EJTIR

Issue 17(1), 2017
pp.46-62
ISSN: 1567-7141
tlo.tbm.tudelft.nl/ejtir

Freight demand generation on commodity and loading unit level

Koen Mommens[1]
MOBI, Vrije Universiteit Brussel, Belgium.

Tom van Lier[2]
MOBI, Vrije Universiteit Brussel, Belgium.

Cathy Macharis[3]
MOBI, Vrije Universiteit Brussel, Belgium.

Freight sustains our daily lives and economy. Information on its characteristics, its production and consumption locations and its modes of transport are consequently of a crucial importance for private and public decision makers. This paper presents a freight generation model for the Belgian territory. Based on data gathered on transport flows by commodity type and loading unit and data on population density and business establishments with their characteristics, generated and attracted freight volumes were obtained for 4934 zones subdividing the country. A generalized linear regression analysis with log link was used to do so. Both generated and attracted volumes are connected to one another by a conditional probability function, resulting in an origin-destination matrix. The analysis is to our knowledge unique as modelled volumes and flows can be distinguished by commodity types and loading unit, and this at very detailed geographical scale. This will lead to new in-depth analyses and added value, including effects of loading unit dependent logistics cost structures..

Keywords*: activity based, freight demand generation, loading unit, commodity type.*

1. Introduction

The large amount of freight transport movements is causing different negative effects (congestion, noise, accidents, pollution, etc.) on both local and global scale. In order to take into account the economic, ecologic and social effects of freight transport, detailed transport system analysis is needed based on freight transport models which take into consideration vehicle characteristics (mode, loading capacity, loading rate, etc.), network characteristics (link type, speed limits, etc.) and local characteristics (receptor densities, demographical characteristics, etc.). The first step in the development of a freight transport model is generation of freight transport demand. Despite wide spread use of freight generation models, it has to be noted that their methodologies are mostly only briefly explained and their accuracy is rarely detailed (Alho and de Abreu e Silva, 2015). Generating freight demand is no straightforward process, and there is no ultimate best methodology. One reason for it is that good data on freight volumes is scarce and very fragmented. National statistics bureaus in the European Union need provide data on commodity flows to EUROSTAT, which publishes them on NUTS2 level. On national level,

[1] Pleinlaan 2, 1050 Brussels, Belgium T: +32 2 629 21 66 E: kmommens@vub.ac.be
[2] Pleinlaan 2, 1050 Brussels, Belgium T: +32 2 629 20 67 E: tom.van.lier@vub.ac.be
[3] Pleinlaan 2, 1050 Brussels, Belgium T: +32 2 629 22 86 E: cathy.macharis@vub.ac.be

however, commodity flow data are often available on NUTS3 level, like it is the case in Belgium and the Netherlands. Still, NUTS3 level is considering entire cities as one zone, which makes it impossible to distinguish commodity flows to or from urban industrial area, CBD's, ports or airports. Moreover, freight can differ a lot in terms of commodity type, weight, volume, value, loading unit, etc. Such detailed information is important for analyses, but often missing in available databases. Most disaggregated and detailed information on freight (flows) is known at enterprise or establishment level, and only by the concerned enterprise itself and/or their logistic service provider if applicable. These data are often unprocurable, referring researchers to aggregated figures on flow-level between larger geographical entities. Those aggregated flows are mostly based on surveys and input-output tables (de Jong and Ben-Akiva, 2007). By combining the aggregated information gathered from transport surveys with disaggregated data on establishment level – which is mostly easier to get than freight transport data – one is nevertheless able to generate freight flows and their respective characteristics between small geographical entities. Using this approach, new valuable input for freight models and consequently new, more precise analyses is produced. In this paper, the problem of estimating freight volumes based on establishment characteristics is addressed and applied to the Belgian territory. The paper presents two main contributions; the incorporation of nine different types of loading units and this in combination with commodity type on a very detailed geographical scale. This incorporation will enable to include loading unit dependent logistic costs (like transhipment costs, storage costs and transport costs) into the modelling process. Secondly, the regression method assumptions are addressed, which is rarely done in freight generation modelling. Violating the assumptions leads to wrong results and low accuracies. In order to do so, first the existing literature on freight generation modelling is studied, with a focus on activity-based generation models. This will be presented in Section 2. Combining the experience retrieved from literature with the data gathered for Belgium, a state-of-the-art methodology to generate volumes is applied in Section 3. Results will be described in Section 4, after which the paper ends with takeaways and conclusions.

2. Literature study

Freight generation models are common practice in freight modelling. Despite their importance within the freight model process – according to the garbage-in-garbage-out-principle – the used methodology and accuracy are rarely addressed. Generally, two techniques can be distinguished; cross classification analysis and multiple regression analysis (Bastida and Holguin-Veras, 2009). Cross classification procedures compute changes in one variable (freight volume/trips) when other variables (employment, surface, sales, etc.) are brought into account. Although similar to multiple regression analyses, this technique is non-parametric as it does not take into account the distribution of the individual values. The works of Bastida and Holguin-Veras (2009) and Oliveira Neto et al. (2011) are forming the first applications of a cross classification technique for the generation of freight, in contrast to earlier applications for passenger transport. In their work, Bastida and Holguin-Veras compare the generated results from the cross classification technique with the ones obtained by using an ordinary least square (OLS) procedure. Their conclusions for the New York Metropolitan Area are that both procedures can be used to forecast freight demand as well as to establish the dynamics between commodity type, employment, economic activity and year sales. Multiple regression analyses, like OLS, are widely used for the generation of freight. Rowinski et al. (2008) compared total employment, commodity specific industry employment, population and truck vehicle miles travelled as disaggregation variables for the counties of New Jersey using OLS. They observed large differences between used variables, between generation and attraction of volumes and between transport modes. Piotte and Jourquin (2011) used in their freight generation model for the Walloon Region (Belgium) economic activity and number of employees as independent variables in their OLS. Zhang et al. (2003) also used employment and population to break down state level origin-destination data to county level data for the Mississippi state (USA). Lim et al. (2014) used population, employment, farmland, sales and net annual electrical generation as independent variables for their OLS based freight generation model for California (USA). By grouping different commodity type categories, they obtained generally higher R-squared values. Khan and Machemelh (2015) performed an OLS for the Williamson County (Texas – USA) considering employment, economic activity, population and land use. Low R^2 values were obtained, suggesting that other explanatory variables could be needed. In Wisetjindawat and Sano (2003) their study surface space was taking into account as independent variable for the OLS, together with economic activity and employment. However, still low R^2 values were obtained. Rather low accuracy is a general problem which freight generation models are facing – referring to Khan and Machemehl (2015), Bastida and Holguin-Veras (2009), Wisetjindawat and Sano (2003) and Ahlo and de Abreu e Silva (2015). Both Khan and Machemehl (2015) and Shin and Kawamura (2005) suggest to consider other independent variables. However, above overview illustrates that despite the differences in independent variables (economic activity, gross floor space, employment, land use and demographics), low accuracy remains an issue. Additionally Shin and Kawamura (2005) propose to target relatively simple supply chains and businesses in order to improve the models accuracy. At the same time, the authors are pointing the limited data availability. This limits the possibilities both in terms of independent variable choice as in terms of targeting specific supply chain or businesses. Disaggregated data should be preferred for regression analyses (Douglas and Lewis, 1970). Disaggregated data of freight volumes and transport are, however, scarce. This is due to the complexity of freight flows (commodity type, costs, loading unit, supply chain types, etc.) and the fragmentation of information amongst the different actors (shippers, carriers, receivers, distribution centres and transport agencies). Zhang (2013) describes this complexity in her research, as she identifies multi-actor, multi-scale, multimodality and different commodity types (including different loading units) in her modelling. She implements and focusses her model on containerized freight in the Netherlands. Davydenko (2015) developed on his turn an freight generation model which targets the variety in logistics chains; namely the production-consumption flow, production-distribution flow and distribution-consumption flow. Each of those chains have different characteristics and different actors. Each of those actors has only a partial view on the freight transport system (Jaller and Holguin-Veras, 2013). Therefore, it is nearly impossible to get disaggregated freight flow data which enable both to model on a micro level scale and to include behavioural aspects. Additionally for relatively detailed data, commercial and privacy are being an issue, and often important financial resources are needed to get the available detailed data (Holguin-Veras et al., 2010). Jaller and Holguin-Veras (2013) made an overview on the freight data needed for different modelling purposes, the actors who are having the data and the financial costs related to gathering those necessary data (through surveys). They concluded that collecting all data needed requires too many resources, and they propose modular approaches. This is why only aggregated data are gathered and consequently used and often disaggregated afterwards using multiple regression methods. When doing so one needs to take into consideration that the produced disaggregated data are not intended to be a substitute for local data for reasons of lack of field verification and data synthesis nature (Rowinski et al., 2008). Moreover, one need to be aware that the used data meets the assumptions of the used method. Most commonly linear regression models are used, as they are simple in terms of interpretation and explanation of the effects of exogenous variables. Violations of the assumptions of OLS and other methods could lead to inaccurate parameter estimates (Douglas and Lewis, 1970). The assumptions are rarely addressed in literature.

Consequently, the choice of the regression technique depends on the used / available data and the degree to which those data meet the assumptions of the regression technique. In literature, a distinction can be made between linear regressions, Poisson log-linear generalized linear models and other generalized linear models when other link- and family functions are specified.

Besides the methods, one can also distinguish two approaches in freight generation, namely; the commodity based approach and the vehicle trip based approach (Khan and Machemelh, 2015). Vehicle trip based models focus on a correct generation of vehicle trips. As such they target a better representation of trips, empty legs and transport logistic decisions. Commodity based models focus on their turn on the generation of commodity flows, and via shipment size and vehicle allocations translate the commodity flows in vehicle trips. This approach does not directly model empty trips and vehicle activities. Moreover it does have difficulties with capturing the full complexity of the freight supply chain, but this holds true for the vehicle trip based approach as well. In turn, the commodity based approach is by consequence better in representing economic behaviour, commodity influences on trip choice and most importantly mode choice. Consequently the commodity based approach is better in representing inventory logistics and intermodal chains and is therefore recommended for situations where different transport modes are considered as competing transport alternatives (Khan and Machemelh, 2015). The output produced from the freight generation model presented in this paper will mainly be used for national, multimodal subjects. Therefore a commodity based approach is preferred.

This literature overview illustrates three points. Firstly, the used methodologies, their assumptions and the accuracies of the models are rarely properly addressed. Secondly, low accuracy is a general problem in the freight generation field. Suggested solutions to improve the accuracy are focussing on the integration of other independent variables and on reducing the analysis to specific supply chain type or industries. This leads to the third point, which is the scarceness of accessible, detailed data necessary to develop freight generation models. In the next section, one will discuss the data availability and the data used for this research. The methodology used for the freight generation will be described as well, addressing both the used methodologies and accuracy.

3. Methodology

The methodology is illustrated in a flow diagram in Figure 1. It consists of different steps. The first step is gathering the input data. The input data needs to be rescaled onto properly created Traffic Analysis Zones (TAZ) for the base year 2012. In total 4 934 zones are subdividing the Belgian territory. All analyses are performed in terms of volumes, with tonnes as units, no monetary or volumetric (m^3) units are used. A commodity based approach was used in order to be able to use the generated volumes for modal shift analyses. The aim of the model is to produce a representative origin-destination matrix containing information on commodity type, value of the goods, loading unit and shipment size. This is to enable transport modelling applications. Including used loading unit – beyond the containerized / non-containerized goods differentiation – is new in literature. Moreover, this research tackles both the issue of addressing methodology and accuracy and the national application on a very detailed geographical scale.

The different steps of this research are the data gathering process and data description, the regression analyses based on the input data, the introduction of the used loading unit and finally the production of an OD matrix on traffic analysis zone level for each shipment. This section will follow the different steps of the methodology accordingly, like also presented in Figure 1.

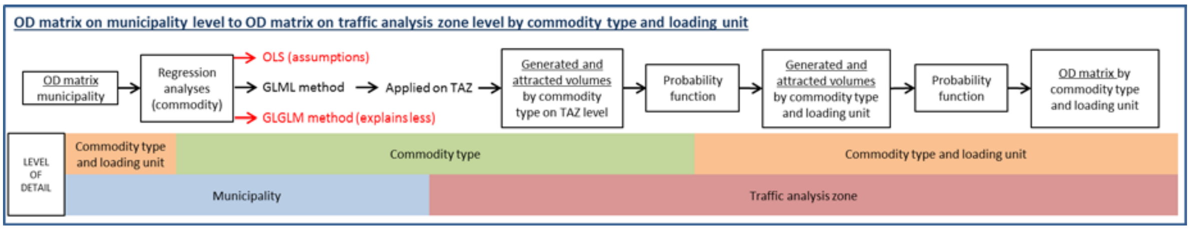

Figure 1. Research structure

3.1 Data collection

The current transport demand for freight transport models in Belgium is based on various aggregated input data, varying amongst the transport models. The Strategic Freight Model for Flanders, a region in Belgium, generates freight on NST level based on employee levels for 518 zones (Verkeerscentrum Vlaanderen, 2006). The freight model for Wallonia – another region in Belgium – developed in the DIDAM project is based on data on NST/R level from the French ECHO study (Guilbault et al., 2008) which were applied for 262 TAZ in Wallonia (Piotte and Jourquin, 2011). The Brussels freight transport model will apply on its turn the values and methodology of the French FRETURB model for 200 TAZ (Mobiel Brussel, 2014). Maes (2013) used the ADA methodology developed by Ben-Akiva and de Jong (2008) to generate her freight demand on NST/R level for 308 TAZ in Flanders. Finally, the national PLANET model generates volumes based on the relationship between the volume generation and the value of domestic production and the value of imports (Desmet et al., 2008) for freight generation for 43 TAZ. Except for the differentiation in terms of containerized and non-containerized freight, none of the Belgian models is considering the used loading unit. Also in transport models for other countries and regions loading unit is never considered properly (Mommens et al., 2016). Zhang (2013) identifies the differences between solid and liquid bulk and containerised freight, where-after her model is applied on containerized goods. Although her research states the importance, her proposed differentiation does not include all relevant loading units (like palletized freight, mobile units and parcels). The proposed model in this paper is taking into account nine types of loading units and ten different commodity type (NST/R), moreover at more detailed geographical level than current Belgian transport models and without relying on foreign input data. Therefore, the existence and availability of different alternative data on freight transport for Belgium was checked. Moreover, contact was made with different experts from around the globe in order to get some insights from their experiences in the field of gathering data and modelling freight.

This approach resulted in a quest for data towards different agencies and interest groups, as information and data are dispersed amongst them. Table 1 gives an overview of the collected data, their geographical level of detail and their source. All mentioned data were collected for the base year 2012.

Table 1. Data overview for freight generation model.

Data	Level of detail	Source
Business establishment location	Address – XY coördinates	National social security office,
Number of employees	XY coördinates establishment	National social security office,, National statistics agency
Economic activity	XY coördinates establishment	National social security office, National statistics agency
Gross floor space	XY coördinates establishment	National geographical institute, National social security office
Population density	Traffic analysis zones	National statistics agency
Volume by commodity type and loading unit	Municipality and gateways	National statistics agency, Port of Antwerp, Ghent, Zeebrugge and Ostend, Airport of Brussels, Liège, Antwerp and Ostend, Vlaams Verkeerscentrum, W&Z SPW Wallonie and NV De Scheepvaart.

Generated and attracted volumes in tonnes distinguished by commodity type and loading unit were obtained at origin-destination-level between municipalities based on the yearly national transport survey conducted by ADSEI (2012), while establishment data (activity, location, number of employees, gross floor surface) were available on coordinate-level, which means that addresses of all establishments were obtained. Moreover, also demographic figures and data on carriers (location(s), number of employees, fleet size and fleet characteristics) were obtained and used. All data are accounting for the base year 2012. Ten types of commodities were distinguished according to the NST/R categories (agricultural products and live animals; foodstuff and animal folder; solid mineral fuels; petroleum products; ores and metal waste; iron, steel, and non-ferrous metals including semi-manufactured products; crude and manufactured minerals and building materials; fertilizers; chemical products and vehicles, machinery and other goods) and nine types of loading units (solid bulk; liquid bulk; containers; other containers; pallets; slings; mobile units; other mobile units and other). Combined, all the data allowed performing different regression analyses to explain generated and attracted volumes by commodity type and by loading unit for the traffic analysis zones.

Additionally, data – on the same level of detail in terms of commodity type and loading unit – were collected for the different gateways. Those gateways are defined as: border crossing for major road, inland waterways and railways; maritime ports (Antwerp, Ghent, Zeebrugge and Ostend) and international airports (Brussels, Antwerp, Liège and Ostend). As such interactions between Belgium and the rest of the world are included with regard to import, export and transit.

The data gathered have different levels of detail (XY coordinates, traffic analysis zones, municipalities). In order to link the data and enable to generate freight volumes and flows on traffic analysis level for transport modelling purposes, one aggregated the data on municipality level. The regression analyses, which will be discussed in the next paragraph, are performed on municipality level and then applied on traffic analysis zone level. The traffic analysis zones are based on statistical geographical units which allow making linkages to the different NUTS-levels. Two different kind of statistical geographical units are used in defining the traffic analysis zones. For urban areas – defined at the municipality level by the study of Luyten and Van Hecke (2007) – the so-called quarters are used. Quarters represent contiguous collections of statistical sectors with similar demographic, administrative and economic characteristics. Multiple quarters are forming so-called sections. Those sections are mostly corresponding to the former municipalities before their fusion. For the rural areas the sections were used as delimitation for the traffic analysis zones. Multiple sections are forming municipalities.

3.2 Regression analyses

As already mentioned, first the establishment data and population density data (number of employees, surface and activity) were aggregated on municipality level. This is the administrative level for which data on generated and attracted volumes per commodity type and loading unit combination were acquired. Then, linear regression analyses are performed with population density and number of employees and surface per activity as independent variables and respectively generated and attracted volume per commodity type and grouped commodity- and loading unit types as dependent variables. Activity is subdivided into 13 different sectors which correspond to an aggregation of the NACEBEL 2008 level I classification. Table 2 presents the 13 categories.

The regression analyses were performed by commodity type. Initially, the intension was to perform regression analyses for each commodity type – loading unit – combination. This was however not possible due to high number of zero-observations, as illustrated in Table 3 for the volume generation. Volume attraction has to cope with similar numbers. The high number of zero-observations is logic as many commodity type – loading unit – combinations are only occasionally observed. A grouping strategy – resulting by grouping NST/R category 0 (agricultural products and live animals) and category 1 (foodstuff and animal folder); 2 (solid mineral fuels) and 4 (ores and metal waste) and 7 (fertilizers) and 8 (chemical products), and by grouping loading units 2 (containers) and 3 (other containers); 6 (mobile units) and 7 (other mobile units) and 5 (slings) and 9 (others) – was tried, but the regression analyses did not produce better results than the once which will be described in the following paragraphs. If only commodity type was taken into consideration in the regression analyses, low numbers of zero-observations are obtained (also illustrated in Table 3 by the column 'All').

Table 2. Data overview for freight generation model.

Data	Source
Accommodation and food service activities	Accommodation and food service activities
Agriculture, forestry and fishing	Agriculture, forestry and fishing
Construction	Construction
Financial and real estate sector	Financial and insurance activities + Real estate activities
Information and communication	Information and communication
Manufacturing	Manufacturing
Mining and quarrying	Mining and quarrying
Professional, scientific and technical activities	Professional, scientific and technical activities
Services	Administrative and support service activities + Public administration and defence; compulsory social security+ Education + Human health and social work activities
Transportation and storage	Transportation and storage
Utility companies	Electricity, gas, steam and air conditioning supply + Water supply; sewerage; waste management and remediation activities
Wholesale and retail trade; repair of motor vehicles and motorcycles	Wholesale and retail trade; repair of motor vehicles and motorcycles
Others	Arts, entertainment and recreation + Other service activities + Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use + Activities of extraterritorial organisations and bodies

Table 3. Number of zero-observations for volume generation out of population of 589 municipalities.

		Loading units									
		0	1	2	3	4	5	6	7	8	All
NST/R	0	419	181	353	451	178	469	384	548	201	64
	1	386	328	438	502	253	536	586	561	308	141
	2	589	162	509	558	435	555	566	579	362	123
	3	417	299	466	528	320	476	577	578	290	143
	4	474	432	477	538	477	589	548	588	441	273
	5	589	530	563	434	476	583	584	385	290	292
	6	429	117	374	487	207	452	533	553	203	44
	7	509	471	581	587	560	585	588	589	557	417
	8	374	211	364	471	208	419	550	555	205	62
	9	448	206	292	420	123	339	275	497	114	26

Regression analyses are often used in freight generation models. The assumptions on which the respective regression analyses are based need to be taken into consideration in order to avoid inaccurate parameter estimates. Often the assumptions are not mentioned in literature, wherefore it cannot be checked if they were met. OLS regression is the dominant technique in freight generation modelling. The assumptions of OLS regression are the following:

- Normality of the error distribution
- Homoscedasticity of the errors

- Statistical independence of the errors
- Linearity and additivity of the relation between dependent and independent variables

The assumptions for an OLS on this analyses were checked and showed that both the assumptions on homoscedasticity and linearity were violated, as illustrated in Figure 2. Moreover, although the Durbin-Watson coefficient is larger than 1 for all analyses, spatial dependence of the errors was often observed as for example municipalities with ports (Ostend, Ghent, Antwerp, Liège) and major distribution centres (Halle, Ninove, Kontich, Sint-Katelijne-Waver, Nivelles) are showing studentized residuals outside the -2 to 2 interval (Figure 3). Therefore OLS analyses could not be applied for our data.

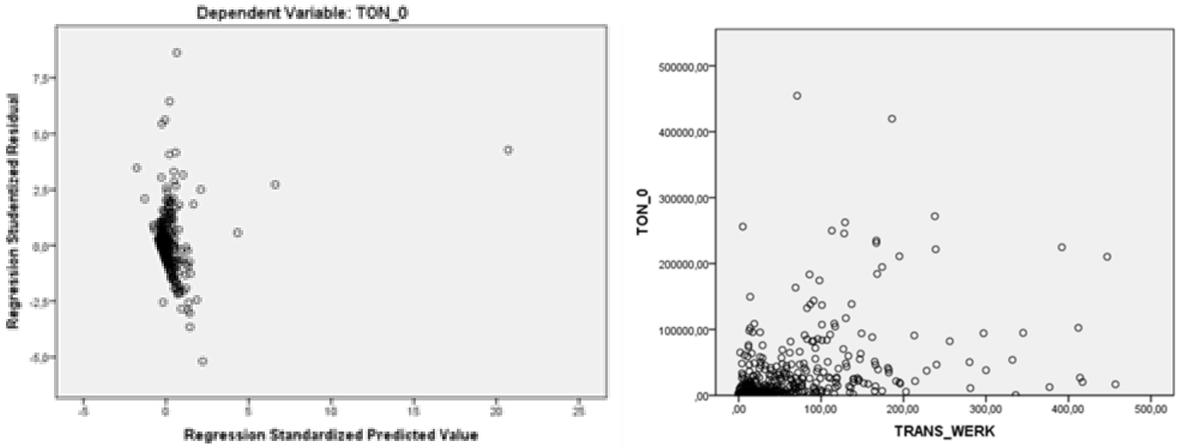

Figure 2. Scatterplots used to check the assumptions on homoscedasticity (left) and linearity (right).

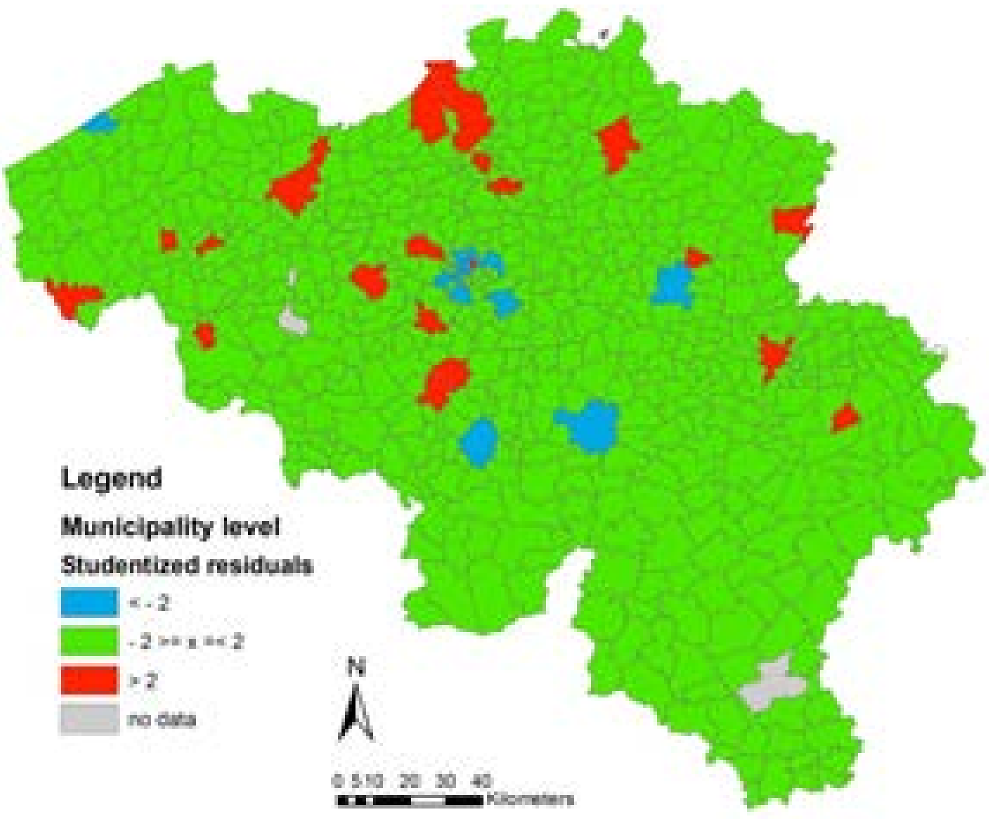

Figure 3. Studentized residual for OLS regression analysis for agricultural products and live animals on municipality level.

In order to cope with these violations two types of generalized linear regression models were used for the same dependent and independent variables. Generalized linear model allows the response variables to have error distributions other than a normal distribution. In that sense generalised linear models are literally generalizing OLS as they permit a relationship between the linear model and the response variable through a link function and as they allow the magnitude of each measurement's variance to be a function of its predicted value. Two link functions were retained; namely a generalized linear regression model with log link (GLML) and a gamma log generalized linear regression model (GLGLM). Generalized linear regression models have four main assumptions:

- Correct specification of the link function
- Statistical independence of the observations
- Correct specification of the distribution of the observations
- Homoscedasticity of the errors

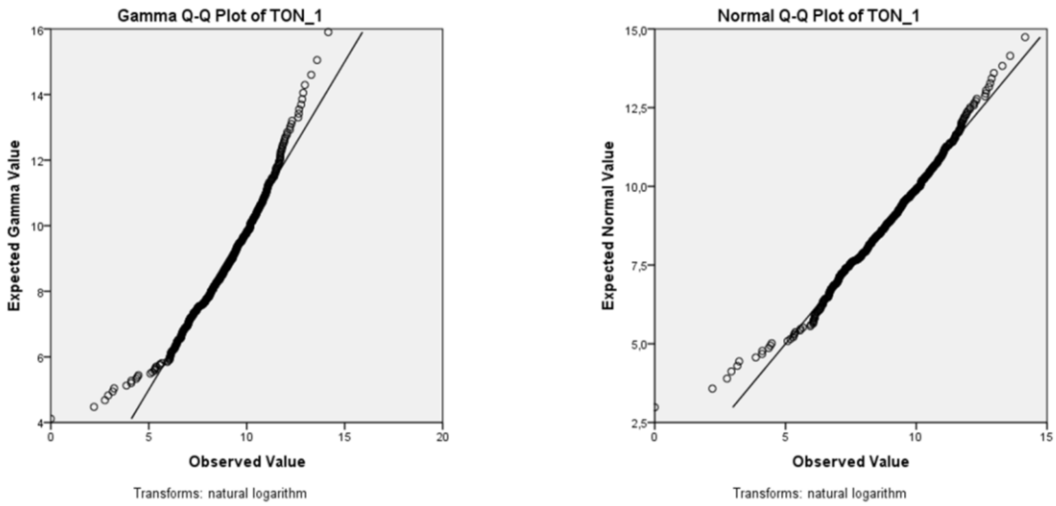

Figure 4. Q-Q plots for commodity type 'foodstuff and animal folder' according the GLGLM regression technique (left) and the GLML regression technique (right).

All four assumptions were met by the gathered data. Both GLML and GLGLM were compared in terms of Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Likelihood Ratio Chi-square values. Distribution, R-squared and residual diagnostics were looked at. Plots enabled to control the distributions and to check for outliers. Figure 4 illustrates for example that the quantiles of the observed values approximately meet the quantiles of the expected values better when the GLML regression technique is used, meaning that the normal distribution fits better for the commodity type 'foodstuff and animal folder' than the gamma distribution. Table 4 illustrates the AIC value, BIC value and Log Likelihood for the GLML and GLGLM method for both the generated and attracted volumes. Moreover, we also looked at leverage points, outliers and influential points. By mapping residuals, we checked for geographical influences. The GLML technique performed better than the GLGLM for each performed analyses. Therefore the generalized linear regression model is chosen as model, and further analyses are based on its results.

Table 4. AIC values, BIC values and log likelihood for generated and attracted volumes following the GLML and the GLGLM method.

	Generated volumes						Attracted volumes					
NST/R	Log Likelih.		AIC		BIC		Log Likelih.		AIC		BIC	
	GLML	GLGLM	GLML	GLGLM	GLML	GLGLM	GLML	GLGLM	GLML	GLGLM	GLML	GLGLM
0	-972	-1057	1959	2129	1994	2159	-837	-902	1692	1824	1732	1867
1	-846	-894	1709	1804	1747	1837	-824	-895	1664	1804	1699	1834
2	-892	-949	1792	1906	1808	1922	-954	-1035	1920	2080	1946	2102
3	-905	-969	1818	1948	1834	1968	-829	-1048	1676	2106	1715	2128
4	-612	-656	1231	1319	1246	1330	-594	-656	1197	1318	1212	1329
5	-575	-609	1159	1226	1173	1241	-667	-685	1341	1382	1357	1405
6	-1023	-1119	2058	2249	2084	2270	-893	-1069	1802	2149	1836	2171
7	-357	-390	720	786	729	795	-345	-359	695	725	705	739
8	-934	-1004	1884	2023	1918	2058	-744	-774	1504	1567	1539	1606
9	-910	-978	1838	1972	1877	2007	-814	-946	1648	1902	1692	1923

Analysing the extent to which the model approximates the original data, R-squared values are spread between 0,313 and 0,691 for the generation of commodities and a slightly better 0,297 and 0,789 for the attraction of commodities. Those are relatively low accuracy values. They are in line with other state-of-the-art studies, as low accuracy is a problem freight generation models in general face – like Khan and Machemehl (2015), Bastida and Holguin-Veras (2009), Wisetjindawat and Sano (2003) and Ahlo and de Abreu e Silva (2015). Table 5 and Table 6 show the R-squared values, parameter estimates and t-stats for the GLML regression analyses by commodity type. Abbreviations 'Empl' and 'Surf' stand for the number of employees and the gross floor surface by activity sector. Only the significant parameters (at 10% significance level) were kept, as many parameters have very low t-stats which may cause overfitting with low explanatory power. This is illustrated by NST/R category 7 (fertilizers). By only taking into consideration the significant parameters, fertilizers are confronted with extremely low R-squared values (0,038 for the generation and 0,128 for the attraction). In order to allow further analyses on all NST/R categories, one opted to keep all parameters for category 7, increasing R-squared values to 0,378 for generation and 0,297 for attraction.

It is difficult to address this issue. Shin and Kawamura (2005) propose to target niche markets or specific supply chains. This could possibly increase the accuracy. However, this suggestion is conflicting with the aim of freight generation modelling on a national scale, as its goal is to predict freight flows to be able to introduce them in transport models and other applications. Khan and Machemehl (2015) and Shin and Kawamura (2005) suggest to look at other independent variables. Currently economic activity, employment, gross floor space, land use and population density are considered in literature. The introduction of those independent variables did not lead to big improvements in the accuracy. It is therefore doubtful that new independent variables will solve the problem. However, if one does want to test a new independent variable, it could be interesting to consider transport infrastructure availability. The possibility to introduce new independent variables is also depending on the availability of data. This lack of available disaggregated data can be considered as the main cause of low accuracy of freight generation models. Despite the fact that it has an impact on the R-squared values of this research, results illustrate that the most important variations are well explained by the model – which is the main goal of a national freight generation model. As such, areas containing seaports, airports, inland ports or main distribution centres are representing realistic higher volumes according to the model. Also cities have significantly higher and different freight patterns than rural areas. And finally, regional differences in terms of freight categories are well explained by the model too.

Table 5. Freight generation parameters.

	0	1	2	3	4	5	6	7	8	9
n	525	448	466	446	316	297	545	172	527	563
R-square	.559	.529	.319	.328	.313	.396	.505	.378	.624	.691
Intercept	7.56 (24)*	8.28 (10.8)*	6.904 (27.48)*	4.954 (15.39)*	6.108 (45.52)*	3.809 (10.09)*	6.112 (26.39)*	4.714 (1.17)	4.231 (9.92)*	5.427 (15.87)*
Population_den	-.354 (-6.05)*	-.273 (-3.28)*						.03 (.13)		-.206 (-3.74)*
Agriculture_Empl	.245 (6.03)*	.292 (5.75)*						.011 (.07)		
Agriculture_Surf								.261 (2.58)*		
Mining_Empl		-.169 (-2.49)*	.613 (8.99)*				.528 (8.14)*	-.196 (-.95)		
Mining_Surf					.0.074 (2.22)*			.045 (.58)		
Manufacturing_Empl	.393 (8.05)*	.48 (7)*	.516 (11.1)*	.677 (9.85)*		.62 (7.96)*	.459 (7.46)*	-.045 (-.25)	.584 (8.99)*	.492 (9.96)*
Manufacturing_Surf								.366 (1.56)		
Utility_Empl	.09 (2.45)*			.157 (2.74)*	.546 (10.89)*		.15 (3.36)*	.018 (.14)	.179 (4.04)*	
Utility_Surf						.121 (4.18)*		.017 (.31)		
Construction_Empl								.936 (2.34)*	.328 (3.1)*	
Construction_Surf								-.029 (-.09)		
Wholesale_Empl		.801 (4.62)*						-.726 (-1.43)		.583 (5.69)*
Wholesale_Surf		-.385 (-2.69)*						-.239 (-.63)		
Transport_Empl	.329 (6.83)*	.366 (5.24)*					.318 (5.7)*	.034 (.17)	.423 (7.07)*	.415 (8.94)*
Transport_Surf								-.023 (-.18)		
Accomodation_Empl								.224 (.68)		
Accomodation_Surf								.026 (.11)		
Information_Empl								.015 (.11)		
Information_Surf								.017 (.26)		
Financial_Empl								.54 (1.57)		-.276 (-3.32)*
Financial_Surf								-.675 (-3)*	.107 (2.23)*	
Science_Empl								-.697 (-2.24)*		
Science_Surf		-.328 (-4.16)*						-.001 (0)		-.081 (-2.18)*
Services_Empl								.574 (1.75)**	-.437 (-5.72)*	
Services_Surf								.038 (0.1)		
Others_Empl								-.065 (-0.42)		
Others_Surf	.035 (2.01)*							.127 (1.69)**		.056 (3.57)*

* significance at 5% significance level
** significance at 10% significance level
n is the number of observations

Table 6. Freight attraction parameters.

	0	1	2	3	4	5	6	7	8	9
n	551	530	503	541	283	352	553	193	554	549
R-square	.701	.671	.456	.688	.314	.426	.669	.297	.761	.789
Intercept	5.708 (21.98)*	2.018 (2.81)*	3.6 (3.52)*	3.466 (5.15)*	3.702 (4.37)*	-2.895 (-1.92)**	0.516 (0.58)	4.729 (1.89)**	4.523 (7.49)*	4.898 (10.22)*
Population_den	-.184 (-3.79)*			-.257 (-6.23)*	.568 (4.39)*			-.233 (-1.63)	-.135 (-3.46)*	-.238 (-5.76)*
Agriculture_Empl	.195 (6.28)*	.228 (8.01)*						-.009 (-.09)	.107 (4.18)*	
Agriculture_Surf					.092 (2.32)*			.073 (1.07)		
Mining_Empl					.291 (3.23)*			-.228 (-1.24)		
Mining_Surf								.029 (.48)		
Manufacturing_Empl	.391 (9.27)*	.353 (8.189)*		.201 (5.23)*		.298 (3.14)*		.018 (.15)	.344 (9.87)*	.259 (5.56)*
Manufacturing_Surf			.355 (4.99)*			.378 (2.99)*	.196 (4.01)*	.112 (.71)		.154 (2.98)*
Utility_Empl	.091 (3.212)*			.064 (2.41)*	.395 (7.1)*			-.028 (-.3)	.112 (4.71)*	
Utility_Surf							.059 (4.57)*	.012 (.3)		.06 (5.55)*
Construction_Empl			.704 (5)*				.29 (3.67)*	.193 (.9)		
Construction_Surf		.141 (2)*	-.3 (-2.26)*					-.002 (-.01)		
Wholesale_Empl	.374 (5.22)*	.301 (3.53)*		.326 (4.27)*			.403 (4.61)*	.028 (.09)	.303 (4.79)*	.323 (4.08)*
Wholesale_Surf								.043 (.17)		
Transport_Empl	.226 (5.79)*	.168 (4.3)*	.25 (3.44)*	.171 (4.72)*		.197 (2.57)*	.178 (4.12)*	-.029 (-.24)	.19 (5.94)*	.378 (10.57)*
Transport_Surf			-.162 (-2.98)*					.168 (1.89)**		
Accomodation_Empl						-.481 (-2.89)*		-.168 (-.8)		
Accomodation_Surf		-.112 (-2.57)*		.093 (2.31)*		.262 (2.08)*		.015 (.08)		.117 (2.7)*
Information_Empl								.009 (.09)		.127 (4.53)*
Information_Surf				.052 (3.74)*				-.01 (-.21)		
Financial_Empl							-.268 (-2.95)*	-.049 (-.22)		-.282 (-3.62)*
Financial_Surf							.106 (2.21)*	.03 (.24)		.103 (2.55)*
Science_Empl	-.158 (-3.04)*	-.228 (-4.26)*						-.07 (-.42)		
Science_Surf					-.148 (-1.74)**			-.002 (-.01)		
Services_Empl				-.155 (-2.11)*			-.292 (-3.46)*	-.071 (-.38)	-.184 (-2.77)*	-.138 (-2.36)*
Services_Surf		.233 (3)*	.223 (2.27)*	.315 (3.52)*		.342 (2.30)*	0.455 (4.23)*	.067 (.27)	.189 (2.37)*	
Others_Empl								.374 (3.29)*		
Others_Surf	.033 (2,61)*	.038 (2.98)*						-.072 (-1.55)	.03 (2.8)*	

* significance at 5% significance level
** significance at 10% significance level
n is the number of observations

The formulas obtained from the regression analysis on municipality level are then applied on the 4 934 traffic analysis zones, as original data is available for all 27 independent variables on that geographical scale. Thanks to this analysis generated and attracted volumes by commodity type for every single traffic analysis zone were obtained.

3.3 Introduction of the used loading unit

The used loading unit was added afterwards based on a probabilistic determination. Loading unit is, although usually neglected in transport modelling, an important factor in logistics as it influences transhipment costs, storage sizes, transport modes and costs, etc. and therefore the entire supply chain and transport system. The used probabilistic determination is based on the probability for a loading unit given the type of commodity and given the municipality which contains the concerned traffic analysis zone. By using real world data local differences in used loading units (like the presence of inland container port or distribution centre for palletized goods)are directly applied. Variation is not based on a modelling process (like discrete choice models or decision trees) as models inherently produce errors. If real data are available, they should be preferred upon modelled values. Using this analysis, generated and attracted volumes for each of the traffic analysis zones by commodity type and by loading unit were obtained. After this the volumes were calibrated.Based on volumes by traffic analysis zone and by commodity type and loading unit, those volumes were translated into origin-destination-combinations.

3.4 Origin-destination-combinations and shipments

Generated and attracted volumes by commodity type and loading unit were calculated for each of the 4 934 traffic analysis zones. However, in order to have transport flows between the traffic analysis zones, generated and attracted need to be linked to one another. Different methods exist for the trip distribution, namely the gravity models and opportunity models. Gravity models are most often used (Desmet et al., 2008; Piotte and Jourquin, 2013; Rowinski et al., 2008, amongst others), and are based on the declining relationship between two locations with increasing distance/time/cost, and positively associated with the amount of economic activity or population for each location. While opportunity models are related to the relative accessibility of opportunities that satisfy the goal of the trip. They are consequently not explicitly related to distance. All are based on accessibility and distance/time/cost, while the originally gathered dataset is providing origin-destination combinations by commodity type and loading unit in volume (tonnes) between all the Belgian municipalities and to/from NUTS2 European regions for respectively export, import and transit. Despite the fact that no distinction by transport mode is made, this dataset allows us to build a case-specific probability function as the traffic analysis zones can be linked to the municipality level which is more realistic than origin-destination relationships based on methods in literature. The breakdown from municipality level to traffic analysis zones is done by combining the probability of flows between municipalities with the probability that goods, by commodity type and loading unit, have a certain traffic analysis zone as destination within the municipality of arrival. This second probability is derived from the consumed volumes by commodity type and loading unit on traffic analysis zone level. Those consumed volumes are created by the freight generation model. By coupling those volumes on municipality level, the probability that goods for that specific commodity type and loading unit are arriving in a traffic analysis zone of that specific municipality is calculated. Together this results in the following aggregation-disaggregation equation:

$$P_i^{kl}(j) = P_m^{kl}(n) \cdot P_n^{kl}(j), \; i \in m, \; j \in n \tag{1}$$

Where: $P_i^{kl}(j)$ is the probability that a volume of commodity type k and loading unit l and with traffic analysis zone i as origin has traffic analysis zone j as destination.

$P_m^{kl}(n)$ is the probability that a volume of commodity type k and loading unit l and with municipality m – which contains traffic analysis zone i – as origin has municipality n as destination. It originates from the ADSEI 2012 data.

$P_n^{kl}(j)$ is the probability that a volume of commodity type k and loading unit l and with municipality n – which contains traffic analysis zone j – as destination has traffic analysis zone j as destination. It originates from the ADSEI 2012 data.

If the above probability function based on volume (in tonnes) is applied, one is confronted with unrealistic fragmentation of small volumes. In order to avoid this error, the probability function is applied on shipments. The probabilities per origin are translated into intervals. As for example; for origin 'A' the shipment has 30% chance to go destination 'B', 20% to 'C' and 50% to 'D', than the intervals for origin 'A' look like: 0%-30% for 'B', 30%-50% for 'C' and 50%-100% for 'D'. By picking for each shipment a random value between 0 and 1 (or in other words between 0% and 100%) a shipment is assigned to the destination of which the interval is containing the random value. In this example; if the random value is 0.67, than the shipment with origin 'A' will be assigned to destination 'D'.

The calculation of the number of shipments departing from a traffic analysis zone is based on optimal average shipment sizes by commodity type and loading unit. The shipment size is generally an optimisation function of both inventory and transport cost, taking into account commodity characteristics and economies of scale in the transport operation (de Jong and Ben-Akiva, 2007). However, in our analysis average shipment sizes must be calculated before the generation of origin-destination combinations to avoid the above fragmentation. It is by consequence impossible to include transport costs in the average shipment size optimisation function, as these depend on transport distance, transport mode and transport chain. The justification of this assumption will be tested once the freight transport model will be running well, like de Jong and Ben-Akiva (2007) did in their research from which the used standard Economic Order Quantity formula is extracted.

$$q_k = \sqrt{\frac{(o * Q * 2)}{(w+d*v)}} \tag{2}$$

Noted that q represents the average shipment size; o, the costant unit cost per order; Q, the annual demand (tonnes/year); w, the storage costs per unit per year; d, the discount rate (per year) and v, the value of the transported goods (per tonne). The value of the goods per NST/R type is retrieved from the Belgian national freight transport model (Desmet et al., 2008). The constant unit cost, storage cost per unit and discount rate is based on the work of Maes (2013) for the Flanders Region.

4. Results and discussion

The presented research has two major results which will contribute to a better modelling of freight transport and transport-related impact assessments. The first major result is a freight generation model explaining generated and attracted freight volumes according to both 10 different commodity types and 9 types of loading units. The introduction of used loading units is to our knowledge new and is an important contribution as the loading unit is influencing storage costs, transhipment costs and transport costs. As such it is having an impact on the entire transport system. The distribution of generated and attracted volumes is produced on a geographical scale of 4 934 traffic analysis zones. The zones are smaller in urban areas, allowing too use the outcome as input for in-depth analyses in the field of city distribution. Figure 5 illustrates the distribution for generated and attracted volume (in tonnes) for all commodity types and loading units on TAZ level. Large volumes are found in and around urban areas and air and seaports and zones with inland ports or large distribution centres. Despite the low R-squared values, this illustrates that the developed model is able to produce realistic variation and realistic volumes for the different zones.

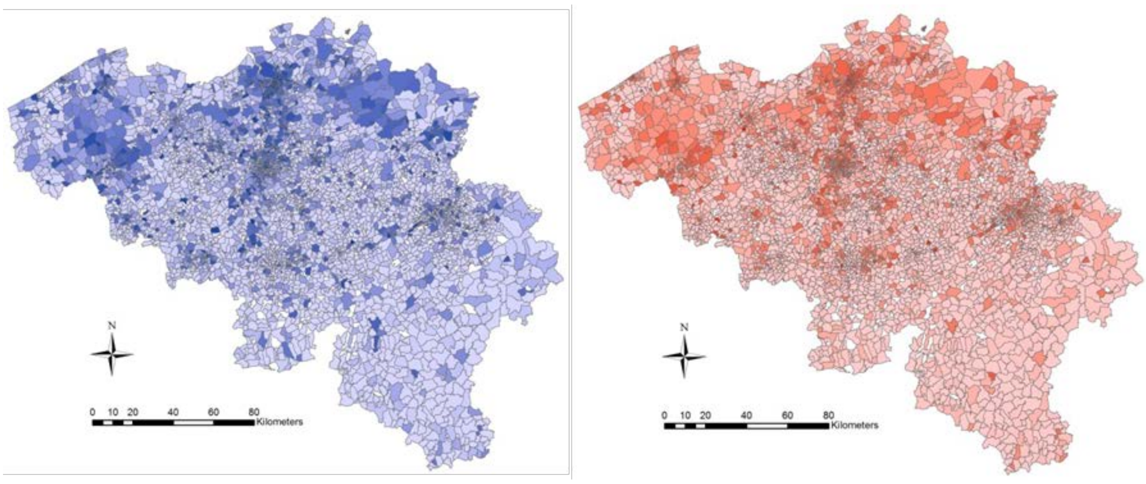

Figure 5. Generated (left) and attracted (right) volumes (in tonnes) for all commodity types and loading units on TAZ level.

Next to the detailed distribution in terms of commodity type, loading unit and geographical scale, the research also resulted in an origin-destination matrix for all modes together between the traffic analysis zones mutually and gateways as interaction with the rest of the world. In addition, the research improves the level of detail of freight flows within Belgium both on the geographical level with almost 5 000 traffic analysis zones as on level of the loading unit (ADSEI, 2012; Maes, 2013; Desmet et al., 2008). This allows more applied research with existing models and moreover it enables the development of new models considering different loading units. All mentioned possibilities for future research will facilitate impact studies on freight transport within Belgium, as they are currently based on aggregated freight flows between larger geographical entities.

5. Conclusion

This work presented a state-of-the-art freight generation model for the Belgian territory and origin-destination matrix between 4 934 traffic analysis zones. The model is developed to produce input for a multi-modal agent-based freight transport model. Like any other multi-modal freight model, mode choice is an important step in the modelling process. That choice is based on parameters of which some are influenced by the used loading unit (Mommens et al., 2016). The used loading unit is however, to our knowledge never properly considered in freight transport modelling. To enable the necessary incorporated of used loading unit into the modelling process disaggregated freight flows need to generate which include information on this used loading unit.

This paper presents how disaggregated freight flows can be generated – based on the knowledge and the limitations in freight generation modelling. This was done by applying regression analyses on the combined datasets containing population density and number of employees and gross floor surface per economic activity. Two main contributions in this part are the addressment of the assumptions of the regression techniques. This is rarely done in literature, despite its importance. Secondly, a generalized linear regression model with log (GLML) link and the gamma log generalized linear regression model (GLGLM) were compared. The GLML technique proved to be a suitable regression technique both in comparison with GLGLM and with the results of other methodologies used in literature.

Based on a data gathering process, it enabled to model generated and attracted volumes (in tonnes) by commodity type (#10) and by loading unit (#9).

The obtained volumes – together with in- and outgoing volumes with the rest of the world – were used to construct an origin-destination matrix which is unique in both geographical level of detail, as well as to regard of the used loading unit. It will be valuable input for new models and new analyses with existing models as it will enable research on the modal choice of different loading units, different loading and unloading infrastructures and loading unit dependent total logistics costs.

Acknowledgement

The research presented in this paper is part of the dec²4susts project ('Towards sustainable transport: the development of an External Cost Calculator for sustainable transport scenarios', IRP 42028), funded by the Vrije Universiteit Brussel in the frame of the university's Interdisciplinary Research Programme.

References

Algemene Directie Statistiek en Economische Informatie – ADSEI (2012).

Alho A.R., de Abreu e Silva J. (2015). Modeling commercial establishments' freight trip generation: 1 a two-step approach to predict weekly deliveries in total of 2 vehicles. *TRB*.

Bastida C., Holguin-Veras J. (2009). Freight Generation Models: Comparative Analysis of Regression Models and Multiple Classificarion analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 51-61.

Ben-Akiva, M., de Jong, G (2008). The Aggregate-Disaggregate-Aggregate (ADA) freight model system. In: Ben-Akiva, M., Meersman, H., Van de Voorde, E. (Eds.) *Recent Developments in Transport Modelling: Lessons for the Freight Sector*, 1st ed. Emerald, UK.

Chin S.M., Hwang H.L. (2007). National Freight Demand Modeling – Bridging the Gap between Freight Flow Statistics and U.S. Economic Patterns. *TRB*.

Davydenko I.Y.. (2015). Logistics Chains in Freight Transport Modelling. Dissertation, Technische Universiteit Delft.

de Jong G., Ben-Akiva M. (2007). A micro-simulation model of shipment size and tranport chain choice. *Transportation research part B*, 41, 950–965.

Desmet R., Hertveldt B., Mayeres I., Mistiaen P., Sissoko S. (2008). The PLANET model: Methodological report, planet 1.0. Working Paper 10-08. Federal Planning Agency, Brussels.

Douglas A. A., Lewis R.J. (1971). Trip Generation Techniques. *Traffic Engineering and Control* 12.10.

European Conference of Ministers of Transport (1993). Terminology on combined transport. Paris, OECD.

Guevara C. A., Thomas A. (2007). Multiple Classification Analysis in Trip Production Models. *Transport Policy*, 14, 514–522.

Guilbault M., Costa G., Franc P., Gouvernal E. Hémery C., Rizet C. (2008). ECHO : Envois – Chargeurs – Opérateurs de transport. Synthèse INRETS 56.

Holguin-Veras, J., Ban J., Jaller M., Destro L., Marquis R. (2010). Feasibility Study For Freight Data Collection (Final Report). Rensselaer Polytechnic Institute and New York Metropolitan Transportation Council.

Jaller M., Holguin-Veras J. (2013). Comprehensive freight demand data collection framework for large urban area. Rensselaer Polytechnic Institute and New York Metropolitan Transportation Council.

Khan M., Machemehl R. (2015). A Truck Trip Generation Model for Williamson County, Texas: Survey Analysis. *TRB*.

Lim R., Qian Z.S., Zhang H.M. (2014). Development of a freight demand model with an application to California. *International Journal of Transportation Science and Technology*, 3(1), 19-38

Luyten S., Van Hecke E. (2007). De Belgische Stadsgewesten 2001' [The Belgian Citydsitricts 2001]. Statistics Belgium, Working Paper, FOD Economie, Algemene Directie Statistiek en Economische Informatie, Leuven, KULeuven, Instituut voor Sociale en Economische Geografie.

Maes T. (2013). Modelling carrier decisions in an activity-based freight transportation framework. Dissertation, University of Hasselt.

Mobiel Brussel (2014). Modélisation des comportements tactiques en matière de mobilité' [Model of tactic behaviour in terms of mobility]. STRATEC, Brussels.

Mommens, K., van Lier, T., Macharis, C. (2016). Loading Unit in Freight Transport Modelling. *Procedia Computer Science*, 83, 921 - 927

Oliveira Neto F.M., Chin S.M., Hwang H. (2011). Aggregate Freight Generation Modeling: Assessing the Temporal Effect of Economic Activity on Freight Volumes using a Two-Period Cross-Sectional Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2285, 145-154.

Piotte J., Jourquin B. (2011). An agent based dynamic road freight transport demand Generation. *Association For European Transport and Contributors*, http://abstracts.aetransport.org/paper/index/id/3712/confid/17. Accessed 17 March 2015.

Rowinski J., Opie K., Spasovic L.N. (2008). Development of Methods to Disaggregate 2002 FAF2 Data Down to County Level of New Jersey. *TRB*.

Shin H., Kawamura K. (2005). Data Need for Truck Trip Generation Analysis: Qualitative Analysis of the Survey of Retail Stores. Transport Chicago, Illinois Institute of Technology.

Stopher P. R., McDonald K. (1983). Trip Generation by Cross Classification: An Alternative Methodology. *TRB*.

van Lier T. (2014). The development of an external cost calculator framework for evaluating the sustainability of transport solutions. Dissertation, Vrije Universiteit Brussel.

Verkeerscentrum Vlaanderen (2006). Ontwikkeling van een multimodaal goederenvervoermodel voor vlaanderen [Development of a multimodal freight transport model for Flanders]. Tech. rep., K+P transport consultants and TRITEL.

Wisetjindawat W., Sano K. (2003). A behavioral modeling in micro-simulation for urban freight transportation. *Journal of the Eastern Asia Society for Transportation Studies*, 5, 2193-2208.

Zhang M. (2013). A Freight Transport Model for Integrated Network, Service, and Policy Design. Dissertation, Technische Universiteit Delft.

Zhang Y., Bowden R.O., Allen A.J. (2003). Intermodal freight transportation planning using commodity flow data. Denver, National Center for Intermodal Transportation.