual characteristics are relatively comparable as can be seen in

Table *4* Speed does not seem to be much affected by the number of floating vehicles, which is very promising in terms of further problems over these data.

Next, the 90th percentiles were analysed, i.e. the sections where the 10% highest speeds were measured. It can be noted that most of these speeds were measured on motorways or freeways. This confirmation is not surprising. Among the other sections in this percentile can be found sections of first-class roads where speeds in this range are measured in greater than the maximum speed limit (90 km/h).

5.3 Highway data analysis

Furthermore, the analysis was carried out on the busiest roads, i.e. on 5 highways in the Czech Republic (namely D1, D5, D8, D10 and D11). The question was whether there would be a clear drop in the number of floating vehicles and whether this change would be reflected in the speeds.

At the beginning, the assumption was made that the highways would be comparably occupied, but slightly higher speeds would be detected on the main D1 highway, which connects the 3 biggest cities in the country and is the longest highway. Subjectively speaking, vehicles generally travel faster on this highway. From previous analyses, the assumption is that speed will not be much affected by the smaller number of detected vehicles.

After the analysis, it can be concluded that there is no decrease in the number of vehicles on the highways in speed either. Surprisingly, the average speed is similar on all highways and time periods, approximately 109km/h. This is interesting since even the mentioned D1 does not stand out from the statistics.

Next, the percentiles were analysed. It can be argued that the 90th percentile corresponds roughly to the highest speed allowed, i.e. 130 km/h. Conversely, the 10th percentile corresponds to a speed of 90 km/h. Again, this can be claimed for all the highways in the examined period.

The numbers of detected vehicles were also examined. Here again percentiles were used as they have a greater telling value for the resulting dataset. A decrease in vehicle counts was evident on all of the resolved highways. There is a noticeable drop in the data between 2020 and 2021. The 90th percentile is slightly higher for the D1 highway (this is based on the hypothesis), while it is slightly lower for the D8 highway (which connects Prague and the north of the country). Conversely, the decreases between 2020 and 2021 are tens of detected sections lower, as can be observed in the example table for the D8 highway.

Finally, it was analysed how many sections are affected by speeds exceeding the 90th percentile, to determine whether these are highway sections only or not (see

Table *5*). Out of the total of about 29,000 segments, which was analysed above, the least number of segments above the 90th percentile was detected in 2021 (7195 segments, i.e. almost 25%), while the most number was detected in 2020 (9867 segments, i.e. over 30%). This means that these are not only highway segments, which is also a reasonable explanation when looking at the value of this percentile. This could be a good sign for the future that the data is not primarily occupied by highways and freeways.

Table 5: Statistical values for highway D8

	Speed [km/hour]	Count [veh/segment]

	AVG	90_PCTL	MAX	10_PCTL	AVG _car	AVG _truck	90_PCTL _car	90_PCTL _truck
2019	110,8	130	236	87	7	12	19	25
2020	105,6	124	182	87	11	18	28	39
2021	105,2	129	162	84	2	1	6	3
2022	107,4	130	162	86	3	1	6	3

Source: Authors

6. Conclusion and future work

A new data source has been introduced. From the beginning, this resource is promising for the future, as this data source does not require the installation of technology along the transport infrastructure. All what is required is a GNSS unit in the car (just a mobile phone or navigation system), which is now a common part of cars.

Attention was paid to the size of the data. Since it is modern to store a large amount of data today, the care must also be taken to use them. Therefore, the attention was further focused on comparing data from floating vehicles for the month of January over a period of four years (2019, 2020, 2021 and 2022).

The data was analyzed in terms of coverage, number of detected vehicles and speed. Finally, five major highways in the Czech Republic were compared and the measured speeds and numbers of vehicles were compared.

If floating car data are reliable, it will be possible to use the data as a full or complementary source to other traffic data. It is possible to use the existing fleet, which is based on data collection using a signal from GPS. It is a commonly used technology, and this is a prerequisite for the future expansion of the floating vehicle fleet. It is thus possible to obtain an additional data source without the large necessary investments only at the cost of storage hardware and data security.

Data from their characteristics will be very heterogeneous throughout the Czech Republic. Even from the above analysis, the data has changed significantly in the last two years due to the impact of the traffic flow. This caused by pandemic of COVID-19 and the data need to be further investigated. If the decline in traffic did not have a significant effect on the data, or if the data outputs were not affected by this, it would be a promising step for the future.

In order to be able to rely on this data as inputs to transport systems, the key question is how many floating vehicles on the network are needed. The problem arises of the critical value of floating vehicles, below which it is not possible to fall below so that the data source is still conclusive. Obviously, this is a key issue for future research.

References

Brož, J., Šrotýř, M., Jeřábek, M. (2022) “Use Cases and C-ITS Communication Analysis as a Step Towards the Connected Mobility”. In: *Sustainable Management of Manufacturing Systems in Industry 4.0*. EAI/Springer Innovations in Communication and Computing. Springer, https://doi.org/10.1007/978-3-030-90462-3_2

Cowan, K., Gates, G. (2002) "Floating Vehicle Data system - realisation of a commercial system" *Eleventh International Conference on Road Transport Information and Control*, pp. 187-189, doi: 10.1049/cp:20020230.

Frydrišek, M. (2020) "Standardizovaná data z plovoucích vozidel jsou zdarma k dispozici široké veřejnosti". *Dopravní noviny*, online, 31.7.2020. Available at https://www.dnoviny.cz/informatika/standardizovana-data-z-plovoucich-vozidel-jsou-zdarma-k-dispozici-siroke-verejnosti

Gunawan, F. E., Chandra, F. Y., Soewito, B. (2015) "Data transmission strategy of probe vehicle in floating car traffic monitoring," *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 303-308, doi: 10.1109/ISITIA.2015.7219996.

Hrubeš, P., Purkrábková, Z., Kovaljov, M., Kovaljová, K., Hampl, J. (2020) "Kontrolní a expertní činnost při zavádění a následném provozu FCD – kalibrace výpočetního modelu III", report no. 97/2019b.

Kessler, L., Huber, G., Kesting, A., Bogenberger, K. (2018) "Comparing Speed Data from Stationary Detectors Against Floating-Car Data", *IFAC-Papers OnLine*, vol.51, issue 9, pp. 299-304, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2018.07.049.

Khoury, J., Haas, C., Mahmassani, H., Logman, H., Rioux, T. (2003) "Performance Comparison of Automatic Vehicle Identification and Inductive Loop Traffic Detectors for Incident Detection". Journal of Transportation Engineering. 129, https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(600).

Mehralian, S., Teshnehlab, M., Nasersharif, B. (2020) "Traffic data analysis using deep Elman and gated recurrent auto-encoder". Neural Network World. 30. 347-363, https://doi.org/10.14311/NNW.2020.30.023

Mazaré, P. E., Tossavainen, O. P., Bayen, A. M., Work, D. B. (2012) "Trade-offs between inductive loops and GPS probe vehicles for travel time estimation: A Mobile Century case study", TRB 2012 Annual Meeting, https://bayen.berkeley.edu/sites/default/files/mazare_12-2746.pdf

Open charge point protocol 2.0.1, OPEN CHARGE ALLIANCE, online, 2021, https://www.openchargealliance.org/protocols/ocpp-201/

Sun, J., Wen, H., Gao, Y., Hu, Z. (2009) "Metropolitan Congestion Performance Measures Based on Mass Floating Car Data," *2009 International Joint Conference on Computational Sciences and Optimization*, pp. 109-113, doi: 10.1109/CSO.2009.374.

"Společnost VARS BRNO a.s. podepsala smlouvu s ŘSD ČR na systém pro zjišťování dynamiky dopravních proudů v ČR", *VARS BRNO a.s*, online, 2020. Available at http://www.vars.cz/smlouva-fcd

Teixeira, J., Fernandes, P., Bandeira, J., Coelho, M. (2017) "Information Management for Smart and Sustainable Mobility". In: *TRB 96th Annual Meeting Compendium of Papers*, Washington, https://ria.ua.pt/handle/10773/22105

Acknowledgements

Data from the Transport Portal of the Czech Republic were used for this project (data source: ŘSD ČR – JSDI – www.dopravniinfo.cz).