Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

Estimating a light commercial vehicle OD matrix based on the vehicle tracking data of heavy good vehicles

András Szele*

Institute for Transport Sciences, Thán Károly 3-5, Budapest 1119, Hungary

Abstract

Light commercial vehicle (LCV) is one of the most dynamically developing vehicle categories, but, unfortunately, there is a general lack of data on LCVs. Perhaps this is the reason behind the traditional traffic modelling practice that LCVs are treated as automobiles and included in the passenger car Origin Destination (OD) matrix, which may lead to false outcomes. We had a sample LCV OD database with 2800 records which describes some specific travel patterns, but there were no data on the production and attraction of the settlements. In the development of the heavy good vehicles (HGV) matrix, significant methodological innovations were made to build it exclusively on big data. The database contained vehicle tracking data of approximately 46% of the heavy good vehicles on the roads of Hungary. The idea was, however, to find a quantifiable link between the traffic of HGVs and LCVs. On the basis of this connection, LCV productions and attractions of the settlements were estimated and with the help of the sample LCV OD database a good quality national LCV matrix was created.

keywords: light commercial vehicle; matrix of the light commercial vehicle; matrix building for the light commercial vehicle

*Tel.: +3630-2064921
E-mail address: szele@kti.hu

1. Introduction

Traffic modeling is an essential tool for planning transport networks. One of the key elements of traffic modeling is the Origin Destination (OD) matrix. In order to forecast the impacts of planned traffic developments accurately, there is a need to periodically update the OD matrices. After a long period of time or with major changes in the economic and territorial structure, it is justified to rebuild the OD matrices. Updating and rebuilding matrices gives a chance to try out new methods.

In Hungary, after 2008, the national interurban road OD matrix was updated and partly rebuilt in 2016. Both road and public transport OD matrices were made, even though the two types of matrices were not connected. During the project, we refined the applied methodology in many aspects and in some cases we put the matrix-building on a completely new basis. One of the important refinements involved the Traffic Analysis Zones (TAZ). In 2016 there were 1722 TAZs in the model, which is considerably more than in 2008 when 955 TAZs were used. The densification improved the reliability of the matrix and will facilitate the modeling work for the smaller areas. The number of TAZs has increased for three reasons: First, each larger city was a single TAZ in 2008, and now we split them into several TAZs. Secondly, we defined a large number of non-settlement traffic analysis zones, such as big factories and industrial parks, shopping malls, and logistics centers (special TAZs). Finally, some of the 2008 TAZs, which contained more settlements in 2008, were split into several more TAZs according to the involved settlements.

Two major methodological innovations were made during the work, both of which are related to commercial vehicle matrices. One of the novelties is that the matrix of Heavy Good Vehicle (HGV) is based on vehicle tracking data. Vehicle tracking systems provide information on deliveries to freight forwarders. The most important information is to know exactly where the vehicle is, but the most systems can monitor the speed, the distance travelled, the amount of fuel in the fuel tank and the open/closed state of the vehicle doors. Vehicle tracking systems are widespread worldwide, and specialized companies follow large fleets. The use of a large amount of data generated in the vehicle tracking is not yet fully formed and not wide-ranging in transport engineering. At the same time, the use of vehicle tracking data in transport engineering offers many obvious advantages. On the one hand, the most important advantage is that sampling can be large and representative. In this case, we have had about 2.5 billion GPS data of 95,000 HGVs, which is two-month travel data of about 46% of the HGVs on the roads of Hungary. It is a great advantage that the origins and destinations of the trips within the TAZs can be accurately located, and the data are not limited to Hungary, but also the external routes of HGVs are included in the database. Fortunately, in Hungary, the toll-charging system for HGVs is also based on vehicle tracking data, so the Hungarian routes of the foreign HGVs in transit were also known. On the other hand, we faced numerous problems due to the not-fully-formed methodology. This involved a cumbersome processing due to the size of the database, a very different structure of the data received, and mainly that vehicle tracking companies only collected data that were important to them, for example, vehicle category data were generally not included and the individual carriages were generally not isolated.

Creating a matrix based on vehicle tracking data is, in principle, a simple task, but one has to deal with a number of so far unknown problems. The presentation of the difficulties is not in the scope of this paper. But the vehicle tracking-based creation of the HGV matrix had a significant role in the other major methodological innovation. This novelty is the main topic of the paper, namely the building of a separate Light Commercial Vehicle (LCV) matrix. It is an old and justified demand to create a separate LCV matrix, but so far, no attempt has been made, which is likely to have two closely related causes. The first problem, which regularly occurs in the literature, is that there is simply no sufficient data on the LCV category. The other one is that because of lack of data it is difficult to create a working methodology.

Fortunately, we got a relatively wide-range data collection during the building of the road vehicle matrices in 2008, which was mainly aimed to collect data about the traveling habits of the HGVs. As a by-product, a significant LCV travel sample was also created. This data was the base for the decision to make a separate LCV matrix. However, we soon had to realize that the absence of data is a real limitation. Although we knew the typical origin and destination points of the LCVs from the sample, we did not know how many LCVs are produced and attracted in each of the settlements and how many interurban carriages an LCV would make one day. Ultimately, therefore, our main problem was the determination of the attraction and production of LCVs.

In the paper I firstly support the need to build separate LCV matrices and then I present the main findings of the literature, display the pitfalls of the methodology and finally demonstrate the most important steps of the successful methodology.

2. Why is it necessary to build a separate LCV matrix?

In traffic modeling, LCVs are conventionally treated as automobiles and included in the passenger car OD matrix, which may lead to false outcomes, mainly in the forecast. It is one of the most dynamically developing vehicle categories, so making separate LCV matrices is a justified professional need. The primary function of LCVs is freight transport, so in many ways they are closer to HGVs (for example, average travel distance, cargoing in travel chains), but also because of their size, agility, toll obligation and their not marginal role in private use, they are very close to passenger cars. Due to the importance and the specialty of the vehicle category, within a few years, it will not be possible to carry out modeling projects without a separate LCV matrix. The other cause for building a separate LCV matrix is the continuous changing of the cargo market. The cargo market is constantly adapting and transforming. According to trends, the number of HGVs is slowly decreasing and the number of LCVs is growing rapidly at the same time. The modal shift in the transportation modes is best described by the change in the number of LCVs and HGVs. However, time series data were only found in Great Britain, which is shown in Figure 1.

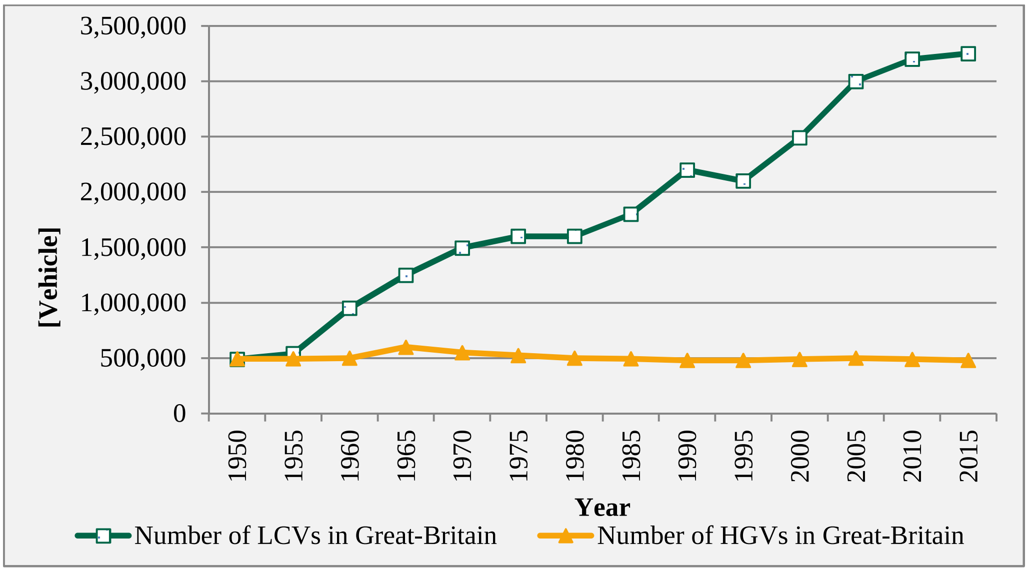

Fig. 1 The change in the number of LCVs and HGVs in Great Britain from 1950 to 2015 (Source: Clark et al. 2014)

The chart clearly shows that the number of LCVs is growing rapidly, as the number of HGVs is stagnant or slowly decreasing. In Britain, in 2015, there were 6.7 times as many LCVs as HGVs. This ratio is about 4.5 times in Hungary today. According to our estimation, there are currently 360,000 LCVs in Hungary. There is no reason to doubt that development in Hungary will be similar to that in Great Britain, which means when the observed growth rate between 2000-2015 is forecast to 2030, the number of LCVs will exceed 500 000, which represents an increase of more than 40% compared to today. This fact does not allow treating LCVs and cars together at the forecast of the matrices, as it is done conventionally. The need to create an LCV matrix is underlined by the fact that the greatest LCV traffic is mostly on the most loaded road sections. The significant growth forecast in the traffic of LCVs will also be concentrated on these road sections, so in many cases, the traffic of LCVs may be a very important component of crossing the capacity limit. Another important aspect is that the mileage of LCVs multiplies more quickly than the mileage of passenger cars. That means that the growth in the number of LCVs is multiplied by the number of kilometers travelled by LCVs. Our traffic counts show that currently LCVs are around 19-20% in road traffic, and a significant growth can be expected in this proportion.

3. Literature

Based on a summary study by Browne and Allen (2006), the most important finding in the literature is the general lack of information on LCVs. According to Browne et al. (2007), a small number of studies are investigating the activity of LCVs and LCV trips are generally not taken into account in freight data collection and modeling either. In Britain, from 1987 LCV activity surveys were carried out. Later, the last documented data collection was from 2003 to 2005 (Van activity baseline survey). During the data collection and the analysis of the collected data by Clark et al. (2014), special emphasis was put on revealing differences between the company and private use of the LCVs. According to the results, 32% of the travels for company-owned LCVs and 39% for private LCVs are for work-related purposes. For company-owned LCVs, the aim of 34% of the trips and 19% of private ones is goods collection/distribution. The average length of travel is bigger amongst the company-owned LCVs than the private LCVs. The distance of the private trips is typically smaller than business trips for both types of ownership. According to the surveys, the peak hours of the LCV traffic are between 7-9 and 16-18 o'clock. At that time, 30% of company-owned and 20-25% of private LCVs were in use. 70% of deliveries are done before 14:00 and 5% are done at night. Examining the vehicle mileage of company vehicles, the construction industry has the highest share. Wholesale and retail trade, repair, tourism, industry and mining, still have a significant proportion, all of which are responsible for 2/3 of all vehicle kilometers. In private use, the construction sector is even more prominent, but also other services appear relatively significant. Empty freighters accounted for 14% amongst the company vehicles and 28% of private vehicles. In spite of the large number of data and detailed analyses in the literature, only a few data were useful at the matrix building.

4. Used and unused data

Modeling work and matrix estimation usually start with data collection. Depending on how the data are collected, the matrix can be either based directly on travel data, or can be formulated on the sample by multiplication, or travel patterns can be created from the sample first and then the matrix will be based on these travel patterns. In our case, there was no possibility of extensive survey because no fully formed methodology or sufficient resources were available. In generating the matrix, the general lack of data did not mean a complete lack of information; we were able to access many different data sources. At the same time, some database considered to be important did not fulfil our hopes. We planned the direct use of vehicle tracking data. However, the pay-as-you-go toll service provider only collects data of vehicles above the permissible gross weight of 3.5 tons as only they pay tolls for use this way. Vehicle tracking companies do not collect vehicle category data at all, so this did not directly assist in the building of an LCV matrix. We planned to carry out an on-site survey, but there were no data on the owners of LCVs and we did not have enough resources to find them. Finally, we also intend to use the data of the wide-range household survey, which was conducted during the work. However, in the end, only 52 people (0,6%) told that last day they had an LCV transport, so the sample was too small. Typically, even the number of LCVs in the settlements was not directly accessible; even our estimation is based on two different databases.

We were fortunate to have a 2,800-record LCV OD database. In 2008, during the National OD Survey, a cordon interview was conducted in 12 Hungarian cities. The selected settlements represented the Hungarian settlements on the basis of their population, economic development and geographical location. The filtered database contained 2,737 LCV OD data. The proportion of trips abroad was 0.18%, these were not taken into account in further studies. The database (LCV OD databases) included the starting and arriving settlements of the current travel. According to the database, 17.6% of the LCV trips remain within the subregion, another 32.7% will remain within the county, but will be in another subregion, and 49.8% of the trips will cross the county border. Travels leaving the county are usually directed to the neighbouring county, with the exception of trips to and from Budapest and the travels between small settlement–small settlement and small settlement–small sized city, as they can go practically to any county of the country. Based on the database, it was possible to identify the chosen destinations for LCVs.

However, no data were available for producing and attracting the transport analysis zones or settlements. In terms of producing and attracting, it is important to mention a special characteristic of the LCVs which comes from the special function of the vehicle category. This characteristic defines the methodology and basically differs from the operation of passenger cars. The characteristic is that the LCVs cargo by going from one address to another, turning up at dozens of locations every day. Most of these trips are local travel, but there are also a large number of interurban trips. Most of the LCVs are therefore traveling in travel chains, i.e. in the case of interurban trips, traveling from one settlement to the second, then to the third, etc. Trips between origins and destinations broken by loadings have been considered as elemental travel. The conventional there-and-return trip at the passenger car matrices is less typical here, and this affects the matrix as well.

As it was mentioned, data about 46% of HGVs using a pay-as-you-go toll system in Hungary during two months was available for us. During this work, we carried out a wide-ranging traffic count that was used to measure the LCV production and attraction of settlements and special TAZs. In the traffic counts, LCV and HGV traffic were counted in 31 settlements and 11 special TAZs. National traffic counts data for 2015 were also available. The use of individual data and the most important steps for creating the LCV matrix are shown in Figure 2.

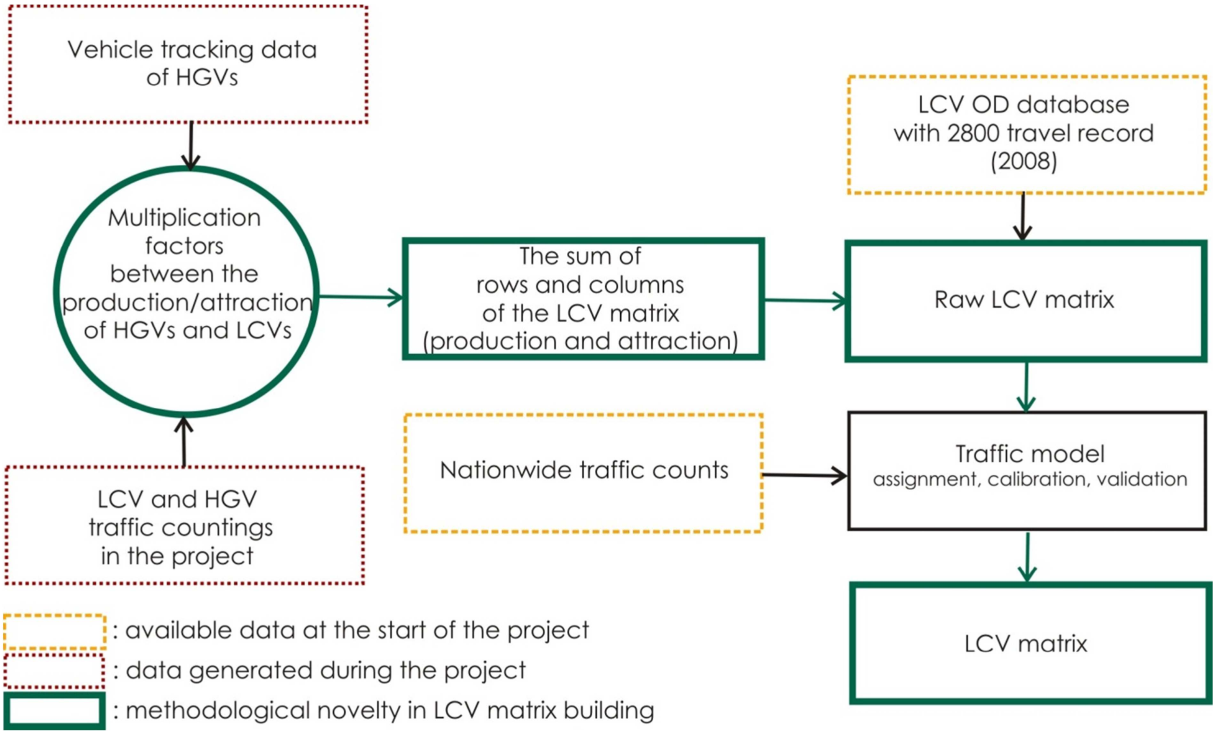

Fig. 2 The main steps of building the LCV matrix (Own source)

5. Determining the LCV production and attraction of settlements

Based on the vehicle tracking data of HGVs, the HGV production and attraction of the settlements could be estimated. The first methodological innovation was that we were looking for a connection between the production and attraction of the HGVs and LCVs at the settlements and special TAZs. The description of this connection and the determination of the corresponding multiplication factors were made possible by the traffic counts carried out during the project. In order to calculate the production and attraction of the settlements, settlement types were identified based on the location of the settlements (geographical regions), their official complex developmental classification and the population. During the counts, all the outbound and inbound routes of the LCV and HGV in selected settlements and special TAZs were counted. The multiplication factors between the traffic of the HGVs and LCVs were determined by dividing the total number of the counted arriving and departing LCVs with the total number of arriving and departing HGVs. The multiplication factor also included through traffic, which is not part of the production and attraction, this problem was treated later.

The average of the multiplication factors in the 31 surveyed settlements was 5.9, and the decisive parts of the multipliers were from 3.07 to 8.24. The very high (above 8) multiplier occurred at the economically underdeveloped settlements, where the lack of industry leads to a lack of HGV traffic. High values (5-8) were found mainly in the settlements of the Budapest agglomeration and in developed settlements. In this case, the low value of the denominator is likely to be due to a lot of restricted areas for HGVs and the lack of transport-intensive industry, and the high value of the numerator is explained by the advanced industry and services which evolve LCV traffic. Very low values (below 3.1) were measured in settlements where there is a lot of transit HGV traffic, so the value of the denominator is higher.

Out of the 31 settlements surveyed, 15 outcomes were between 3.1 and 5. In these settlements there was no significant transit traffic, so we assumed that the natural proportion of LCV and HGV traffic will also be within this range of values. Taking into account the above explanations, the overall value of the multiplication factor for settlements was finally 4.00. The multiplier of underdeveloped villages was 4.67, and in the case of developed villages it was 3.84. At shopping malls 10.40, big factories and industrial parks 1.55, and finally at logistics centers 1.20 became the multiplier value. With the multiplier factors we converted the HGV production and attraction of settlements to the LCV production and attraction of settlements, thus calculating the sum of rows and columns of the LCV matrix for settlements and special TAZs. However, to build a matrix, it was necessary to know how many LCVs from each settlement can start when they start their first trip.

6. Number of LCVs starting daily on the settlements and the number of daily interurban trips by LCVs

The values of these two parameters were estimated in one step, without any available data. Another parameter was taken into account for the estimate, namely the proportion of LCVs doing interurban trips in each city. The values of the parameter system that determines the estimation are shown in Table 1.

Table 1. The parameter system and its values that determine the number of interurban trips

	small settlements	settlements in the agglomerations	small and medium sized cities	cities	Budapest
Non-moving LCVs	5%	5%	5%	5%	5%
LCVs with only local trips	5%	5%	10%	40%	60%
LCVs with two daily interurban trips	75%	75%	70%	40%	20%
LCVs with three daily interurban trips	15%	15%	15%	15%	15%

Two-trips-daily LCVs were the most prominent category and their proportion decreased with the size of the settlement. Here we included LCVs for private use, the traffic for shopping malls and a significant part of traffic for industrial parks and settlements. Three-trips-daily LCV category meant to represent the role of LCVs in transport. The actual day-to-day LCV traffic from individual settlements was determined based on the above parameter values. So the sum of rows and columns of the LCV matrix (production and attraction), as well as the number of LCVs starting each day on interurban trips was available. The next step was to determine the matrix cell values, i.e. the travel destinations for LCVs, besides the above baseline values.

7. The travel destinations for LCVs

We made a program based on the 2800-record LCV OD database that was based on the LCV production of the settlements creating the trips. As input data, the number of LCVs starting from settlements was used, and the result was an LCV matrix. The trips in the matrix were well reflected in the travel patterns of the database; however, the individual destinations were randomly selected from the appropriate settlements. The program has made it possible to create separate sub-matrices for the different LCV categories in terms of their daily interurban trips. It was an important and useful opportunity to create quasi travel chains through the excel program. This is how we have created the trips of the three-trips-daily LCVs. Most of the trips of LCVs have been taken into account in a classic way, as one there and one back way.

8. The process and the results of the traffic modeling

The matrix created at the settlement level was transformed to the TAZ level, from that point onwards the work continued with the TAZs. In addition to the LCV traffic between the settlements, there are other LCV movements, such as cross-border trips or the traffic to special TAZs, which were created manually. In the traffic modelling we followed the steps of the classical process, such as assignment, calibration, and validation. For the validation, the trip distance distribution (Fig. 3.) obtained from the LCV OD database was very important. Due to the cordon character of the data collection, the results did not include local travel, but they did not need to create an interurban matrix. For the sake of comparison, the average trip distance of passenger cars and LCVs, jointly and HGVs with a maximum permissible weight of over 12 tonnes was obtained from the 2008 matrices.

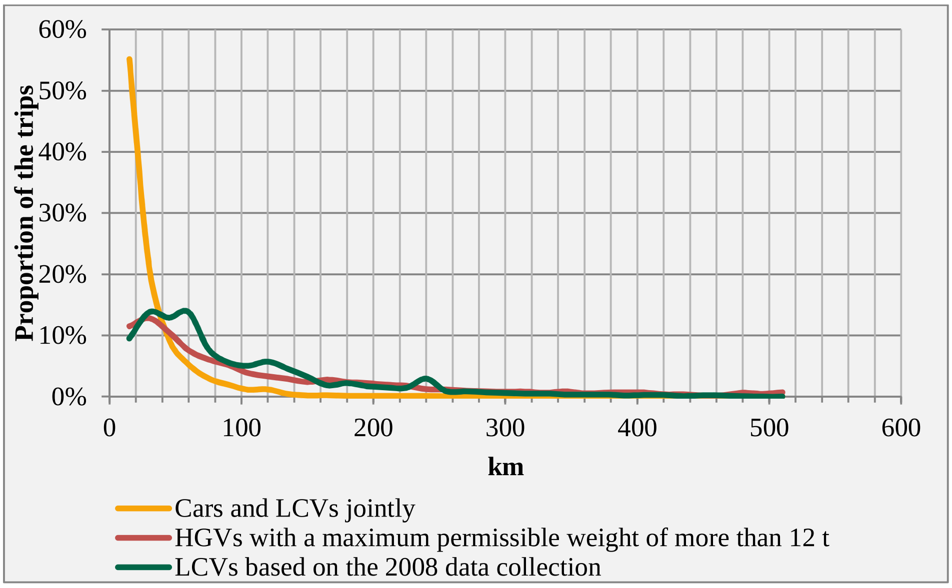

Fig. 3 Trip distance distribution of the cars and LCVs jointly, HGVs with a maximum permissible weight of more than 12t and LCVs based on the 2008 data collection (Own source)

The figure clearly shows that the trip distance distribution of LCVs is very similar to the HGVs with a maximum permissible weight of over 12 tons: there are few short trips, relatively many long trips of up to several hundred kilometers and travels between 20-150 km cardinality. The average trip length of LCVs was 93 km.

To finish the matrices, we conducted calibration and validation for seven times. According to the results, approximately 360,000 LCVs take about 296,000 interurban trips on an average working day in Hungary. The total mileage was 20,745,000 km. During the calibrations, the average trip distance increased steadily, reaching a value of 70.1 km in the final matrix, which is significantly higher than the 45.8 km of passenger cars, but significantly lower than the aimed 93 km, the result of the data collection in 2008. During the calibrations and necessary repairs, we did an attempt to significantly increase the average trip distance of the matrix to reach the 2008 data, simultaneously with the approximation of the trip distance distribution in Figure 3. The nationwide traffic counts LCV traffic data were used for validation. Figure 4 shows the trip distance distribution of the calibrated LCV matrix comparing with the result of the LCV OD database 2008.

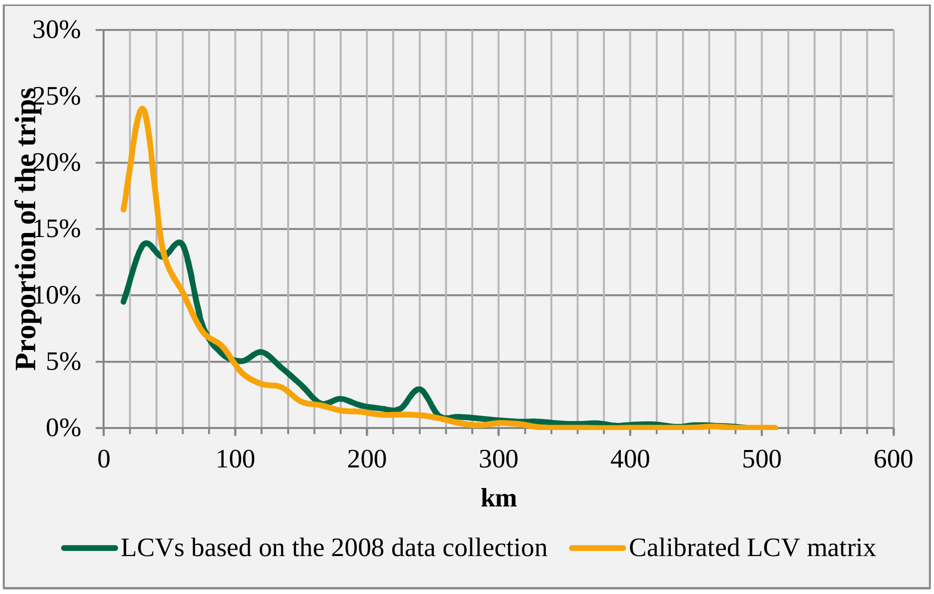

Fig. 4 Trip distance distribution of the measured and calibrated LCV trips (Own source)

The production and attraction calculated for settlements dropped considerably from 588,000 to 296,000 during calibration. Behind the general decline, there is a slight increase (in Budapest, in international travel and in the special TAZs), a small number of matches (mainly in settlements around Budapest) and a generally significant decrease. The productions and attractions of many settlements (i.e. the row and column sum of the matrix) fell to three-fourths part of the original, while the production and attraction of Budapest grew by 44%. Behind the significant drop in my view, there were two main reasons.

The first reason is that in the absence of a wide-range database, the road traffic data of the nationwide traffic counts were used for calibration. This caused problems as the results of the nationwide traffic counts and the road traffic data of the traffic counts performed by us in 123 cross sections between 4 and 20 hours showed significant differences. The data from nationwide traffic counts were generally significantly smaller, agreed in some places and in some places exceeded the results of our own counts. On average, traffic data in nationwide traffic counts accounted for 78.2% of our own counts. The difference is perhaps explained by the fact that the LCV vehicle category is generally quite difficult to distinguish in traffic counts. Automated count stations can not distinguish a significant part of LCVs from cars, thus calculating the proportion of LCVs by model counts. It is a fact that on the basis of the results of the nationwide traffic counts, the upper limit of the traffic of LCVs on the Hungarian motorway and highway network is the result of the current calibrated matrix, which is at least 20% lower in LCV traffic over the entire network compared to the traffic in reality.

The other, probably more important reason is the methodology used and it can also be seen as a criticism at the same time. During matrix building, we generated essentially urban-level production and attraction values, but we did not pay enough attention to the proper directions of production and attraction, so the LCV traffic from the agglomerations towards the cities was significantly underestimated. The methodology mainly focused on long-distance traffic due to the typically large average trip distance. Such trips mainly loaded the motorway and highway network and are displayed on the road linking the motorway and highway network with the settlements. The calibration cross sections were mainly highlighted on the motorway and highway network, so calibration was mainly related to long distance traffic and according to the results of nationwide traffic counts, it is much smaller than we assumed. Generally, there were no calibration cross-sections for the suburban LCV traffic, so this traffic remained small. In large part, this could have led to a significant reduction in the production and attraction of the settlements.

In my opinion, both the methodology of the matrix building and the results of the nationwide traffic counts are basically appropriate. The created matrix during the calibration process corresponded to the results of the nationwide traffic counts, so the results are suitable to reach the nationwide goals. However, to have appropriate results for smaller regions, further efforts need to be made.

9. Conclusion

Findings indicate that production and attraction of the two vehicle categories (LCV, HGV) are closely related and vehicle tracking data of HGVs can be used to estimate the production and attraction of LCVs. On this basis, trip distribution of LCVs, as well as the production and attraction of the settlements can be analysed. Conclusions are expected to be highly transferable, as this is one of the first studies on the estimation of a separate LCV OD matrix.

The results show that it is possible to create an LCV matrix without a wide-range database, but because of it, in the methodology there are many estimates that need to be clarified. According to the experience of this work, for example, the proportion of the three-trips-daily LCVs should be considered. It would be good to get results closer to the real trip distance distribution and average trip distance. It would be important to involve the special TAZs and foreign TAZs into the three-trip travel chains and to think about creating four-trip travel chains as well.

In the next modeling projects, it is useful to consider that the measurements of nationwide traffic counts show a significant 20% reduction in traffic compared to what is actually happening. The reasons for this should be considered deeply. However, the most important development requirement is to find the right balance of suburban and long-distance traffic in production and attraction, because we believe that the production and attraction values defined by the methodology are basically correct.

Due to the weight and rapid development of the LCV vehicle category, within a few years time, it will not be possible to carry out meaningful traffic modeling work without the use of a separate LCV matrix. We still do not know much about the LCVs — data collection, and based on these, a basic database would be needed. We hope we have contributed to encouraging further efforts with our methodological experiment.

Acknowledgements

This project (Enhancing the participation of Hungary in the TEN-T Corridors) has been funded by the EU Connecting Europe Facility.

References

Browne M., Allen J. 2006, Best Urban Freight Solutions II (BESTUFS II), Best Practice in data collection, modelling approaches and application fields for urban commercial transport models, Urban freight data collection - synthesis report, Deliverable 3.1, BESTUFS (Best Urban Freight Solutions) II project, TREN/04/FP6TR/S07.31723/506384

Browne, M., Allen, J., Woodburn, A., Piotrowska, M. 2007, Light Good Vehicles in Urban Areas, Literature Review WM9, Transport Studies Group, University of Westminster

Clark, G., Johnson, A., Nankivell, J., Turpin, M. 2014, Van travel trends in Great Britain, RAC Foundation, London

Department of Transport: Van activity baseline survey 2008, http://webarchive.nationalarchives.gov.uk/20110503210608/http://www.dft.gov.uk/pgr/statistics/datatablespublications/freight/vanactivitybaseline08/ (last access: 30/06/2017)