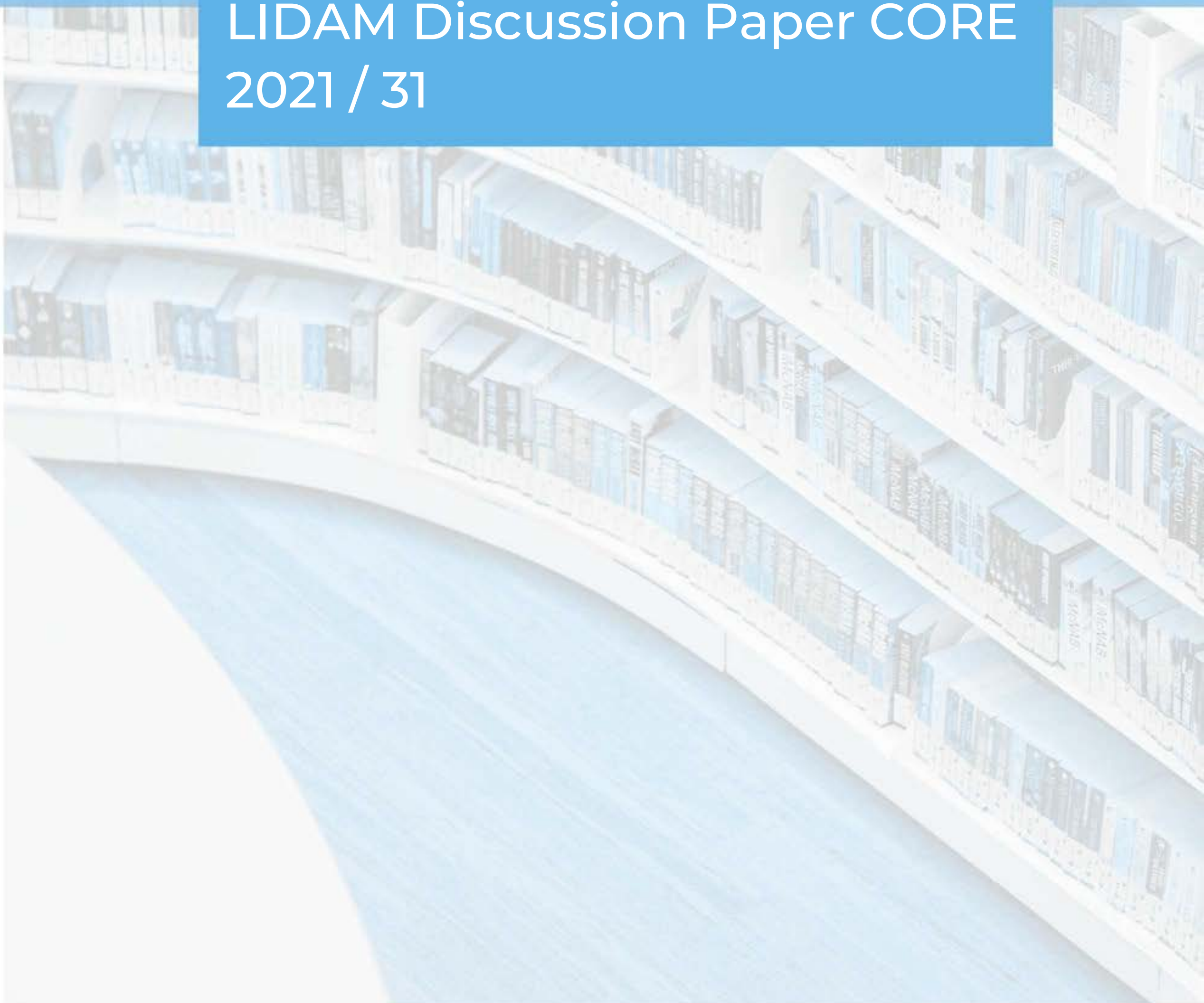
INTEGRATE AN ACCESSIBILITY MEASURE IN THE MODAL CHOICE OF STRATEGIC FREIGHT TRANSPORT MODELS

Bart Jourquin

LIDAM Discussion Paper CORE
2021 / 31



UCLouvain

LIDAM
CORE | Louvain Institute of Data Analysis and Modeling in economics and statistics

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-discussion-papers.html

Integrate an accessibility measure in the modal choice of strategic freight transport models

Bart JOURQUIN
Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium
bart.jourquin@uclouvain.be

Abstract

Modal choice models for strategic freight transportation studies covering large inter-regional or international areas generally rely on basic explanatory variables such as transportation costs and transit times.

Using origin-destination matrixes, it is also possible to compute an accessibility measure that can further be used as an additional explanatory variable.

This paper shows that the inclusion of an accessibility measure in the utility functions used for a logit model significantly improves its predictive power. Moreover, when the refined model is used to compute cost and transit time elasticities, the obtained (absolute) values are somewhat lower.

The use of an accessibility measure in the modal choice model has thus a double advantage for policymakers: it improves the predictive power of the freight transport model, giving more accurate traffics on the modal networks, and it avoids overestimations of own and cross-elasticities.

Keywords

Freight transport model; Modal choice; Accessibility; Elasticity

1 Introduction

Modal choice models are important tools supporting transport policy decisions. Used to compute elasticities, for instance, they can be very useful in cost-benefit or traffic impact analysis for planned infrastructure (de Jong, Gunn and Ben-Akiva, 2004; Liedtke and Carrillo Murillo, 2012; Beuthe *et al.*, 2014; Rothengatter, 2019). However, when applied to inter-regional or international areas, they are often setup using utility functions that rely on basic explanatory variables, such as transportation costs or transit times.

Cost and transit time are obviously not the only drivers in the modal choice decision process. Other variables, such as the levels of safety, flexibility, frequency, losses during transport and reliability play a role in decision-making (Beuthe and Bouffioux, 2008). In-depth surveys carried out with shippers, carriers and receivers conclude that network accessibility also plays an important role (Kim, Nicholson and Kusumastuti, 2014; Holguín-Veras *et al.*, 2021; Madaan *et al.*, 2021). The large ECHO survey conducted in France (Guilbault and Soppé, 2009) indicates that more than 30% of the shippers consider network accessibility as a criterion for their modal choice. Accessibility can be defined as a measure of the capacity of a location to be reached from (or to be reached by) different locations (Rodrigue, 2020). Therefore, the topology of the transport networks, and in particular its connectivity, is a key element in the determination of accessibility. Furthermore, accessibility also influences manufacturing output, and thus the demand for transport (Yi and Kim, 2018). Including an accessibility measure in modal choice models should therefore improve its robustness and explanatory power. Such a measure can be computed using a weight representing the attractiveness of a node, itself weighted using a decay function, that can for instance take the trip lengths as input. This length can rather easily be gathered or computed along with costs and transit times.

Some attempts to incorporate a spatial dimension in a destination choice model can be found in the literature(Cervero, 2002; Hammadou *et al.*, 2008). For instance, Hammadou *et al.*, (2008) propose to use spatial metrics to compute travel times by car in the region of Antwerp (Belgium). However, to the best of our knowledge, nothing has been published about the inclusion of a spatial dimension in modal choice models (for freight transport) using multivariate utility functions.

This paper examines to what extent accessibility, when added to transportation cost and transit time as a third explanatory variable, improves the performance of a modal choice model. It also presents a set of own and cross-elasticities computed using the improved model and compared to the values obtained without an accessibility measure. It appears that accessibility substantially improves the predictive power of the modal choice model and that the computed elasticities are lower than those obtained without this additional independent variable. Both results obviously are interesting for policymakers, who can rely on more precise estimated traffics and better estimations of the costs and benefits of a new infrastructure project, for instance.

However, a particular attention must be paid to two aspects of the model to setup.

Firstly, costs and transit times are often correlated, which is problematic. Indeed, when independent variables are correlated, changes in one variable are associated with shifts in other variables. In such cases, the coefficient estimates are unstable: their value and even sometimes their sign “swing” between the independent variables. Moreover, they also may have a weak statistical power (Greene, 2012; Adeboye, Fagoyinbo and Olatayo, 2014). The resulting estimators may thus have unexpected signs. For instance, an estimator with a positive value for the transportation cost by truck would signify that the probability to use trucks increases with their operational costs, which is obviously not what is expected.

Secondly, the computation of an appropriate accessibility index merits a particular attention. A lot has been written about accessibility measures and their ability to evaluate the accessibility impacts of national land-use and transport scenarios including the related social and economic impacts. It goes beyond the scope of this paper to present an extensive literature review on this topic. The interested reader can find such a review in Geurs and Ritsema van Eck (2001), for instance. It comes out that most publications deal with passenger transport and are often limited to urban areas. Examples of accessibility measures applied to freight transport apply the General Accessibility Measure (see section 1.2) representing the degree of interconnection between a particular reference location and all or a set of other locations in the area (Thomas *et al.*, 2003; Reggiani, Bucci and Russo, 2011). The decay (or impedance) function used to compute the accessibility measure can take many forms, but the negative exponential form has been used more often because it is supposed to more closely model travel behavior, especially for long distance interactions (Handy and Niemeier, 1997; Thomas *et al.*, 2003). This choice is, however, seldom challenged or even discussed (Reggiani, Bucci and Russo, 2011). Therefore, a set of several functional forms will be tested in this paper to choose the most appropriate.

The methodological aspects of the model are introduced in section 1, while section 2 describes the dataset used for the European case study. The results and the robustness of the model are discussed in section 3. Before a general conclusion, a set of own and cross elasticities are presented in section 4.

1 Methodological discussion

Logistic modeling (Cramer, 2002) is the most common approach for predicting modal choices in transportation economics. The logit mode choice relationship states that the probability of choosing a particular mode for a given trip is based on the relative values of several factors such as cost, level of service or travel time. The general form of a (multinomial) logit model can be written as:

$$P_m = \frac{e^{U_m}}{\sum_n e^{U_n}} \tag{1}$$

Where:
P_m : Probability of choosing mode m
U_m : Utility of mode m
U_n : Utility of mode n

A first model is set up, which utility function contains only costs C and transit times T. One could argue that the transport time is incorporated in the transport cost through the value of the goods during transport, but we decided not to do so for two reasons. Firstly, it is known that the value of the transported goods (the opportunity costs) is much lower than the value of time for freight transport (Blauwens and Van de Voorde, 1988; De Jong *et al.*, 2004). Secondly, policymakers are also interested in the impact of time savings. Keeping both transportation costs and transit times as separate independent variables makes thus sense. As C and T are alternative (mode) specific rather than being individual (origin-destination pair) specific, the analysis applies the McFadden's conditional logit model (McFadden, 1973). The conditional logit differs from the multinomial logit as α and β are not mode specific. The utility function for the first model can thus be written as:

$$U_m = \alpha C_m + \beta T_m + \delta_m \tag{2}$$

Where:
C_m = Transportation cost for mode *m*
T_m = Transit time for mode *m*
α, β, δ_m : Parameters to estimate

Obviously, it is expected that the α and β coefficients have a negative sign, as a cost increase or a longer transit time makes a transportation mode less attractive. Finally, δ_m is the calculated intercepts for each mode. As one mode must be considered as the reference mode (road transport in this paper), δ_{road} is set to 0.

Finally, note that, when transport demand is embedded in origin-destination (OD) matrixes containing aggregated data (of total annual transport quantities for instance), a weighted logit methodology should be applied, whereby the transported tonnages weight the mode choice observations in the Log-Likelihood functions (Rich, Holmblad and Hansen, 2009; Jourquin and Beuthe, 2019).

The improved model that is presented in this paper includes an accessibility measure in the utility function, that becomes:

$$U_m = \alpha C_m + \beta T_m + \sigma_m A_m^o + \delta_m \tag{3}$$

Where:
A_m^o : Accessibility measure of location *o* for mode *m* (see section 1.2).
σ_m : Parameter to estimate

Unlike *C* and *T*, which are alternative specific, A^o is specific to each location. Therefore, a separate σ_m parameter must be estimated for each mode.

Even if the utility functions for both models are rather simple, a special attention must be paid to the correlation between C_m and T_m (multicollinearity) and to the definition of the accessibility measure A_m^o.

1.1 Handling multicollinearity

Several methods can be found in the literature to overcome multicollinearity (Greene, 2012; Dormann *et al.*, 2013), among which the Box-Cox transforms (Box and Cox, 1964) of independent variables. Beside the fact that this technique can be useful to improve the Log-Likelihood of a model, it is indeed sometimes considered as an elegant and efficient way to help overcome multicollinearity and to re-establish the expected signs of the estimators (Fridstrøm and Madslien, 1994; Gaudry, 2016). The Box-Cox transformation (equation 4) belongs to the family of power transforms, used to create a monotonic transformation of data using power functions.

$$BC(x,\lambda)=\begin{cases}\frac{x^{\lambda}-1}{\lambda}, if\ \lambda\neq 0\\ \log(x), if\ \lambda=0\end{cases} \tag{4}$$

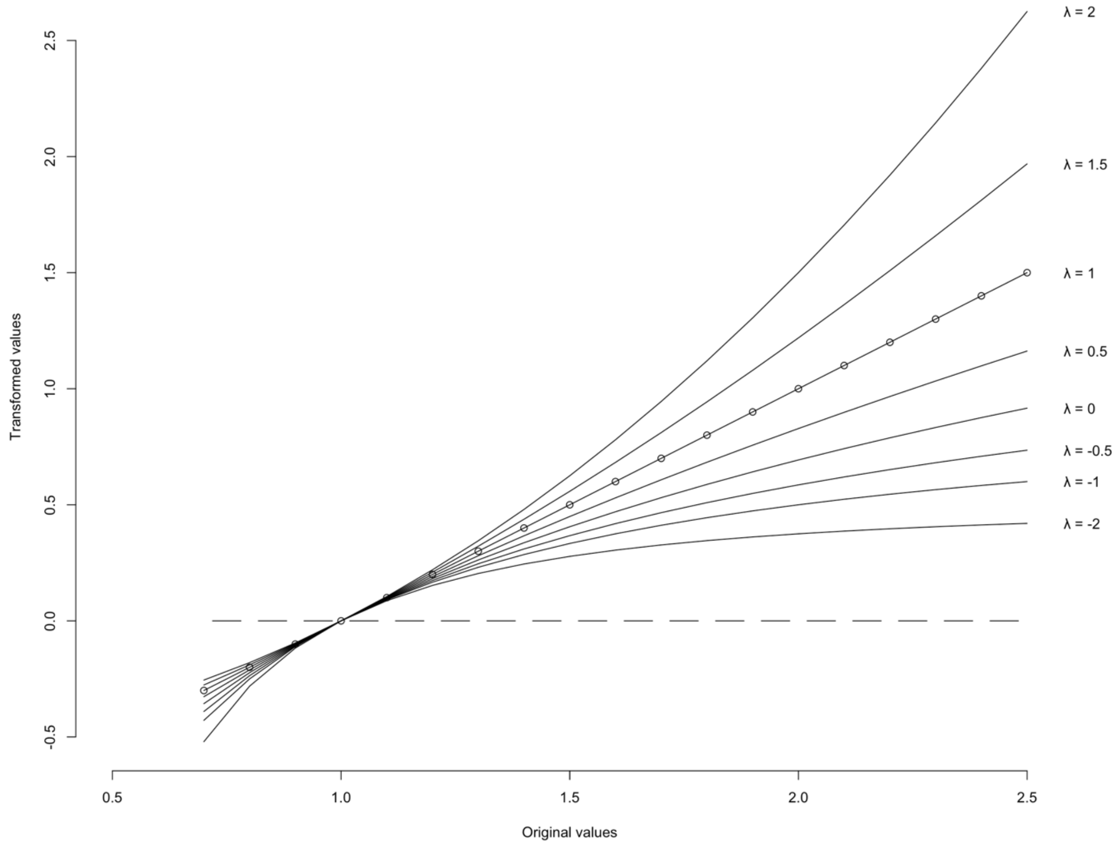


Figure 1: Box-Cox transformations for several values of λ

By "bending" the functional form (Figure 1), the chosen value for λ offers a lot of flexibility, providing more degrees of freedom to the modeler. To include the Box-Cox transforms of *C* and *T*, the utility functions of both models can thus be rewritten to become[1]:

$$U_m=\alpha\left(\frac{C_m^{\lambda_c}-1}{\lambda_c}\right)+\beta\left(\frac{T_m^{\lambda_t}-1}{\lambda_t}\right)+\delta_m \tag{5}$$

$$U_m=\alpha\left(\frac{C_m^{\lambda_c}-1}{\lambda_c}\right)+\beta\left(\frac{T_m^{\lambda_t}-1}{\lambda_t}\right)+\sigma_m A_m^o+\delta_m \tag{6}$$

Where λ_c and λ_t are the Box-Cox parameters to estimate.

1.2 Computing the Accessibility Measure

1 Obviously, $\log(C_m)$ or $\log(T_m)$ must be used if λ_c or $\lambda_t = 0$.

Even if a very simple accessibility measure (the number of centroids that can be reached by each mode from a given origin) already seems to give interesting results (Quittelier, 2021), the General Accessibility Measure (Morris, Dumble and Wigan, 1979; Pooler, 1995; Gutiérrez, Monzón and Piñero, 1998), a well-established measure in the literature, will be used here. It can be written as:

$$A^{0}=\sum_{d} w_{d} f\left(l_{od}\right) \tag{7}$$

Where:
A^o : Accessibility at origin o
w_d : Weight representing the attractiveness of destination d
L_{od} : Measure of separation (length for instance) between o and d
$f(\ldots)$: Decay function

The accessibility measure that is used here is thus defined as a measure of the capacity of a location to be reached from. Hence, the model will consider that the modal choice decision is taken at the origin node, which is a limit of the exercise presented in this paper.

The values computed by the formula presented in equation 7 directly depend on the weights w_d (transported quantities). However, as explained earlier, a weighted logit is implemented here. Therefore, a relative accessibility index (Thomas *et al.*, 2003) will be used, which can be written as:

$$\Lambda^{o}=\frac{\sum_{d} w_{d} f\left(L_{od}\right)}{n \sum_{d} w_{d}} \tag{8}$$

Where n is the number of nodes included in the study area.

Knowing the values of w_d (the transported tons[2] from o to d), L_{od} (the trip length between o and d), the number of centroids n, and the decay function to use, Λ^o can be computed. Thus, the utility function (6) becomes:

$$U_{m}=\alpha\left(\frac{C_{m}^{\lambda_{c}}-1}{\lambda_{c}}\right)+\beta\left(\frac{T_{m}^{\lambda_{t}}-1}{\lambda_{t}}\right)+\sigma_{m} \Lambda_{m}^{o}+\delta_{m} \tag{9}$$

Finally, if multicollinearity is undoubtedly a problem that must be tackled for C and T, one could argue that the same problem occurs between Λ^o, C and T, because Λ^o is computed using L, which is also strongly correlated with C and T. However, the value of L is diluted in the decay function and w, the attractiveness of destinations plays a preponderant role. A model using C, T and L, applied to the same dataset that the one that is outlined in section 2, is presented in (Jourquin, 2021). It appears that the inclusion of L doesn't improve the predictive power of the model while the inclusion of Λ^o has a clear impact (section 3.4).

2 One could consider that the volume of the goods or their perishability could be used as w_d, but only aggregated data expressed in tons are available for the type of transport models discussed in this paper.

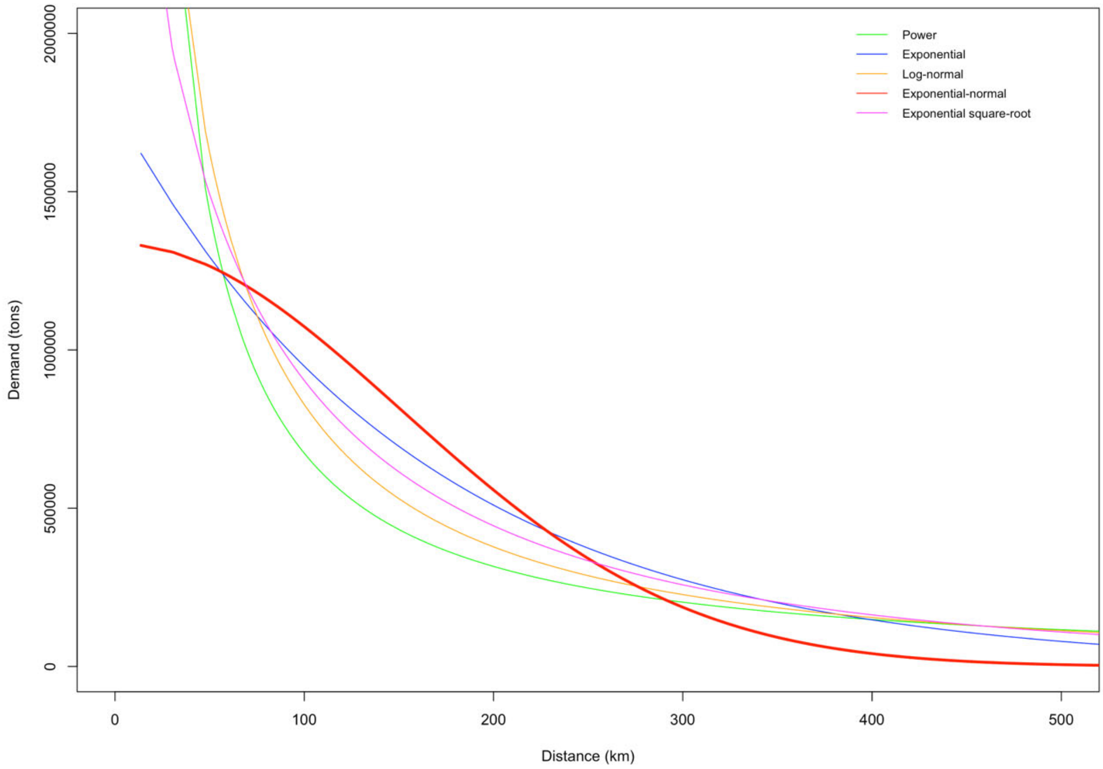


Figure 2: Tested decay functions

Most authors use the (negative) exponential form represented by equation 11 for the decay function, as it is considered as the most closely tied to travel behavior (Handy and Niemeier, 1997; Thomas *et al.*, 2003) Other functional forms can be considered, and some of them could be more suitable for freight transport. Therefore, the five specifications presented in Reggiani, Bucci, and Russo (2011) and illustrated in Figure 2, will be tested (see section 3.2):

1 - Power: $$f\left(L_{od}\right)=L_{od}^{-\gamma_1} \tag{10}$$

2 - Exponential: $$f\left(L_{od}\right)=e^{-\gamma_2 L_{od}} \tag{11}$$

3 - Log-normal: $$f\left(L_{od}\right)=e^{-\gamma_3\left(\log\left(L_{od}\right)\right)^2} \tag{12}$$

4 - Exponential normal: $$f\left(L_{od}\right)=e^{-\gamma_4 L_{od}^2} \tag{13}$$

5 - Exponential square root: $$f\left(L_{od}\right)=e^{-\gamma_5\sqrt{L_{od}}} \tag{14}$$

In these equations, γ_1, γ_2, γ_3, γ_4 and γ_5 are the parameters to estimate and L_{od} is a measure of separation, in this case the trip length between *o* and *d*.

2 Dataset description

The dataset used in the context of this paper is gathered from the ETISPlus European FP7 research program (Szimba *et al.*, 2012). Combining data, analytical modeling with maps and an online interface for accessing the data, it provides an information system useful for assessing European transport policies. Public access to several deliverables and many data is provided, among which origin-destination matrixes and digitized networks.

In the context of this paper, the OD matrixes for the year 2010 at the NUTS-2 regional level are used. They are available for roads, railways, and inland waterways (IWW) transport, with annual volumes for 10 categories of commodities (NST/R chapters 0 to 9, see Table 1). In this paper, a subset of the OD matrixes is used, focused on continental European transport. It includes the countries belonging to the European Union (EU) or to the European Economic Association (EEA), apart from islands and very peripheral zones within some countries.

NST/R	Description
0	Agricultural products and live animals
1	Foodstuffs and animal fodder
2	Solid mineral fuels
3	Petroleum products
4	Ores and metal waste
5	Metal products
6	Crude and manufactured minerals, building materials
7	Fertilizers
8	Chemicals
9	Machinery, transport equipment, manufactured articles and miscellaneous articles

Table 1: NST/R categories of commodities

The datasets corresponding to the different categories of commodities (Table 2) are quite different in size (number of OD cells), volumes (Mios tons/year) and modal split (market shares). This diversity is useful to test the robustness of the models presented in section 1.

The modal matrixes are combined, and the modal choice models are applied to these merged matrixes. Separate models are solved for each group of commodities.

	OD cells			Volumes (10^6 t)			Market shares (%)		
NST-R	**Road**	**IWW**	**Rail**	**Road**	**IWW**	**Rail**	**Road**	**IWW**	**Rail**
0	40619	1137	3771	361,6	24,2	34,2	86,1%	5,8%	8,1%
1	40042	1151	1532	478,1	25,7	9,3	93,2%	5,0%	1,8%
2	19761	559	4115	47,5	37,2	158,4	19,5%	15,3%	65,2%
3	27137	664	4199	190,1	67,0	26,5	67,0%	23,6%	9,3%
4	27596	977	5816	569,0	31,7	72,9	84,5%	4,7%	10,8%
5	36317	677	5377	189,4	11,7	43,5	77,4%	4,8%	17,8%
6	36771	1356	5714	1059,5	111,3	179,3	78,5%	8,2%	13,3%
7	27136	738	5392	43,0	5,8	76,8	34,2%	4,6%	61,2%
8	39834	1072	5493	302,0	37,0	38,8	79,9%	9,8%	10,3%
9	44408	1075	5846	1030,5	66,7	93,5	86,5%	5,6%	7,8%

Table 2: Content of the demand matrixes

In addition to the OD matrixes, digitized networks are available in the ESRI shapefile format. Even if the downloaded files can be visualized directly in a GIS software, the ETISPlus networks cannot be used for assignments. They are indeed not conceived for this, and some important manipulations are needed to make them "assignment compatible"(Jourquin, 2017). The resulting networks contain more than 58,000 road links, 1,600 inland waterways (IWW) links and 10,000 railroad links.

3 Application and model performance

The methodology outlined in section 1 is applied to the dataset described in section 2, following the workflow illustrated in Figure 3.

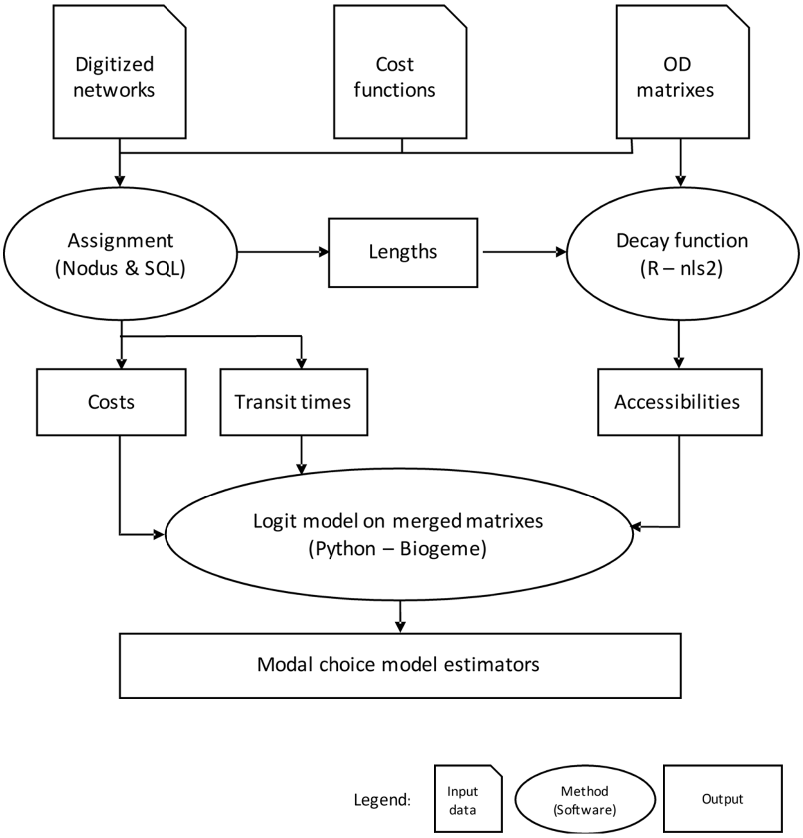


Figure 3: Model workflow

3.1 Compute costs and transit times

The OD matrixes and the networks are imported in Nodus (http://nodus.uclouvain.be) to compute the values of the explanatory variables needed for the modal choice models. The software allows retrieving, for each OD pair, each mode and each group of commodities, the total cost of transport *C*, the total transit time *T* and the trip length *L*. These are thus constructed variables, retrieved from the results of an assignment of each observed modal OD matrix on its corresponding digitized network.

The loading and unloading costs *ld_cost* and *ul_cost* are fixed costs, though they vary with the mode and the transported goods. Note that the transshipment costs associated with intermodal transport chains are not explicitly modeled in this exercise. The loading factors of the vehicles are taken from the ECCONET research project (Beuthe *et al.*, 2014). They are exogenous but specific for each group of commodities and type of vehicle (trucks, trains or barges) included in the model. The traveling unit cost, or moving cost, *mv_cost,* depends on the length and on the average commercial speed for the considered mode, as well as on the transported commodity *g*. It considers all the carrier's costs during transport: labor, capital invested in vehicles, fuel, maintenance, insurance, handling and storage costs, services directly linked to a transport, plus all residual indirect costs like those of administrative services. The duration of the transportation also impacts the inventory cost of transported goods proportional to their value. For a given link *l* belonging to a network of mode *m*, it is computed as:

$$mv_{cost\,l,m}^{\;g} = \frac{Average\,speed_m}{Speed_{l,m}} * length_l * unit\,mv_{cost\,m}^{\;g} \qquad (15)$$

As the unit *mv_cost* also contains time-related costs, the *Average speed / Speed* ratio allows for considering higher/lower than average costs on slower/faster segments of the network. Indeed, *Average Speed*$_m$ represents the average speed for mode *m* on the network and *Speed*$_{l,m}$ is the average speed on link *l*.

The total cost C_m^g of a route between an origin and a destination for a vehicle of mode *m* transporting commodities of type *g* is thus equal to:

$$C_m^g = ld_{cost\,m}^{\;g} + ul_{cost\,m}^{\;g} + \sum_l^R mv_{cost\,l,m}^{\;g} \qquad (16)$$

where *R* is the set of successive links representing the route.

Similarly, the total transit time has fixed elements (the loading and unloading durations (*ld_duration* and *ul_duration*) and a variable part (the travel duration that depends on length and speed on the successive links along the route). Thus:

$$T_m^g = ld_{duration\,m}^{\;g} + ul_{duration\,m}^{\;g} + \sum_l^R mv_{duration\,l,m} \qquad (17)$$

with $mv_{duration\,l,m} = length_l / speed_{l,m}$ (18)

3.2 Choice of a decay function

The γ_x parameters of equations 10 to 14 are estimated using the R "nls2" package (Grothendieck, 2013), which determines the nonlinear least square estimates of the parameters of a nonlinear model. The five decay functions are estimated for the three transportation modes and the ten groups of commodities listed in section 2. As shown in Figure 2, and unlike the other formulations, the exponential normal decay function has an S-shape. This is probably why it was considered as being the less efficient decay function by Reggiani, Bucci, and Russo (2011). Their conclusion could be explained by the fact that they used the R^2 of the models to compare the performances of the different formulations. However, R^2 is known not to be an appropriate performance measure for non-linear functions (Kvalseth, 1983; Spiess and Neumeyer, 2010). Indeed, if R^2 is used for nonlinear models, the following problems occur:

- R^2 is consistently high for both excellent and appalling models.
- R^2 will not always rise for better models.
- If R^2 is used to pick the best model, it leads to the proper model only 28-43% of the time (Spiess and Neumeyer, 2010).

Therefore, another measure, such as the residual standard error should be used instead. Table 3 compares the relative residual standard error ($\frac{RSE_{model}}{RSE_{best\,model}}$) of each specification (1-5, as given by equations 10 to 14). It comes out that the exponential normal decay function (4) always performs best for road transport. For the two other modes, no decay function appears to clearly outperform the others. Altogether, the exponential normal formulation will further be used in this paper to compute the accessibility measure.

	Road					IWW					Rail				
NST/R	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
0	4.0%	0.6%	2.5%	**0.0%**	1.4%	**0.0%**	0.2%	0.3%	0.7%	0.2%	0.1%	0.7%	**0.0%**	0.7%	**0.0%**
1	5.0%	1.0%	3.0%	**0.0%**	1.9%	0.0%	0.1%	**0.0%**	0.3%	**0.0%**	0.2%	0.1%	0.2%	**0.0%**	0.1%
2	2.0%	0.5%	1.6%	**0.0%**	1.3%	**0.0%**	0.4%	**0.0%**	0.6%	0.2%	0.1%	**0.0%**	0.1%	0.1%	**0.0%**
3	3.3%	0.5%	2.4%	**0.0%**	1.2%	**0.0%**	0.1%	0.1%	0.2%	0.1%	0.1%	**0.0%**	0.2%	0.1%	0.2%
4	3.5%	0.9%	2.8%	**0.0%**	1.7%	**0.0%**	**0.0%**	0.1%	**0.0%**	0.1%	0.1%	**0.0%**	0.1%	**0.0%**	0.1%
5	1.8%	0.2%	1.2%	**0.0%**	0.6%	0.8%	**0.0%**	0.3%	0.8%	0.5%	**0.0%**	0.1%	0.1%	1.1%	0.1%
6	4.1%	0.8%	2.8%	**0.0%**	1.7%	0.8%	**0.0%**	0.3%	0.5%	**0.0%**	0.1%	0.3%	0.1%	0.4%	**0.0%**
7	2.1%	0.6%	1.8%	**0.0%**	1.3%	**0.0%**	**0.0%**	0.3%	**0.0%**	0.2%	0.5%	0.7%	0.3%	**0.0%**	0.1%
8	5.3%	1.0%	3.3%	**0.0%**	1.9%	**0.0%**	1.0%	**0.0%**	1.1%	**0.0%**	**0.0%**	**0.0%**	0.3%	0.2%	0.3%
9	4.2%	0.7%	2.6%	**0.0%**	1.4%	**0.0%**	0.4%	**0.0%**	0.2%	**0.0%**	0.2%	0.5%	0.1%	0.5%	**0.0%**

Table 3: Relative residual standard error for the 5 decay functions (per mode and NST/R)

3.3 Logit model estimation

Table 4 presents the modal choice models estimators and their corresponding t-test for model 1 (without connectivity, Equation 5) while Table 5 contains those of the second model (Equation 9). The parameters are estimated using Biogeme (Bierlaire, 2020), an open-source Python package designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models. The only constraint imposed to both models is that, as already discussed, α and β must be negative.

Estimators **NST/R**		α	β	δ_2	δ_3	λ_c	λ_t
0	*Value*	-0.414	-0.939	-2.98	-1.97	0.393	-1.08
	t-test	-13.9	-5.7	-36.2	-64.8	22.3	-9.62
1	*Value*	-1.49	-1.26	-1.78	-3.17	0.0237	-0.17
	t-test	-14.6	-12.5	-7.76	-17.6	0.92	-5.9
2	*Value*	-2.9	0	-3.87	-1.41	0.0654	0.807
	t-test	-28.9	0	-55.8	-40.6	6.15	0
3	*Value*	-0.415	-0.833	1.05	0.438	0.585	0.155
	t-test	-14.2	-9.43	4.6	2.6	31.5	6.62
4	*Value*	-4.93	-2.62	1.49	2.63	-0.216	0.00496
	t-test	-9.49	-13.4	4.55	9.09	-4.05	0.229
5	*Value*	-0.522	-0.00451	-2.85	-0.324	0.351	1.15
	t-test	-12	-2.91	-41.3	-7.42	13.6	16
6	*Value*	-0.236	-0.781	2.35	2.09	0.529	0.277
	t-test	-10.2	-16.7	12.9	14.8	21.1	20.8
7	*Value*	-3.43	0	-4.95	-1.38	-0.0994	0.127
	t-test	-28.4	0	-67.7	-39	-10.5	0
8	*Value*	-1.23	-0.000624	-3.4	-1.44	0.225	1.69
	t-test	-22.3	-3.41	-66	-53.3	16.7	26.6
9	*Value*	-0.806	-0.0124	-1.54	-0.353	0.143	1.24
	t-test	-8.72	-5.06	-16.8	-5.28	3.76	26.6

Table 4: Estimated parameters for the model without connectivity

It appears that almost all estimators are (highly) significant. The only exceptions are the estimators for transit time β and their corresponding Box-Cox modifier λ_t for NST/R 2 and 7 in the first model and for NST/R 2 only in the second model. This is because the optimal λ of the Box-Cox transform producing a β having the expected negative sign cannot always be identified using a pure max-likelihood approach. Alternative strategies exist that try to overcome this problem (Jourquin, 2021), but their implementation is not straightforward. As a complete discussion about this topic deviates from the objective of this paper which is focused on the added value of the accessibility measure in the utility function, the estimators presented in Table 4 and Table 5 are further used.

Estimators NST/R		α	β	σ_1	σ_2	σ_3	δ_2	δ_3	λ_c	λ_t
0	*Value*	-0.346	-1.76	7.07	1.59	3.46	-2.18	-1.79	0.402	-0.992
	t-test	-11.1	-9.73	36.3	43.2	89.7	-20.8	-39.3	18.6	-12
1	*Value*	-1.25	-3.04	3.4	0.671	1.78	-3.17	-4.31	0.0616	-0.932
	t-test	-12	-8.63	17.4	68.1	53.8	-15.5	-25.2	1.89	-8.04
2	*Value*	-1.26	0	3.59	1.54	1.42	-3.2	-0.856	0.255	0.0651
	t-test	-14.4	0	44.2	53.4	59.3	-34.9	-18.1	13.2	0
3	*Value*	-1.38	-0.00228	2.21	0.745	1.86	-4.19	-3.27	0.398	1.32
	t-test	-20.2	-1.6	30.2	76.3	60.2	-38	-47.2	28.4	9.92
4	*Value*	-74.4	-4.63	4.24	0.264	2.72	4.12	4.35	-2.03	-0.214
	t-test	-6.58	-36.5	31	50.4	67.8	21.5	30.1	-20.5	-12.9
5	*Value*	-1.11	-0.000431	2.34	0.846	2.68	-3.73	-0.706	0.138	1.41
	t-test	-12.7	-1.07	48.7	55.7	90.9	-57.5	-20.3	5.52	7.48
6	*Value*	-0.559	-0.758	3.25	2.47	2.51	-0.358	0.209	0.388	0.129
	t-test	-13.3	-11.9	32	64.3	78.3	-1.72	1.4	18.4	6.5
7	*Value*	-3.66	-1.01	2.76	1.46	3.1	-4.82	-1.36	-0.128	-0.81
	t-test	-21	-6.06	51	62.4	72.4	-30.1	-11.4	-6.68	-6.63
8	*Value*	-1.43	-0.000248	1.79	1.44	2.26	-4.05	-1.81	0.184	1.82
	t-test	-22.4	-2.47	25.1	75.6	65.5	-72.7	-65.1	13.6	21.4
9	*Value*	-1.44	-0.00132	3.77	0.608	2.87	-2.89	-1.33	-0.0512	1.6
	t-test	-8.76	-2.57	28.8	86.4	78.2	-35.1	-22.1	-1.33	18.7

Table 5: Estimated parameters for the model with connectivity

3.4 Comparative performances

The Log-likelihood and the Akaike information criterion of model 2 are better for all groups of commodities. As these performance measures don't give a comprehensive idea of the added value of the accessibility measure in the model, an analysis of the correlation coefficients *r* between the computed tonnages by the model and those gathered ("observed") from the ETIS modal matrixes for each OD pair, each group of commodities and each mode is presented in Table 6.

The figures of the last row ("Aggregate") represent the weighted average correlation (depending on the volumes transported for each NST/R group). It appears that, altogether, the *r* coefficient for road transport increases from 0.94 to 0.98, but it is mainly the increases for IWW (from 0.67 to 0.88) and rail transport (from 0.69 to 0.83) that are interesting. The improvements are even more substantial for some modes and categories of commodities, such as IWW transport of NST/R 7 (fertilizers) from 0.17 to 0.85 or rail transport of NST/R 5 (metal products) from 0.55 to 0.85 for instance. It is thus clear that the integration of an accessibility measure in the utility function substantially improves the performance of the modal choice model.

	Model 1 (cost & transit time)			**Model 2 (cost, transit time & accessibility)**		
NST/R	**Road**	**IWW**	**Rail**	**Road**	**IWW**	**Rail**
0	0.98	0.76	0.52	0.99	0.84	0.79
1	0.99	0.52	0.33	1.00	0.87	0.47
2	0.62	0.59	0.96	0.69	0.99	0.99
3	0.84	0.85	0.66	0.95	0.97	0.69
4	0.91	0.97	0.66	0.99	0.99	0.89
5	0.96	0.27	0.55	0.98	0.67	0.85
6	0.97	0.54	0.58	0.98	0.76	0.77
7	0.50	0.17	0.96	0.79	0.85	0.98
8	0.95	0.62	0.64	0.97	0.83	0.67
9	0.93	0.75	0.42	0.98	0.93	0.64
Aggregate	0.94	0.67	0.69	0.98	0.88	0.83

Table 6: Correlations between observed and computed quantities at the OD level

The performance of both models can also be compared to what is obtained using *C*, *T* and *L* (instead of Λ^o) in the utility function (Jourquin, 2021). Table 7 shows that, unlike Λ^o, *L* doesn't improve the model. In some cases, the correlations are even somewhat lower than those computed with *C* and *T* only.

	Model 1 (cost & transit time)			**Cost, transit time & length**		
NST/R	**Road**	**IWW**	**Rail**	**Road**	**IWW**	**Rail**
0	0.98	0.76	0.52	0.98	0.76	0.52
1	0.99	0.52	0.33	0.99	0.52	0.33
2	0.62	0.59	0.96	0.62	0.53	0.96
3	0.84	0.85	0.66	0.82	0.82	0.67
4	0.91	0.97	0.66	0.91	0.96	0.66
5	0.96	0.27	0.55	0.96	0.24	0.57
6	0.97	0.54	0.58	0.97	0.52	0.58
7	0.50	0.17	0.96	0.51	0.15	0.96
8	0.95	0.62	0.64	0.95	0.61	0.64
9	0.93	0.75	0.42	0.93	0.75	0.42

Table 7: Added value of L in the model

The figures presented in Table 6 neglect the role that the network topology can play. As no observed count data is available along the segments of the networks, the traffic computed by separate assignments of each mode (the OD matrix of a mode assigned to its own network) is used as a proxy of the actual transport operations on each link. These reference flows are then compared to those obtained by a multimodal assignment procedure performed using Nodus. The modal split is computed by Nodus using a specific modal choice plugin that uses the estimators presented in Table 4 and Table 5 along with, for model 2, the relevant computed accessibility measure for each centroid.

Table 8 gives the resulting correlation coefficients between the two sets of flows[3] assigned to the same links. There figures show that model 2 always performs noticeably better.

NST/R	Model 1 (cost & transit time) Road	Model 1 IWW	Model 1 Rail	Model 2 (cost. transit time & accessibility) Road	Model 2 IWW	Model 2 Rail
0	0,96	0,73	0,44	0,97	0,75	0,70
1	0,99	0,75	0,18	0,99	0,95	0,39
2	0,72	0,92	0,95	0,73	1,00	0,99
3	0,81	0,96	0,76	0,92	0,99	0,85
4	0,90	0,94	0,75	0,98	0,98	0,87
5	0,93	0,69	0,79	0,96	0,84	0,86
6	0,96	0,80	0,69	0,98	0,84	0,83
7	0,52	0,43	0,96	0,79	0,93	0,99
8	0,95	0,88	0,78	0,96	0,91	0,81
9	0,96	0,89	0,59	0,98	0,91	0,74
Aggregate	0,97	0,89	0,88	0,99	0,95	0,93

Table 8: Correlations between "observed" and computed traffic on the networks

4 Computed elasticities

It is now clear that the introduction of an accessibility measure significantly improves the performance of the modal-choice model. Policymakers can thus rely on more precise traffic estimations for the evaluation of the impact of a new transport infrastructure for instance.

Beside this, it is also worth considering the impact of the improved model on the computed elasticities. In economics, elasticity is a measure of a variable's sensitivity to a change in another variable. After the reference volume Q_1 being computed with the reference value of an independent variable V_1, the value of the latest is modified (increased by 5% for instance). This new value V_2 is reused in the modal choice model to compute the new volume Q_2, keeping all the estimators unchanged. This provides the figures that are necessary to compute arc-elasticities (Equation 19):

$$\varepsilon = \frac{(Q¿¿1 - Q_2)/Q_1}{(V¿¿1 - V_2)/V_1¿}¿ \tag{19}$$

Table 9 gives the computed own and cross elasticities computed using both models. The extreme values (range) are printed in italic while the average weighted values are in bold underneath. It goes beyond the scope of this paper to present an in-depth discussion about the obtained elasticities, but there is no doubt that they are comparable to what is published elsewhere (De Jong, 2003; Rich, Holmblad and Hansen, 2009; Beuthe, Jourquin and Urbain, 2014). The values computed for model 1 can also be directly compared to those published by Jourquin and Beuthe (2019) as they were obtained using the same dataset and a similar (but not identical) Box-Cox transform of the *C* and *T* variables.

It appears that the computed elasticities are somewhat lower (in absolute value) for the model that includes the accessibility measure. This can be explained by the fact that the

3 In order not to bias *r*, the links connected to only two other links of the same mode, and from which it is not possible to change direction are removed from the calculus. Indeed, the flow on these links is always equal to the flow on their preceding and following links.

addition of a third explanatory variable improves the performance of the model by extracting its explanatory power not only from the intercepts δ_2 and δ_3, but also, partially, from C and T. Even if the impact on the elasticities is rather limited, it is not negligible. This indicates that the elasticities that can be found in the literature for strategic freight transport models which modal choice is based on costs and transit times only tend to be overestimated. Obviously, this can have an important impact on the conclusions of cost-benefit analysis or socio-economic impact of a projected new infrastructure, especially if its estimated profitability is low.

	Model 1 (cost & transit time)						**Model 2 (cost. transit time & accessibility)**					
	Cost			**Transit time**			**Cost**			**Transit time**		
	Road	**IWW**	**Rail**	**Road**	**IWW**	**Rail**	**Road**	**IWW**	**Rail**	**Road**	**IWW**	**Rail**
Road	*-0.10 to -2.41*	*0.59 to 1.91*	*0.60 to 2.06*	*0.00 to -0.33*	*0.00 to 1.87*	*0.00 to 1.81*	*-0.20 to -1.65*	*0.14 to 1.39*	*0.07 to 2.97*	*0.00 to -0.33*	*0.00 to 1.81*	*0.00 to 1.82*
	-0.28	**1.03**	**1.08**	**-0.12**	**0.48**	**0.45**	**-0.24**	**0.86**	**0.93**	**-0.09**	**0.33**	**0.36**
IWW	*0.04 to 0.38*	*-0.43 to -2.56*	*0.03 to 0.28*	*0.00 to 0.21*	*0.00 to -1.93*	*0.00 to 0.16*	*0.04 to 0.27*	*-0.47 to -1.94*	*0.04 to 0.20*	*0.00 to 0.06*	*0.00 to -1.04*	*0.00 to 0.05*
	0.07	**-1.01**	**0.11**	**0.08**	**-1.01**	**0.05**	**0.07**	**-0.89**	**0.06**	**0.03**	**-0.40**	**0.03**
Rail	*0.03 to 2.06*	*0.03 to 1.51*	*-0.78 to -2.07*	*0.00 to 0.28*	*0.00 to 0.19*	*0.00 to -1.78*	*0.02 to 1.30*	*0.01 to 0.70*	*-0.32 to -2.63*	*0.00 to 0.13*	*0.00 to 0.12*	*0.00 to -1.08*
	0.18	**0.24**	**-1.17**	**0.15**	**0.09**	**-0.90**	**0.15**	**0.13**	**-0.92**	**0.07**	**0.04**	**-0.43**

Table 9: Computed elasticities

Conclusion

In this paper, a modal choice model is presented in which network accessibility is used as a third explanatory variable, beside transport cost and transit time.

A special attention is paid to two specific aspects of the model. Firstly, as transport cost and transit time are correlated, multicollinearity must be tackled. For this purpose, a Box-Cox transform is applied to both variables.

Secondly, several possible functional forms of the decay function used to compute the accessibility measure are tested. If the negative exponential form is generally cited in the literature as being appropriate for transport over medium and long distances, we conclude that the exponential-normal curve fits better.

To assess the added value of the introduction of a (relative) accessibility measure in the modal choice model, two performance indicators are presented. The first one compares the correlation between the observed transported volumes (per mode and per group of commodities for every cell of the origin-destination matrix) and those computed by the modal choice models. It comes out that the correlations of the model including the accessibility measure are significantly better. A similar exercise is performed with the "observed" and computed traffic along the roads, inland waterways, and railroads of the digitized networks. Here also, the correlation coefficients are systematically better for the improved model.

Finally, cost and transit-time elasticities are computed. The analysis shows, as may be expected, that the (absolute values of the) elasticities computed with the improved model are somewhat lower.

The inclusion of an accessibility measure in the modal choice model has thus a double advantage for policymakers: it improves the predictive power of the model, and it avoids some overestimations of own and cross-elasticities. Both benefits may increase the accuracy of cost-benefit or socio-economic analysis.

This research was conducted using exclusively publicly available data and open-source software. Ideally, the model estimation process should seamlessly be integrated with the workflow of the modeler, i.e., using a single programing language. However, if one excludes the use of SQL to prepare the data extracted from Nodus and Java to implement the modal choice model in Nodus as a plugin, both R and Python are used. The first one is chosen to estimate the parameters of the decay functions used to compute the accessibility measures and the second is required by Biogeme for the estimation of the logit model. Some additional work should be devoted to solving both facets using the same language.

More fundamentally, the accessibility measure that is used here is defined as a measure of the capacity of a location to be reached from. Hence, the model considers that the modal choice decision is taken at the origin node only, by the shipper, for instance. This is, however, not always true as the customer at the destination node may have a say in the decision. Further investigations are thus needed to properly combine the accessibilities computed at both the origins and the destinations.

5 References

Adeboye, N.O., Fagoyinbo, I.S. and Olatayo, T.O. (2014) 'Estimation of the Effect of Multicollinearity on the Standard Error for Regression Coefficients', *IOSR Journal of Mathematics*, 10(4), pp. 16–20. doi:10.9790/5728-10411620.

Beuthe, M. *et al.* (2014) 'Climate change impacts on transport on the Rhine and Danube: A multimodal approach', *Transportation Research Part D: Transport and Environment*, 27. doi:10.1016/j.trd.2013.11.002.

Beuthe, M. and Bouffioux, Ch. (2008) 'Analysing Qualitative Attributes of Freight Transport from Stated Orders of Preference Experiment', *Journal of Transport Economics and Policy*, 42(1), pp. 105–128.

Beuthe, M., Jourquin, B. and Urbain, N. (2014) 'Estimating Freight Transport Price Elasticity

in Multi-mode Studies: A Review and Additional Results from a Multimodal Network Model', *Transport Reviews* [Preprint]. doi:10.1080/01441647.2014.946459.
Bierlaire, M. (2020) *A short introduction to PandasBiogeme*. TRANSP-OR 200605, p. 22. Available at: http://transp-or.epfl.ch/documents/technicalReports/Bier20.pdf.
Blauwens, G. and Van de Voorde, E. (1988) 'The valuation of time savings in commodity transport', *International Journal of Transport Economics*, 15(1), pp. 77–87.
Box, G.E.P. and Cox, D.R. (1964) 'An Analysis of Transformations', *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), pp. 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x.
Cervero, R. (2002) 'Built environments and mode choice: toward a normative framework', *Transportation Research Part D: Transport and Environment*, 7(4), pp. 265–284. doi:10.1016/S1361-9209(01)00024-4.
Cramer, J.S. (2002) *The Origins of Logistic Regression*. SSRN Scholarly Paper ID 360300. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=360300 (Accessed: 7 December 2021).
De Jong, G. (2003) 'Elasticities and policy impacts in freight transport in Europe', in.
De Jong, G. *et al.* (2004) 'New values of time and reliability in freight transport in the Netherlands', *Association for European Transport* [Preprint].
Dormann, C.F. *et al.* (2013) 'Collinearity: a review of methods to deal with it and a simulation study evaluating their performance', *Ecography*, 36(1), pp. 27–46. doi:10.1111/j.1600-0587.2012.07348.x.
Fridstrøm, L. and Madslien, A. (1994) 'Own account or hire freight: a stated preference analysis', in *7th International Conference on Travel Behaviour*. Valle Nevado, Chile. Available at: https://www.researchgate.net/publication/268221233_Own_account_or_hire_freight_a_stated_preference_analysis.
Gaudry, M. (2016) 'Méthodes Box-Cox, algorithmes de TRIO et demande de transport : trois consignes occamiennes pour faire rendre sens aux coefficients de régression de modèles simples et logistiques discrets ou agrégés'. doi:10.13140/RG.2.1.1102.5044.
Geurs, K.T. and Ritsema van Eck, J.R. (2001) *Accessibility measures: review and applications. Evaluation of accessibility impacts of land-use transport scenarios, and related social and economic impacts*. RIVM report 408505 006. Bilthoven, The Netherlands: Rijksinstituut voor Volksgezondheid en Milieu.
Greene, W.H. (2012) *Econometric analysis*. Seventh edition, international edition. Pearson (Pearson series in economics).
Grothendieck, G. (2013) *Non-linear regression with brute force* [R]. Available at: https://cran.r-project.org/package=nls2.
Guilbault, M. and Soppé, M. (2009) *Apports des enquêtes chargeurs: connaissance des chaînes de transport de marchandises et de leurs déterminants logistiques*. Bron: INRETS (Actes INRETS, n° 121).
Gutiérrez, J., Monzón, A. and Piñero, J.M. (1998) 'Accessibility, Network Efficiency, and Transport Infrastructure Planning', *Environment and Planning A: Economy and Space*, 30(8), pp. 1337–1350. doi:10.1068/a301337.
Hammadou, H. *et al.* (2008) 'How to Incorporate the Spatial Dimension in Destination Choice Models: The Case of Antwerp', *Transportation Planning and Technology*, 31(2), pp. 153–181. doi:10.1080/03081060801948126.
Handy, S.L. and Niemeier, D.A. (1997) 'Measuring Accessibility: An Exploration of Issues and

Alternatives', *Environment and Planning A: Economy and Space*, 29(7), pp. 1175–1194. doi:10.1068/a291175.
Holguín-Veras, J. *et al.* (2021) 'Freight mode choice: Results from a nationwide qualitative and quantitative research effort', *Transportation Research Part A: Policy and Practice*, 143, pp. 78–120. doi:10.1016/j.tra.2020.11.016.
de Jong, G., Gunn, H. and Ben-Akiva, M. (2004) 'A meta-model for passenger and freight transport in Europe', *Transport Policy*, 11(4), pp. 329–344. doi:10.1016/j.tranpol.2004.03.001.
Jourquin, B. (2017) 'Estimation of Travel Time Elasticities with a Multimodal Freight Transport Model Calibrated and Validated on Publicly Available Regional Data', in *14th International Nectar Conference*. Madrid. Available at: https://www.researchgate.net/publication/317580054_Estimation_of_travel_time_elasticities_with_a_multimodal_freight_transport_model_calibrated_and_validated_on_publicly_available_regional_data.
Jourquin, B. (2021) 'Mode choice in strategic freight transportation models: a constrained Box–Cox meta-heuristic for multivariate utility functions', *Transportmetrica A: Transport Science*, pp. 1–21. doi:10.1080/23249935.2021.1937375.
Jourquin, B. and Beuthe, M. (2019) 'Cost, transit time and speed elasticity calculations for the European continental freight transport', *Transport Policy*, 83, pp. 1–12. doi:10.1016/j.tranpol.2019.08.009.
Kim, H.-C., Nicholson, A. and Kusumastuti, D. (2014) 'Freight Transport Mode Choice and Mode Shift in New Zealand: Findings of a Revealed Preference Survey', in Macharis, C. et al. (eds) *Transport and Sustainability*. Emerald Group Publishing Limited, pp. 165–192. doi:10.1108/S2044-994120140000006007.
Kvalseth, T.O. (1983) 'Note on the R2 measure of goodness of fit for nonlinear models', *Bulletin of the Psychonomic Society*, 21(1), pp. 79–80. doi:10.3758/BF03329960.
Liedtke, G. and Carrillo Murillo, D.G. (2012) 'Assessment of policy strategies to develop intermodal services: The case of inland terminals in Germany', *Transport Policy*, 24, pp. 168–178. doi:10.1016/j.tranpol.2012.06.002.
Madaan, J. *et al.* (2021) 'Mode choice problem in context of freight transportation: A systematic review and bibliometric analysis', *International Journal of Logistics Systems and Management*, 1(1), p. 1. doi:10.1504/IJLSM.2021.10039381.
McFadden, D. (1973) 'Conditional Logit Analysis of Qualitative Choice Behaviour', in Zarembka, P. (ed.) *Frontiers in Econometrics*. New York, NY, USA: Academic Press New York, pp. 105–142.
Morris, J.M., Dumble, P.L. and Wigan, M.R. (1979) 'Accessibility indicators for transport planning', *Transportation Research Part A: General*, 13(2), pp. 91–109. doi:10.1016/0191-2607(79)90012-8.
Pooler, J.A. (1995) 'The use of spatial separation in the measurement of transportation accessibility', *Transportation Research Part A: Policy and Practice*, 29(6), pp. 421–427. doi:10.1016/0965-8564(95)00013-E.
Quittelier, P. (2021) *Modèles de choix modal pour le transport de marchandises*. Master Thesis. UCL - Louvain School of Management. Available at: http://hdl.handle.net/2078.1/thesis:31450.
Reggiani, A., Bucci, P. and Russo, G. (2011) 'Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network', *International Regional Science Review*, 34(2), pp. 230–252. doi:10.1177/0160017610387296.

Rich, J., Holmblad, P.M. and Hansen, C.O. (2009) 'A weighted logit freight mode-choice model', *Transportation Research Part E: Logistics and Transportation Review*, 45(6), pp. 1006–1019. doi:10.1016/j.tre.2009.02.001.

Rodrigue, J.-P. (2020) *The geography of transport systems*. Fifth edition. Abingdon, Oxon ; New York, NY: Routledge.

Rothengatter, W. (2019) 'Megaprojects in transportation networks', *Transport Policy*, 75, pp. A1–A15. doi:10.1016/j.tranpol.2018.08.002.

Spiess, A.-N. and Neumeyer, N. (2010) 'An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach', *BMC Pharmacology*, 10(1), p. 6. doi:10.1186/1471-2210-10-6.

Szimba, E. *et al.* (2012) 'ETISplus Database Content and Methodology'. doi:10.13140/RG.2.2.16768.25605.

Thomas, I. *et al.* (2003) 'Accessibility to freight transport networks in Belgium: a geographical approach', *Tijdschrift voor Economische en Sociale Geografie*, 94(4), pp. 424–438. doi:10.1111/1467-9663.00271.

Yi, Y. and Kim, E. (2018) 'Spatial economic impact of road and railroad accessibility on manufacturing output: Inter-modal relationship between road and railroad', *Journal of Transport Geography*, 66, pp. 144–153. doi:10.1016/j.jtrangeo.2017.11.017.