Paper ID 707

Outputs and characteristics of the traffic flow speed model based on floating car data

Zuzana Purkrábková[1*], Pavel Hrubeš[2]

1. Czech Technical University in Prague Faculty of Transportation Sciences, Czech Republic, purkrzuz@fd.cvut.cz

2. Czech Technical University in Prague Faculty of Transportation sciences, Czech Republic, hrubepav@fd.cvut.cz

Abstract

In the Czech Republic, there is a new, freely available source of data on the speeds of traffic flows on the road network, based on the movement of floating vehicles. This opens up new possibilities for community and commercial use, as well as research, especially in assessing the qualitative characteristics of this resource and the possibilities and areas of application of its use. It is a unique source of traffic data covering the entire country and it is necessary to examine its reliability in time and space resolution. In the future, this source could be used directly to influence and manage the optimization of the transport network operation, so it is necessary to subject it to a detailed qualitative examination. In addition to these qualitative analyzes, the paper also deals with similar research and the presentation of the user interface. Finally, possible uses and other directions of research are discussed.

Keywords:

Data collection, floating car data, geographic information software, traffic data

Introduction

Today, roads are predominantly equipped with detectors and powerful devices with other communication systems that provide traffic control, traffic flow optimization or, for example, driver information. In general, the basic building block of all telematics systems is traffic data.

In the Czech Republic, there is a new source of traffic data published by an organization under the responsibility of the state, the Directorate of Roads and Motorways, which is based only on floating car data. These are real-time data from the motorway, road, and city network in the Czech Republic, which are regularly published in a very detailed cycle, a period of one minute. To create this data, the need for the installation of additional sensors or other technology along the roads has arisen. The source is created on the basis of agreed contractual relationships with traffic data integrators and owners of large fleets of vehicles in operation.

Due to the relatively large and dense road network in the Czech Republic, such a data source is very promising and useful. Thanks to the mentioned characteristics, this is a very significant qualitative shift in spatiotemporal knowledge of traffic behavior on roads in the Czech Republic, especially compared to standard models of static traffic surveys, which in the timeline affects the dynamics of traffic places, ie sections where there are no measurements or observations. The new data source is currently operating in pilot operation, where the calculation model is being modified, calibrated, and tuned, as the data of floating fleet vehicles have a number of specifics that must be accepted, respected, or filtered directly from the calculation and outputs. The authors of the article deal with this source in detail and bring current findings and conclusions.

Generally, floating cars are vehicles equipped with a location device that sends data to a database, as defined in more literature, such as [1]. Floating vehicles are mobile probes on the road network, with which we are able to collect data from real and current traffic on the transport network. They are most often equipped with a GNSS device and a communication unit that sends the position together with other parameters. Alternatively, it is possible to monitor vehicles in general based on a signal from mobile phones, however, the first variant allows the contractual provision of acquisition not only with data on the position itself but with many other parameters about the vehicles themselves. Floating car data (FCD) are often used to measure travel time.

Literature review

The idea of collecting data from floating vehicles dates back to the end of the last century, when the German companies Siemens and Philips wanted to install a similar technology [2]. Most current research deals with the collection and use of data in the city. Very often, data from taxi fleet vehicles are compared. Various papers have processed this data in the past, such as data from Berlin [1], Stockholm [3] or Beijing [4]. These vehicles have very good penetration in larger cities. The data are slightly distorted by the characteristics of drivers, where their knowledge of local roads, differences in traffic during the day and specific areas of the beginning or end of the route (often located in touristic or business areas) is assumed.

It is possible to watch several projects in the suburbs. Research by Kerner at all [5] describes a system of data collection from floating vehicles throughout the network, where it discusses the required percentage of data and their quality, as well as addresses the connectivity of the system and the possibility of data transmission. The main goal is to reduce the amount of data sent so that the transport center is not overwhelmed by duplicate data.

More similar projects can be found in authors in the Netherlands. An initial analysis of the possibilities of this data was performed by Turksma [6]. The possibilities of using data from floating vehicles, as well as the technical and economic feasibility of measuring these data for urban areas were discussed. The authors discuss the necessary parameters for traffic management on the Dutch motorway around Rotterdam. Paper [7] solves the fusion of floating car data and stationary detectors and tries to eliminate distortion of individual sources. The authors worked with data from one specific company (Be-Mobile), so they did not have a complete sample of floating vehicles.

A number of research projects are also addressing the issue, such as the UMneo project [8] for real-time traffic management and the future replacement of induction loops, the Bavarian project for improving traffic information [9] or Italian research on the future interconnection of smart junctions [10].

There are many national or cross-border projects, where real time data is gathered in each country and harmonized We can mention project Crocodile [11], Czech national RODOS [12], or for example UMneo [13]. A team of experts at FTS CTU in Prague has already dealt with the Czech source. The primary basis for the creation of a data source is a study [14], which deals with the importance and possibilities of data from floating vehicles and related technology. Three control reports have been issued for the operated model, which analyze data from different periods, the last one [15] from July 2020.

Data acquisition application

In 2019, an application for providing traffic information based on the collection of floating car data in a pilot project was launched in the Czech Republic. Specifically, it is a project "Development and delivery of a system for area continuous monitoring of traffic current dynamics on the communication network of the Czech republic". Organization VARS BRNO a.s., which is the creator of many telematics applications in the Czech Republic, succeeded in the tender for the implementation of this project. VARS is responsible for all system design and implementation, integration of FCD data inputs, model calibration and running, operation of the graphical output interface and continuity of outputs on the information environment of the Directorate of Roads and Motorways. Cooperation with the PTV Group was used to develop the model itself.

A research group from CTU in Prague, Faculty of Transportations sciences was also involved in the initial phase of project preparation, determination of technical specifications and control mechanisms. Currently the role of this group is mainly from the point of view of an expert observer. Furthermore, their function is also to control the functioning quality of the entire model.

Data interface

It is possible to sign up for the actual collection of model output data by signing a contract with the Directorate of Roads and Motorways (RSD = abbreviation used). Collection of this data is free and every entity has the opportunity to request it. After negotiating this contract, it is necessary to set up an Internet address to which the data will be sent by output.

In the Internet interface, it is possible to select a spatial or otherwise defined area of interest, ie to select consumption based on the entered parameters. The interface allows you to define the class of communication, geographical area, degree of traffic and select only sections detecting the column (binary yes / no). It is also possible to further specify in the note the period of sending data or a more specific request for data collection. The last option is to select the data format, where you can select DATEX, HTTP or Native XML, HTTP. The model does not yet speak a language other than Czech. An example of the interface can be seen in the picture Figure 1.

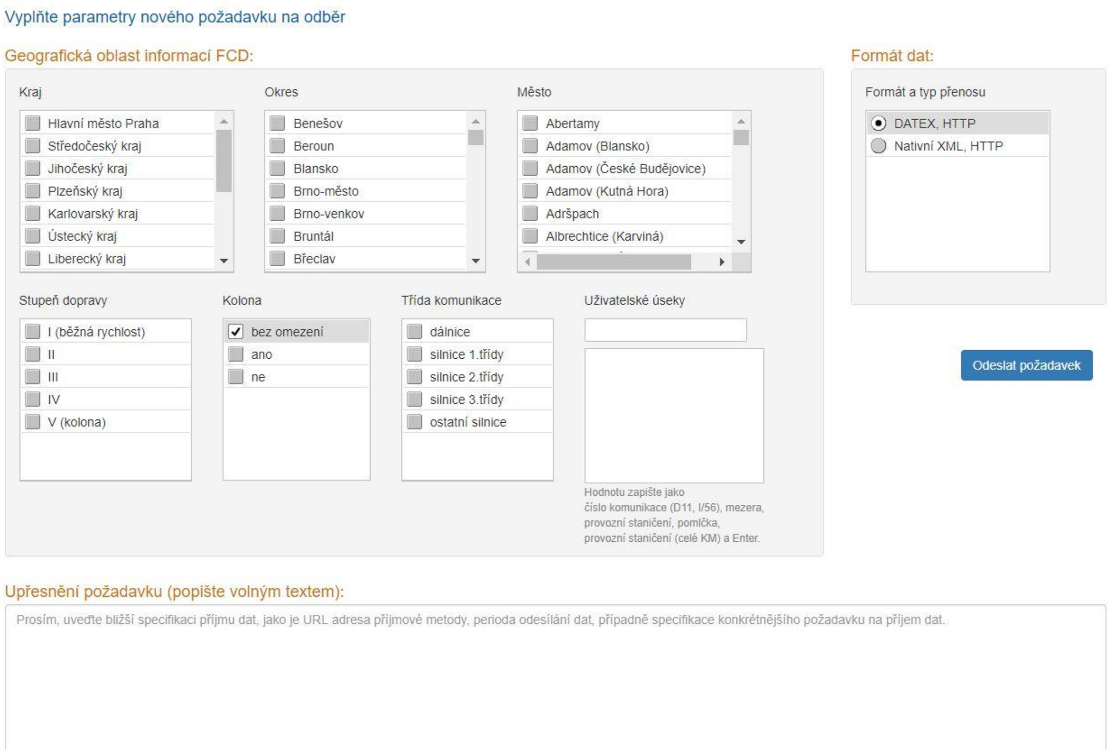

Figure 1: User interface

Real time data

After the specified and approved subscription, the data is sent to the selected address. The calculation model itself works in minute intervals, in which several parameters are calculated for each monitored section from the available received FCD messages. The data are already processed by the integrator and fully anonymous, so it is not possible to monitor individual vehicles in the network in the outputs, only the characteristics of the transport product and other parameters. The provided and sent output data of the model contain a total of 15 attributes, which are summarized here:

Defined TMC segment

The computational model is calculated in the detailed spatial resolution of the Global Network, product from CEDA, a.s., which was created and updated in cooperation with CEDA and RSD, and is used in the Czech Republic under a departmental license by the Ministry of Transport. The open output itself for the public is published in the spatial resolution of the product Localization Tables of the Czech Republic, which is a localization database certified by the Association for Traffic Information TISA. The individual road sections are defined in the spatial view as TMC segments created for applications in the RDS-TMC system. This is an unambiguous designation of individual road segments, with the help of which we can determine a specific section to which traffic information has been assigned, in this case model data calculated. The TMC segment is not a spatial copy of the communication line, but a virtual line of the network model edge above the communication. This identification attribute has a text format, the marking is a combination of numbers and letters. An example of such a segment with a map background can be seen in Figure 2 (TMC segment no. TS13779T25637 on highway D11, direction to Prague, km 79.1).

Figure 2: TMC segment no. TS13779T25637

Day and time

Another important parameter of the model output data is the time stamp, ie the designation of the day and time to which the data was calculated. The date is in the format YYYY-MM-DD and the time in the format HH: MM: SS is always in Coordinated Universal Time UTC. For possible local comparison during data processing and analysis, it is therefore necessary to take into account a possible shift of time according to the relevant time zone and summer or winter period.

Speed and Count

Speed on the segment is a key parameter for floating car data model output. Two values of speeds can be found in the data. The first one is speed of traffic flow in current condition so the speed for this moment of time. The second (free_flow_speed) one is speed of traffic flow in free traffic flow. Both of these parameters are measured in km / h.

The Count parameter specifies the value of the current number of FCD vehicles in the segment, ie the number of vehicles that were included in the calculation of the model speed on the section.

Travel time and delay

The following are three values associated with time. The first is the value of the travel time, the so-called travel time given in seconds. This is the time required to traverse the segment under current conditions.

Similarly, as with speed, the travel time in the free traffic flow (free_flow_time) parameter is specified for time. This is used to calculate the delay.

The last value is the delay, which is also calculated as the difference between the current travel time and the travel time in free traffic flow.

Reliability and Reaction time

A statistically interesting value is the reliability parameter. It tells us how reliable the output data of the model is. This parameter characterizes the quality of data. This is a value between 0 and 1, which reflects the percentage success of this parameter. So if this parameter is equal to 1, it means that the data is 100% trustworthy. Quality is evaluated with more parameters such as number of the floating car data, reaction time, inputs and more.

Reaction time is time required for system response. It is calculated as the difference between the availability time of inputs and the calculation time.

Traffic level

Traffic level is a qualitative parameter for measurement of the quality of traffic service. Traditionally in the USA, it is rated on the A-F scale, where A means free flow. Similarly, a rating of 1-5 is used in Europe, where 1 is equivalent to A. In our data, the conventional marking 1-5 is used.

Parameters connecting with congestions

The last values are parameters connecting with congestion. The first is binary information about whether there is congestion on the segment. Conventionally marked 0/1 in the data.

Additional values are recorded if the binary congestion response is 1. The "Congestion_from" parameter specifies information from where congestion is detected on the segment. It is given in meters (for example 870). The "Congestion_length" finally estimates the length of congestion in meters. These two values are in string format.

General consideration of the model

With basic knowledge of the data format and the principle of model creation, it is possible and desirable to check these outputs and possibly calibrate the model or directly modify it, according to knowledge from operation and comparison with other sources of knowledge about the traffic on the transport network. It is really clear that this is a very interesting source of comprehensive traffic data, which can significantly contribute to high-quality and full information of drivers and the public in general in the future.

At the same time, this situation places great emphasis on real knowledge of all the limits and limitations of this resource, so that it is not attenuated or even damned due to improper use and application. For public and guaranteed publication of data, it is desirable to have the data verified and, of course, to know how and why the data is used, as well as to draw feedback on improving the output. It is impossible to assume an automatic understanding of the data by the public or professionals because it is really something new.

To give these considerations as an example, the fact that no data is reported on a segment does not mean that there is no vehicle in the segment (let's start from a fundamental graph of speed versus intensity, where intensity can be low at high speed and at very low speed). Likewise, a sudden reduction in speed does not automatically mean congestion on a section, especially if it is detected in only one vehicle, etc. Therefore, it is necessary and desirable to constantly verify the model data and analyze the model outputs for a long time. The texts below show exemplary examples that are needed from a research point of view to address the quality of the outputs by calibration or modification of the model itself.

Spatial network analysis

As already mentioned, the data from the model are represented for the purposes of research in spatial characteristics by means of TMC segments and it is possible to display them, for example, by means of freely distributable software QGIS. In general, the spatial component of the data is very suitable for the illustrative presentation of results, for traffic data this is absolutely necessary for understanding specific individual events, their impact on the environment, and the spread in the transport network. For these needs, it is necessary to perform an analysis of the spatial network itself, whether the output model is correctly assigned to all necessary sections, whether there are no discontinuities in the network, or confusion of the connection of individual sections, etc...

In this area, the problems of missing opposite segments, segments not connected to the presumed non-existent roads, or, for example, chaotic areas, ie an area with several segments that overlap each other and create a regular follow-up graph, were identified. In all cases, this issue was pointed out and it was further recommended to examine these segments and special cases. It is not clear that, for example, a missing opposite segment automatically means an error in the representation of the network, but it can be an image of the reality of one-way communication or a model expression of one communication with a different number of segments for opposite directions.

It is possible to view the spatial layout of segments within the Czech Republic by linking a spatial part to data. Based on the tools of the GIS software, it is then possible to create further spatial analyzes. Figure 3 shows the data coverage throughout the Czech Republic, where the blue color shows a very low number of vehicles. On the other hand, the yellow one means good road coverage.

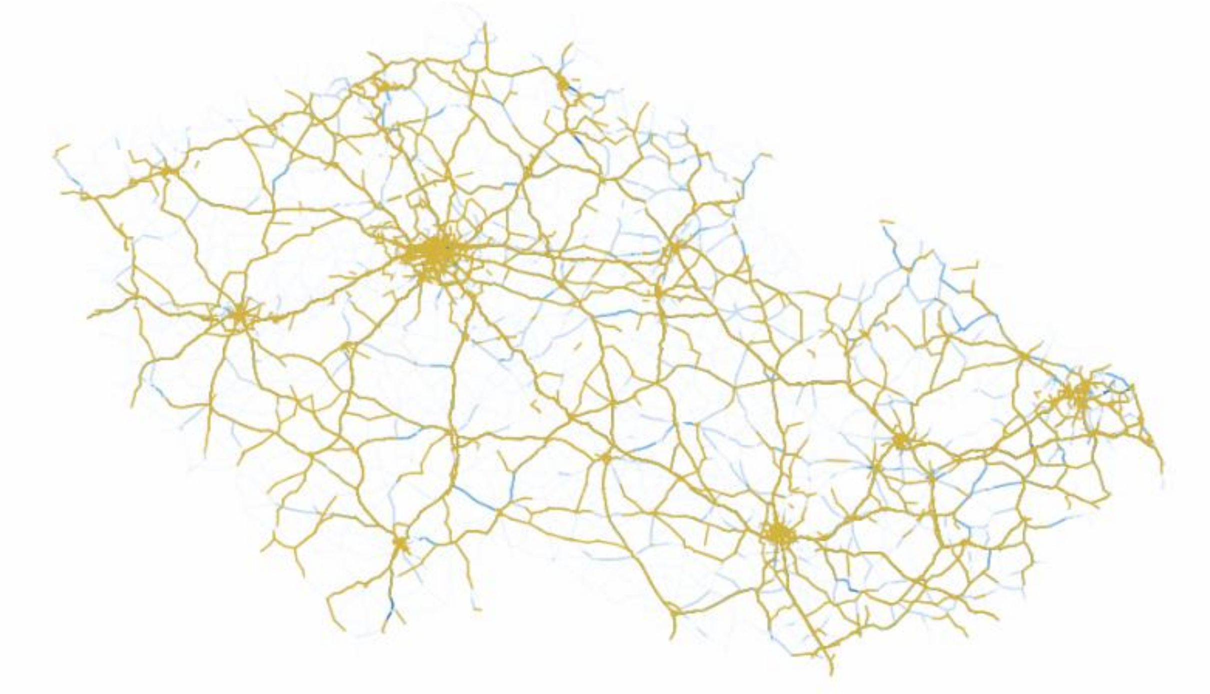

Figure 3: Data coverage in the territory of Czech republic

Assessment of the "vehicle speed" characteristic

To determine the quality of the data, the outputs of the model from floating vehicles were compared separately as time series and also with the data of stationary detectors, especially on the main roads, where they are available as standard. Here it is necessary to monitor whether these are mainly atypical waveforms compared to the expected curves. Again, when significant discrepancies are found, it is necessary to further discuss each case and subject it to a closer examination. It is always necessary to fine-tune segments with low penetration within the model, ie low occurrence of the number of vehicles included in the computational model, where these sections often have a very unstable course of data and distort the success of the model in quality assessment.

During the monitored periods were found, for example, segments with a constant speed over time, which again may not necessarily be a model error but may be a data image of the actual operation. It is therefore precisely these all anomalies that need to be further investigated and resolved.

A typical problem with output values is the different stability (variable scatter size) of the output values of the data source during the day. It is clear from the nature of the fleet's operation, from the uneven distribution of the share of lorries and cars in space and time that there will be considerably fewer vehicles at night and in the morning than during the day. Therefore all this has a significant impact on the final calculation. A small amount of data distorts the telling value when with a lower number of floating vehicles, the resulting speed is much more fluctuating than with average values with a higher number of vehicles. This instability can be seen in Figure 4. The figure shows the course of speeds on one TMC section of the highway during one day when the time course is displayed on the x-axis and the y-axis shows the speed taken from the floating car data for the displayed segment. Data from the floating vehicle model are displayed in red. They are compared with data from conventional detectors (two detectors for each lane of the two-lane road, colored in green).

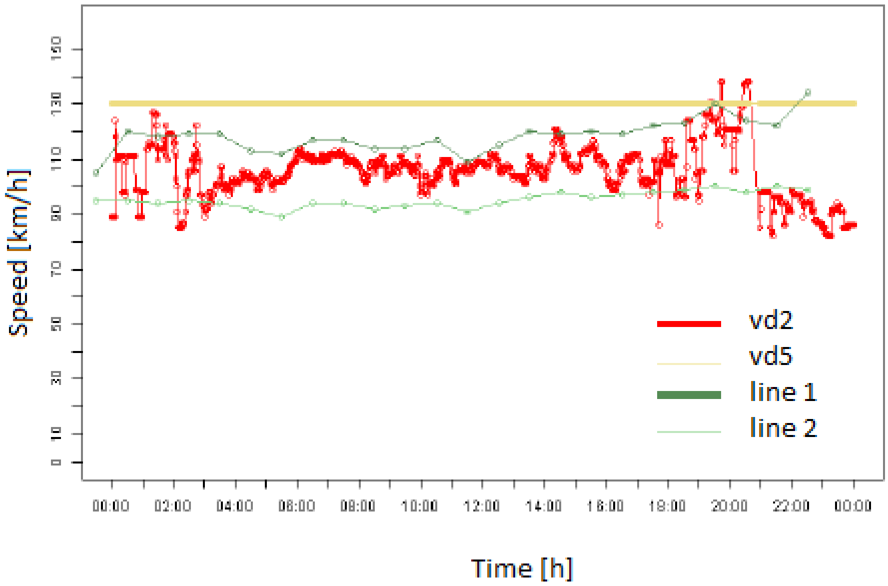

Figure 4: Volatility of floating data during the day

The discussion was also provoked by segments where a significantly higher speed of specific individual vehicles was detected within the calculation than is the maximum permitted by law, and thus the average final calculation was also higher than the usual speed average at any time during the day. It was discussed whether it is appropriate to publish this information as an output of the model. Each significant deviation distorts the quality of the output. In this spirit, it is certainly not expedient to present above-limit emergencies but the characteristic traffic flow. It is necessary to specify the current but usual speed, and not the result of a calculation based on several drivers proving the power of their machines.

However, it is similar with lower speeds than usual, when this can significantly affect the character of the fleet and specific vehicles currently present in the area. An example is whether the construction vehicles, which are included in the calculation of the model, have a very significant effect on speed, but they are on a parallel road, so the main road is completely free to pass and alright at the same time.

Assessment of the "vehicle count" characteristic

As well as the parameter of the speed itself, it is necessary to monitor and control the parameter of the number of vehicles included in the calculation. Even in this case, a number of sections with a non-standard course were observed, for example, where the number of vehicles was constant during the days. This is a completely relevant value for a smaller number of detected vehicles, but at higher values it was recommended to further investigate and monitor this behavior.

Comparison of FCD speed and static detectors

Very interesting results were found when comparing data from floating vehicles and data from conventional detectors. These data were again provided by RSD. It is obvious that with a sufficient number of vehicles, the data are indicative of the entire traffic flow. This is shown, for example, by the comparison in Figure 5. The axes are similar to the first figure, the x-axis shows the time during one day, the y-axis shows speed taken from the floating vehicles. The red line data is data from floating vehicles, the green data is data from lane detectors. The picture also shows when the traffic flow slowed down due to the morning rush hour. This is a section of the main motorway near Prague towards the center of Prague when traffic rises heavily at rush hour.

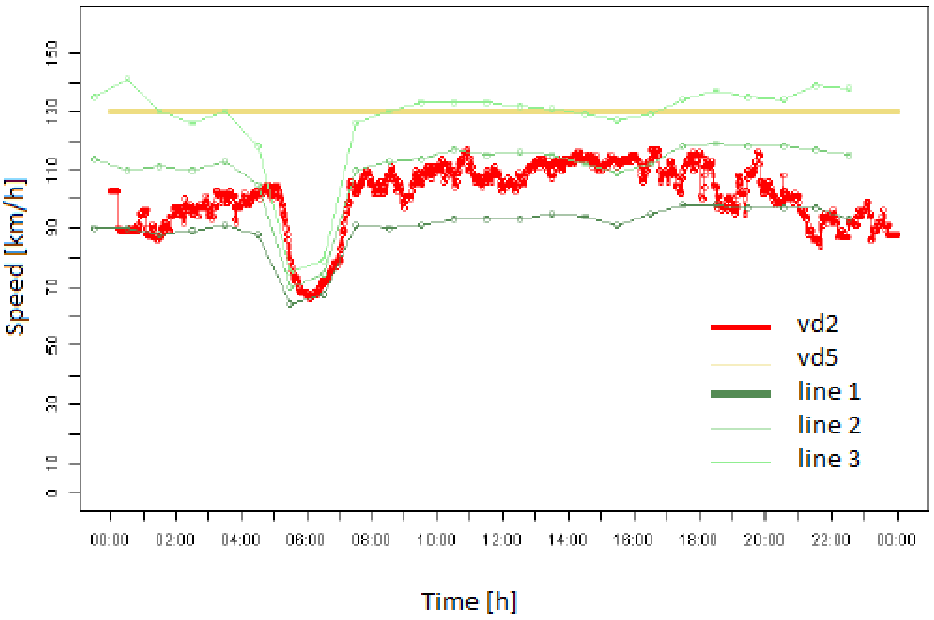

Figure 5: The velocity on the highway D1 direction Prague

Conclusion

The paper presents a new data source, which is currently available in the Czech Republic and is running in pilot mode and also available to the general public. This extensive data source is completely new in the Czech Republic and it is advisable to investigate it further. Until then, there was no source describing the entire road network. The authors aim to analyze and research this resource in the long term, to support its use commercially and in the community, to ensure that this national investment is justified and, if possible, remains operational as a significant alternative to infrastructure detection technologies.

It is not fully open data in the true sense of the word. In the current operation, a contract with the operator is required for data collection and therefore it is not possible to download directly from the web interface, contractual relations are not hindered in any way. This approach is justified for a feedback discussion on how this resource is used. The offered data formats meet the European requirements for DATEX II, so in the future coordination with other European systems or applications based on similar data sources will be possible.

The data attributes provided have been described above and some have been analyzed in more depth. During the start-up and calibration of the model, some phenomena presented and analyzed here were observed and analyzed and are still discussed with the system operator. The overall spatial network of the transport system was analyzed, to which data from floating vehicles are applied. This led to recommendations for updating specific locations of this network. The qualitative deviations of the model with lower vehicle penetration were also pointed out. This data source may continue to encounter this problem if the vehicle fleet of the computational model is not further expanded.

The source discussed in the future promises numerous uses not only in road transport itself but in closely related fields such as logistics, just-in-time processes, or in connection with MaaS, etc. It was found that data with sufficient penetration plot the traffic flow and capture emergencies, mainly by a sharp decrease in speed, which was also observed in comparison with data from conventional detectors.

The research also points to the suitability of further and constantly subjecting this data to detailed research and to designing new modern telematics solutions. In the future, based on this data, it would be possible to transmit real-time information directly to road users, influence and control traffic flow on the nationwide network (ie divert traffic from supersaturated traffic to looser using variable traffic signs or other support systems) with a guarantee and state decision, not only as recommended by community applications with unknown preference and bias algorithms. And now, at the time of the COVID-19 pandemic, when the movement of company fleet vehicles decreased, this issue is also necessarily discussed and the sensitivity of the model to this parameter is examined. Furthermore, the authors are also engaged in research into the emergence and use of alternative routes in the network under exceptional and standard conditions, modeling this issue to optimize traffic in the network.

Acknowledgement

Data from the Transport Portal of the Czech Republic were used for this project (data source: ŘSD ČR – JSDI – www.dopravniinfo.cz). This research was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS SGS21/080/OHK2/1T/16 - Detection of detours of an accident on main roads based on floating car data.

References

1. Jürgen M, K. Thiessenhusen, P. Wagner (2003). Using FCD-data for measuring route-choice probabilities. In Proceedings *Behavioural Responses to ITS*

2. Huber W., M. Ladke, R. Ogger (1999). Extended floating-car data for the acquisition of traffic information

3. Rahmani M., E. Jenelius, H. N. Koutsopoulos (2015). Non-parametric estimation of route travel time distributions from low-frequency floating car data, *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 343-362, doi: https://doi.org/10.1016/j.trc.2015.01.015

4. Zhu D, Y. Bai (2016). Floating car data processing model based on Hadoop-GIS tools. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Tianjin, pp. 1-4

5. Kerner B. S, C. Demir, R. G. Herrtwich, S. L. Klenov, H. Rehborn, M. Aleksic, A. Haug (2005). Traffic state detection with floating car data in road networks, In Proceedings *2005 IEEE Intelligent Transportation Systems*, Vienna, pp. 44-49, doi: 10.1109/ITSC.2005.1520133

6. Turksma S. (2000). The various uses of floating car data. *Tenth International Conference on Road Transport Information and Control*, London, UK, pp. 51-55, doi: 10.1049/cp:20000103

7. Houbraken M. et al. (2015). Real-time traffic monitoring by fusing floating car data with stationary detector data. *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Budapest, pp. 127-131, doi: 10.1109/MTITS.2015.7223247

8. UMneo (2019). UMneo real-life pilot on floating car data: first results. The European ITS Platform, [online 09.02.2021]. Available: https://www.its-platform.eu/highlights/umneo-real-life-pilot-floating-car-data-first-results

9. Floating Car Data Enhances Traffic Accuracy in Bavaria. [online 09.02.2021]. Available: https://inrix.com/case-studies/floating-car-data-enhances-traffic-accuracy-in-bavaria/

10. Astarita V., V. Giofré, D. C. Festa, A. Vitale (2020). Floating Car Data Adaptive Traffic Signals: A Description of the First Real-Time Experiment with ”Connected Vehicles. *Electronics*, vol. 9. 1, pp. 1-2, doi: http://dx.doi.org/10.1002/andp.19053221004

11. The Crocodille Coridor (2021), [online 09.07.2021]. Available: https://crocodile.its-platform.eu/

12. viaRODOS (2021), [online 09.07.2021]. Available: https://rodos.vsb.cz/Overview.aspx

13. UMneo real-life pilot on floating car data: first results (2019). Online. Available: https ://www.its-platform.eu/highlights/umneo-real-life-pilot-floating-car-data-first-results

14. Hrubeš P., M. Langr, P. Derbek, D. Saiko, M. Volný (2010). Zmapování služeb a dat v oblasti FCD (Floating Car Data) pro využití v rámci informačních systémů ŘSD. ČVUT v Praze, Fakulta dopravní, Ústav řídicí techniky a telematiky, Research report no. LSS 384/10.

15. Hrubeš P., Z. Purkrábková, M. Kovaljov, K. Kovaljová, J. Hampl (2020). Kontrolní a expertní činnost při zavádění a následném provozu FCD – kalibrace výpočetního modelu III, 2020, ČVUT v Praze, Fakulta dopravní, Ústav dopravní telematiky, Research report no. 97/2019b.