

Integrating 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 of 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 inclusion of an accessibility measure has thus a double advantage for policymakers: it improves the predictive power of the freight transport model, giving more accurate estimated traffics on the modal networks, and it avoids overestimations of own and cross-elasticities, which has an obvious importance in the framework of transport policy decisions.

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 (Beuthe et al., 2014b; de Jong et al., 2004, 2016; Liedtke and Carrillo Murillo, 2012; Rothengatter, 2019). However, when applied to inter-regional or international areas, they are often set up using aggregated demand data (origin-destination matrixes containing annual transported tons between region's centroids for instance) and utility functions that rely on basic explanatory variables, such as transportation costs or transit times.

Transportation costs 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 Bouffieux, 2008). In-depth surveys carried out with shippers, carriers and receivers conclude that network accessibility also plays an important role (Holguín-Veras et al., 2021; Kim et al., 2014; 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 a modal choice model should therefore improve its robustness and explanatory power. Such a measure can be computed using a weight representing the attractiveness of a centroid, itself weighted using a decay function that can for instance take the trip lengths as input. These lengths can rather easily be gathered or computed along with transportation 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.

Therefore, this paper examines to what extent accessibility, when used along with transportation cost and transit time as a third explanatory variable, improves the performance of a simple modal choice model. It also presents a set of own and cross-elasticities computed using the improved model, which are compared to the values obtained without the accessibility measure.

In this context, a particular attention must be paid to two aspects of the model to setup.

Firstly, costs and transit times are most 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 (Adeboye et al., 2014; Greene, 2012). The resulting estimators may also 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 not what is expected.

Secondly, the computation of an appropriate accessibility index merits some 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 2.2) representing the degree of interconnection between a particular reference location and all or a set of other locations in the area (Reggiani et al., 2011; Thomas et al., 2003). The decay (or impedance) function used to compute the accessibility measure can take many forms, but the negative exponential form has been often used because it is supposed to more adequately 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 et al., 2011). Therefore, several functional forms will be tested in this paper.

The methodological aspects of the model are introduced in section 2, while section 3 describes the dataset used for the case study. The results and the robustness of the model are examined in section 4. Before a general conclusion, a set of own and cross elasticities are presented and discussed in section 5.

2 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. As a reminder, the general form of a multinomial logit model can be written as:

$$P_m = \frac{e^{U_m}}{\sum_n e^{U_n}} \quad (1)$$

Where:

P_m : Probability of choosing mode m

U_m : Utility of mode m

U_n : Utility of mode n

A first simple model is set up here, 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 impacts of time savings. Keeping both transportation costs and transit times as separate independent variables makes thus sense. The utility function for this first model can be written as:

$$U_m = \alpha C_m + \beta T_m + \delta_m \quad (2)$$

Where:

C_m = Transportation cost for mode m

T_m = Transit time for mode m

α, β, δ_m : Parameters to estimate

As C and T are alternative specific rather than being individual 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.

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 are 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, as the transport demand is embedded in origin-destination (OD) matrixes containing aggregated data (of total annual transported quantities for instance), a weighted logit methodology is applied, whereby the transported tonnages weight the mode choice observations in the Log-Likelihood functions (Jourquin and Beuthe, 2019; Rich et al., 2009).

The second model includes an accessibility measure in the utility function:

$$U_m = \alpha C_m + \beta T_m + \sigma_m A_m^o + \delta_m \quad (3)$$

Where:

A_m^o : Accessibility measure of origin location o for mode m (see section 2.2).

σ_m : Parameter to estimate

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

Both models are rather simple and presented here in an Occam's razor spirit. A special attention must, however, be paid to the correlation between C_m and T_m (multicollinearity) and to the definition of the accessibility measure A_m^o . Both aspects are discussed in the next subsections.

2.1 Handling multicollinearity

Several methods can be found in the literature to overcome multicollinearity (Dormann et al., 2013; Greene, 2012), among which the Box-Cox transforms (Box and Cox, 1964) of the independent variables. Beside the fact that this technique can be useful to improve the Log-Likelihood of a model, it is sometimes considered as an elegant and efficient way to help to 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 inflection of data using power functions.

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

By “bending” the functional form, the used value for λ offers a lot of flexibility, providing more degrees of freedom to the modeler: values of $\lambda < 1$ produce convex curves, while using a $\lambda > 1$ results in a concave inflection. To include the Box-Cox transforms of C and T , the utility functions of both models (equations 2 and 3) can thus be rewritten¹:

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

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

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

2.2 Computing the accessibility measure

Even if a very simple accessibility measure such as the number of centroids that can be reached by each mode from a given origin can be used (Quittelier, 2021), the General Accessibility Measure (Gutiérrez et al., 1998; Morris et al., 1979; Pooler, 1995), a well-established measure in the literature, will be applied here. It can be written as:

$$A^o = \sum_d w^d f(L^{od}) \quad (7)$$

Where:

A^o : Accessibility at origin o

w^d : Weight representing the attractiveness of destination d

L^{od} : Measure of separation (trip length in the case presented in section 4) between o and d

$f(\dots)$: Decay function

This accessibility measure is thus defined as a measure of the capacity of a location to be reached from. Its value directly depends on the weights w^d (transported quantities to destination d). However, as explained earlier, a weighted logit is implemented here. In order to avoid a “double counting” of w^d in the model, a relative accessibility index (Thomas et al., 2003) will be used, which can be written as:

¹ Obviously, as specified in equation 4, $\log(C_m)$ or $\log(T_m)$ must be used if λ_c or $\lambda_t = 0$.

$$\Lambda^o = \frac{\sum_d w^d f(L^{od})}{n \sum_d w^d} \quad (8)$$

Where n is the number of centroids in the origin-destination matrix. Therefore, the utility function (6) becomes:

$$U_m = \alpha \left(\frac{C_m^{\lambda_c} - 1}{\lambda_c} \right) + \beta \left(\frac{T_m^{\lambda_t} - 1}{\lambda_t} \right) + \sigma_m \Lambda_m^o + \delta_m \quad (9)$$

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 can also be strongly correlated with C and T , especially if the trip length is used as a measure of separation. However, the value of L is diluted in the decay function while w^d plays a preponderant role.

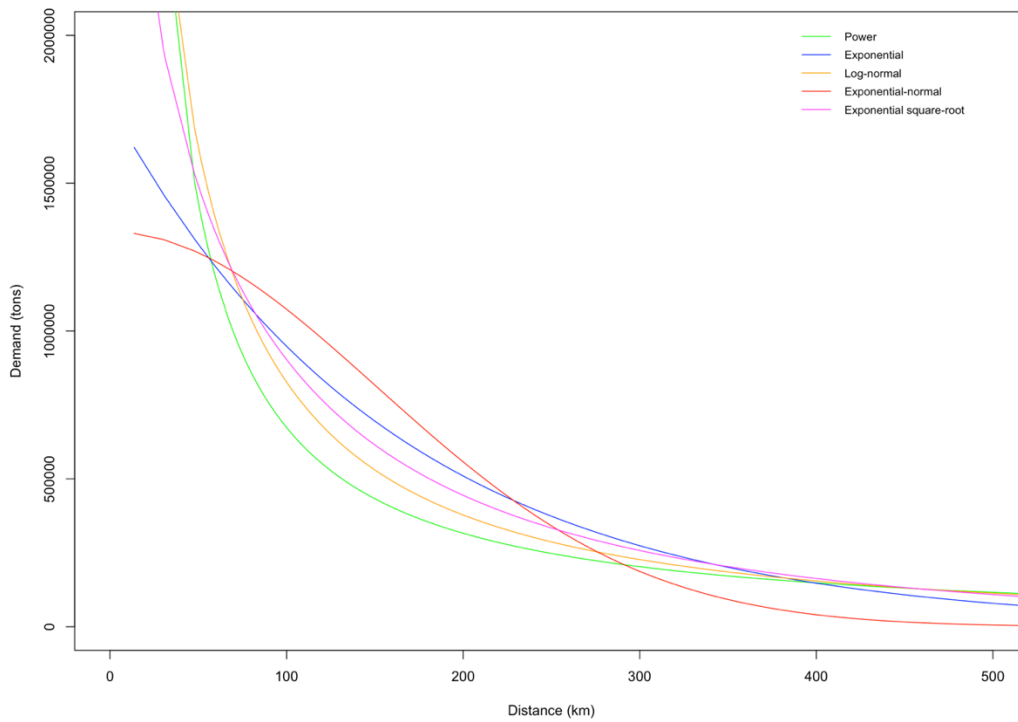


Figure 1: Shapes of the tested decay functions

Most authors use the negative exponential form represented by equation 11 for the decay function, as it is considered as being closely tied to travel behavior (Handy and Niemeier, 1997; Thomas et al., 2003). But other functional forms can be considered. Therefore, the five specifications presented in Reggiani, Bucci, and Russo (2011) and illustrated in Figure 1, will be tested in section 4.2:

I - Power: $f(L^{od}) = (L^{od})^{-\gamma_1} \quad (10)$

$$\text{II - Exponential:} \quad f(L^{od}) = e^{-\gamma_2 L^{od}} \quad (11)$$

$$\text{III - Log-normal:} \quad f(L^{od}) = e^{-\gamma_3 (\log(L^{od}))^2} \quad (12)$$

$$\text{IV - Exponential normal:} \quad f(L^{od}) = e^{-\gamma_4 (L^{od})^2} \quad (13)$$

$$\text{V - Exponential square root:} \quad f(L^{od}) = e^{-\gamma_5 \sqrt{L^{od}}} \quad (14)$$

In these equations, γ_1 , γ_2 , γ_3 , γ_4 and γ_5 are the parameters to estimate.

3 Dataset description

The dataset used for the case presented in 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 transportation networks. Unfortunately, to the contrary to what was initially announced, the data has never been updated.

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 groups of commodities (NST/R chapters 0 to 9, see Appendix). 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 (Figure 2).

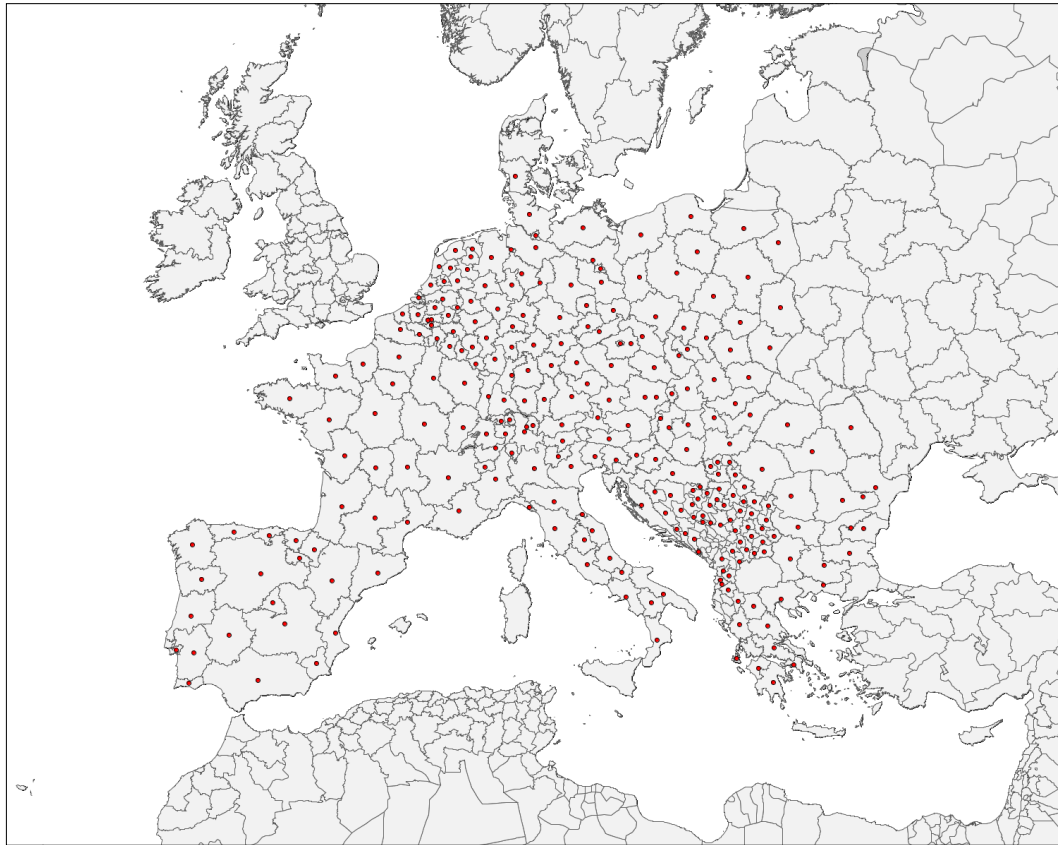


Figure 2: NUTS2 regions and retained centroids

The datasets corresponding to the different groups of commodities (Table 1) are quite different in size (number of OD cells), volumes (Mios tons/year) and modal split (market shares). The models presented in section 2 will be solved for each group of commodities separately, which is useful to test robustness.

NST-R	OD cells			Volumes (10 ⁶ t)			Market shares (%)		
	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 1: Content of the OD matrixes

The modal matrixes are combined, and the modal choice models will be applied to these merged 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.

4 Application and model performance

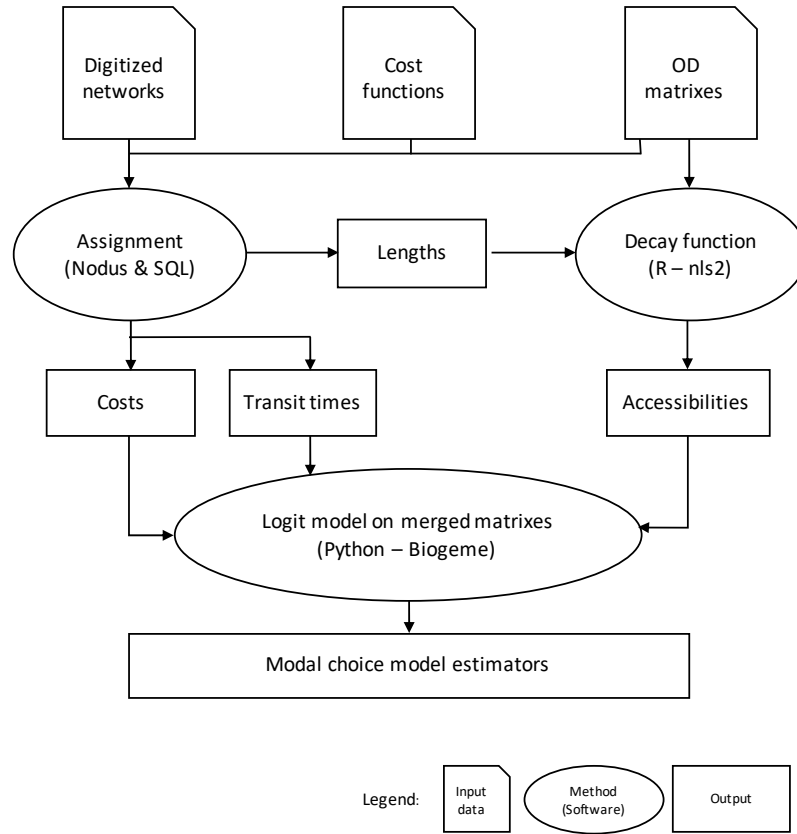


Figure 3: Model workflow

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

Figure 3.

4.1 Compute costs and transit times

The OD matrixes and the transportation 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 indeed retrieving, for each OD pair, each mode m and each group of commodities g , the total cost of transport C_m^g , the total transit time T_m^g and the trip length L_m^g . These are thus constructed variables, retrieved from the results of an assignment of each modal OD matrix on its corresponding digitized transportation network.

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 equal to:

$$C_m^g = ld_cost_m^g + ul_cost_m^g + \sum_l^R mv_cost_l^g \quad (15)$$

Where:

ld_cost : loading cost
 ul_cost : unloading cost
 mv_cost : moving cost
 R : the set of successive links representing the route

The loading and unloading costs are fixed costs, though they vary with the mode and the transported goods. The transshipment costs associated with intermodal transport chains are not explicitly modeled in this paper. The loading factors of the vehicles are taken from the ECCONET research project (Beuthe et al., 2014b). 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. It considers all the carrier's costs during transport: labor, capital invested in vehicles, fuel, maintenance, insurance, 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:

$$mv_cost_{l,m} = \frac{Average\ speed_m}{Speed_{l,m}} * length_l * unit\ mv_cost_m \quad (16)$$

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 transportation network and $Speed_{l,m}$ is the average speed on link l .

Similarly, the total transit time T_m^g 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:

$$T_m = ld_duration_m + ul_duration_m + \sum_l^R mv_duration_{l,m} \quad (17)$$

$$\text{with } mv_duration_{l,m} = length_l / speed_{l,m} \quad (18)$$

4.2 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 3. As shown in Figure 1, 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 nonlinear functions (Kvalseth, 1983; Spiess and Neumeyer, 2010). Indeed, 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 must be used instead. Table 2 compares the relative residual standard error ($\frac{RSE_{model}}{RSE_{best\ model}}$) of each specification (I-V, as given by equations 10 to 14). It comes out that the differences between the specifications are limited, but that the exponential normal decay function (IV, equation 13) always performs best for road transport. For the two other modes, no decay function appears to outperform the others. Altogether, the exponential normal formulation will be used in this paper to compute the relative accessibility measure $\mathcal{A}^o$.

NST/R	Road					IWW					Rail				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
0	1.04	1.01	1.03	1.00	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.00	1.01	1.00
1	1.05	1.01	1.03	1.00	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.02	1.01	1.02	1.00	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
3	1.03	1.00	1.02	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.04	1.01	1.03	1.00	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.02	1.00	1.01	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.01	1.00
6	1.04	1.01	1.03	1.00	1.02	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.02	1.01	1.02	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00
8	1.05	1.01	1.03	1.00	1.02	1.00	1.01	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
9	1.04	1.01	1.03	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00

Table 2: Relative residual standard error for the 5 decay functions (per mode and NST/R)

4.3 Logit model estimation

Now that all the explanatory variables (C , T and $\mathcal{A}^o$) can be computed, the modal choice models can be estimated. Table 3 presents the estimators for the model without accessibility while Table 4 contains those of the model with the accessibility measure. 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		α	β	δ_{iww}	δ_{rail}	λ_c	λ_t
NST/R							
0	Value	-0.414	-0.939	-2.980	-1.970	0.393	-1.080
	t-test	-13.900	-5.700	-36.200	-64.800	22.300	-9.620
1	Value	-1.490	-1.260	-1.780	-3.170	0.024	-0.170
	t-test	-14.600	-12.500	-7.760	-17.600	0.920	-5.900
2	Value	-2.900	0.000	-3.870	-1.410	0.065	0.807
	t-test	-28.900	0.000	-55.800	-40.600	6.150	0.000
3	Value	-0.415	-0.833	1.050	0.438	0.585	0.155
	t-test	-14.200	-9.430	4.600	2.600	31.500	6.620
4	Value	-4.930	-2.620	1.490	2.630	-0.216	0.005
	t-test	-9.490	-13.400	4.550	9.090	-4.050	0.229
5	Value	-0.522	-0.005	-2.850	-0.324	0.351	1.150
	t-test	-12.000	-2.910	-41.300	-7.420	13.600	16.000
6	Value	-0.236	-0.781	2.350	2.090	0.529	0.277
	t-test	-10.200	-16.700	12.900	14.800	21.100	20.800
7	Value	-3.430	0.000	-4.950	-1.380	-0.099	0.127
	t-test	-28.400	0.000	-67.700	-39.000	-10.500	0.000
8	Value	-1.230	-0.001	-3.400	-1.440	0.225	1.690
	t-test	-22.300	-3.410	-66.000	-53.300	16.700	26.600
9	Value	-0.806	-0.012	-1.540	-0.353	0.143	1.240
	t-test	-8.720	-5.060	-16.800	-5.280	3.760	26.600

Table 3: Estimated parameters for the model without accessibility

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 model that includes the accessibility measure. This is most probably 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 3 and Table 4 will be further used.

The Box-Cox parameter λ_c for the first model (Table 3) is lower than 1 for all groups of commodities, which means that the negative effect of a cost increase is marginally decreasing. Adding the accessibility measure in the model (Table 4) doesn't change this. Altogether, except for NST/R 4, the value of λ_c doesn't change significantly between both models.

The interpretation of λ_t is different as its value is higher than 1 for NST/R groups 5 (metal products), 8 (chemical products) and 9 (machinery and containers), meaning that, for these commodities, the negative effect of a longer trip marginally increases with transit time.

Estimators		α	β	σ_{road}	σ_{iww}	σ_{rail}	δ_{iww}	δ_{rail}	λ_c	λ_t
NST/R										
0	Value	-0.346	-1.760	7.070	1.590	3.460	-2.180	-1.790	0.402	-0.992
	t-test	-11.100	-9.730	36.300	43.200	89.700	-20.800	-39.300	18.600	-12.000
1	Value	-1.250	-3.040	3.400	0.671	1.780	-3.170	-4.310	0.062	-0.932
	t-test	-12.000	-8.630	17.400	68.100	53.800	-15.500	-25.200	1.890	-8.040
2	Value	-1.260	0.000	3.590	1.540	1.420	-3.200	-0.856	0.255	0.065
	t-test	-14.400	0.000	44.200	53.400	59.300	-34.900	-18.100	13.200	0.000
3	Value	-1.380	-0.002	2.210	0.745	1.860	-4.190	-3.270	0.398	1.320
	t-test	-20.200	-1.600	30.200	76.300	60.200	-38.000	-47.200	28.400	9.920
4	Value	-74.400	-4.630	4.240	0.264	2.720	4.120	4.350	-2.030	-0.214
	t-test	-6.580	-36.500	31.000	50.400	67.800	21.500	30.100	-20.500	-12.900
5	Value	-1.110	0.000	2.340	0.846	2.680	-3.730	-0.706	0.138	1.410
	t-test	-12.700	-1.070	48.700	55.700	90.900	-57.500	-20.300	5.520	7.480
6	Value	-0.559	-0.758	3.250	2.470	2.510	-0.358	0.209	0.388	0.129
	t-test	-13.300	-11.900	32.000	64.300	78.300	-1.720	1.400	18.400	6.500
7	Value	-3.660	-1.010	2.760	1.460	3.100	-4.820	-1.360	-0.128	-0.810
	t-test	-21.000	-6.060	51.000	62.400	72.400	-30.100	-11.400	-6.680	-6.630
8	Value	-1.430	0.000	1.790	1.440	2.260	-4.050	-1.810	0.184	1.820
	t-test	-22.400	-2.470	25.100	75.600	65.500	-72.700	-65.100	13.600	21.400
9	Value	-1.440	-0.001	3.770	0.608	2.870	-2.890	-1.330	-0.051	1.600
	t-test	-8.760	-2.570	28.800	86.400	78.200	-35.100	-22.100	-1.330	18.700

Table 4: Estimated parameters for the model with accessibility

The case of NST/3 R (petroleum products) is the only one for which the Box-Cox “bending” changes from a convex to a concave shape between both models. Indeed $\lambda_t = 0.155$ for the first models and $\lambda_t = 1.320$ when the accessibility measure is introduced, while λ_c for both models remain similar. This means that accessibility plays an important role in the modal choice decision for petroleum products. Unlike for the other groups of commodities, petroleum products are mainly concentrated on two origins nodes only, i.e., the ports of Antwerp and Rotterdam, while the destinations are spread all over the studied area. Modal accessibility from these two nodes is thus crucial.

Finally, the estimators for the relative accessibility measure σ_m are all positive, indicating that a better accessibility for a mode makes it more attractive, which is an expected result. It is also worthwhile to note that $\sigma_{road} > \sigma_{rail} > \sigma_{iww}$ for all groups, except for NST/R 2 (solid fuels), for which σ_{iww} is slightly higher than σ_{rail} . This may be linked to the fact that the connectivity is unequal between transportation modes. Indeed, among the almost 345,000 OD pairs included in the origin-destination matrix, 98% are effectively disserved by road, 14% by rail and only 3% by inland waterways.

4.4 Comparative performances

The Log-likelihood and the Akaike information criterion of the model with accessibility 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 Pearson's correlation coefficients r between the computed modal tonnages by the model and those gathered from the ETISPlus modal matrixes for each OD pair, each group of commodities and each mode is presented in Table 5.

The figures of the last row (“Aggregate”) represent the weighted average value of r , which

depends 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 groups 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 the accessibility measure in the utility function substantially improves the performance of the modal choice model.

NST/R	$r_{C,T}$			r_{C,T,Δ^0}		
	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 5: Pearson's correlations between observed and computed quantities at the OD level

NST/R	$r_{C,T,L} - r_{C,T}$		
	Road	IWW	Rail
0	0.00	0.00	0.00
1	0.00	0.00	0.00
2	0.00	-0.06	0.00
3	-0.02	-0.03	0.01
4	0.00	-0.01	0.00
5	0.00	-0.04	0.02
6	0.00	-0.02	0.00
7	0.01	-0.02	0.01
8	0.00	-0.01	0.00
9	0.00	0.00	0.00
Aggregate	0.00	-0.02	0.00

Table 6: Correlation difference between the $U(C, T, L)$ and $U(C, T)$ models

To confirm this, a model with a utility function including a Box-Cox transformed L (instead of Δ^0) as a third independent variable is tested². Table 6 shows that, unlike what appears in Table 5 when Δ^0 is used as a third explanatory variable, L doesn't improve the model. In several cases the correlation is even somewhat lower than the one computed with C and T only.

² Biogeme fails to properly estimate the $U(C, T, L)$ model as the three variables are heavily correlated. Therefore, the figures presented in Table 6 were obtained using the methodology presented in (Jourquin, 2021).

NST/R	Model 1 (cost & transit time)			Model 2 (cost. transit time & accessibility)		
	Road	IWW	Rail	Road	IWW	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 7: Correlations between “observed” and computed traffic on the networks

The figures presented in Table 5 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 the software using a specific modal choice plugin that uses the estimators presented in Table 3 and Table 4 along with, for the second model, the computed relative accessibility measure for each centroid. Table 7 gives the resulting correlation coefficients between the two sets of flows³ assigned to the same links. The figures show that the model with accessibility always performs better. The “Aggregate” line is particularly interesting as the figures are computed regardless of the type of commodities. They are thus related to the traffic on the networks (at least for non-empty vehicles), which is particularly interesting for policy-makers.

5 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 modal split and 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 ε (Equation 19):

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.

$$\varepsilon = \frac{(Q_1 - Q_2)/Q_1}{(V_1 - V_2)/V_1} \quad (19)$$

Table 8 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 (Beuthe et al., 2014a; De Jong, 2003; Rich et al., 2009). The cost elasticities computed for the model without accessibility can also be directly compared to those published by Jourquin and Beuthe (2019) as they were obtained using (almost) the same dataset and a similar (but not identical) Box-Cox transform of the C and T variables. They are written in parentheses in Table 8.

	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	<i>-0.10</i>	<i>0.59</i>	<i>0.60</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>-0.20</i>	<i>0.14</i>	<i>0.07</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>
	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>
	<i>-2.41</i>	<i>1.91</i>	<i>2.06</i>	<i>-0.33</i>	<i>1.87</i>	<i>1.81</i>	<i>-1.65</i>	<i>1.39</i>	<i>2.97</i>	<i>-0.33</i>	<i>1.81</i>	<i>1.82</i>
	-0.28	1.03	1.08	-0.12	0.48	0.45	-0.24	0.86	0.93	-0.09	0.33	0.36
	(-0.29)	(1.07)	(1.13)									
IWW	<i>0.04</i>	<i>-0.43</i>	<i>0.03</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.04</i>	<i>-0.47</i>	<i>0.04</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>
	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>
	<i>0.38</i>	<i>-2.56</i>	<i>0.28</i>	<i>0.21</i>	<i>-1.93</i>	<i>0.16</i>	<i>0.27</i>	<i>-1.94</i>	<i>0.20</i>	<i>0.06</i>	<i>-1.04</i>	<i>0.05</i>
	0.07	-1.01	0.11	0.08	-1.01	0.05	0.07	-0.89	0.06	0.03	-0.40	0.03
	(0.07)	(-1.05)	(0.12)									
Rail	<i>0.03</i>	<i>0.03</i>	<i>-0.78</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.02</i>	<i>0.01</i>	<i>-0.32</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>
	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>	<i>to</i>
	<i>2.06</i>	<i>1.51</i>	<i>-2.07</i>	<i>0.28</i>	<i>0.19</i>	<i>-1.78</i>	<i>1.30</i>	<i>0.70</i>	<i>-2.63</i>	<i>0.13</i>	<i>0.12</i>	<i>-1.08</i>
	0.18	0.24	-1.17	0.15	0.09	-0.90	0.15	0.13	-0.92	0.07	0.04	-0.43
	(0.20)	(0.24)	(-1.26)									

Table 8: Computed cost and transit time elasticities for both models⁴

It appears that the computed elasticities obtained with the second model are somewhat lower (in absolute value). This can be explained by the fact that the inclusion of a third explanatory variable improves the performance of the model by extracting its explanatory power not only from the intercepts δ_{iww} and δ_{rail} , 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 interregional or international freight transport models which modal choice formulas rely on costs and transit times only tend to be a bit overestimated. Obviously, this can have an important impact on the conclusions of a cost-benefit or a socio-economic analysis of a projected new infrastructure for instance, especially if its estimated profitability is low.

⁴ Values in parenthesis are aggregated values taken from Jourquin and Beuthe, 2019

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 costs and transit times are correlated, multicollinearity must be tackled. For this purpose, a Box-Cox transform is applied to both independent variables.

Secondly, several possible functional forms for the decay function to use 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 somewhat 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 transportation networks. Here also, the correlation coefficients are systematically better for the model embedding the accessibility measure.

Finally, cost and transit-time elasticities are computed. The analysis shows that the (absolute values of the) elasticities computed with the improved model are somewhat lower than those obtained without taking accessibility into account.

While keeping the econometric model simple, 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 programming 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 used to estimate the parameters of the decay functions needed 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 only, by the shipper for instance. This is, however, not always true as the customer at the destination 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.

Finally, the transportation cost functions could be refined to include more elements, such as external costs for instance. The latter will indeed undoubtedly be more and more integrated into the attention of decision-makers in the near future.

6 References

- Adeboye, N.O., Fagoyinbo, I.S., Olatayo, T.O., 2014. Estimation of the Effect of Multicollinearity on the Standard Error for Regression Coefficients. *IOSR J. Math.* 10, 16–20. <https://doi.org/10.9790/5728-10411620>
- Beuthe, M., Bouffieux, Ch., 2008. Analysing Qualitative Attributes of Freight Transport from Stated Orders of Preference Experiment. *J. Transp. Econ. Policy* 42, 105–128.
- Beuthe, M., Jourquin, B., Urbain, N., 2014a. Estimating Freight Transport Price Elasticity in Multi-mode Studies: A Review and Additional Results from a Multimodal Network Model. *Transp. Rev.* <https://doi.org/10.1080/01441647.2014.946459>
- Beuthe, M., Jourquin, B., Urbain, N., Lingemann, I., Ubbels, B., 2014b. Climate change impacts on transport on the Rhine and Danube: A multimodal approach. *Transp. Res. Part Transp. Environ.* 27. <https://doi.org/10.1016/j.trd.2013.11.002>
- Bierlaire, M., 2020. A short introduction to PandasBiogeme (No. TRANSP-OR 200605), Series on Biogeme.
- Blauwens, G., Van de Voorde, E., 1988. The valuation of time savings in commodity transport. *Int. J. Transp. Econ.* 15, 77–87.
- Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B Methodol.* 26, 211–243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Cervero, R., 2002. Built environments and mode choice: toward a normative framework. *Transp. Res. Part Transp. Environ.* 7, 265–284. [https://doi.org/10.1016/S1361-9209\(01\)00024-4](https://doi.org/10.1016/S1361-9209(01)00024-4)
- Cramer, J.S., 2002. The Origins of Logistic Regression (SSRN Scholarly Paper No. ID 360300). Social Science Research Network, Rochester, NY.
- De Jong, G., 2003. Elasticities and policy impacts in freight transport in Europe.
- De Jong, G., Bakker, S., Pieters, M., Wortelboer, P., 2004. New values of time and reliability in freight transport in the Netherlands. *Assoc. Eur. Transp.*
- de Jong, G., Gunn, H., Ben-Akiva, M., 2004. A meta-model for passenger and freight transport in Europe. *Transp. Policy* 11, 329–344. <https://doi.org/10.1016/j.tranpol.2004.03.001>
- de Jong, G., Tavasszy, L., Bates, J., Grønland, S.E., Huber, S., Kleven, O., Lange, P., Ottemöller, O., Schmorak, N., 2016. The issues in modelling freight transport at the national level. *Case Stud. Transp. Policy* 4, 13–21. <https://doi.org/10.1016/j.cstp.2015.08.002>
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Fridstrøm, L., Madslien, A., 1994. Own account or hire freight: a stated preference analysis, in: 7th International Conference on Travel Behaviour. Valle Nevado, Chile.

- 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. <https://doi.org/10.13140/RG.2.1.1102.5044>
- Geurs, K.T., 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 (No. RIVM report 408505 006). Rijksinstituut voor Volksgezondheid en Milieu, Bilthoven, The Netherlands.
- Greene, W.H., 2012. Econometric analysis, Seventh edition, international edition. ed, Pearson series in economics. Pearson.
- Grothendieck, G., 2013. Non-linear regression with brute force.
- Guilbault, M., Soppé, M., 2009. Apports des enquêtes chargeurs: connaissance des chaînes de transport de marchandises et de leurs déterminants logistiques, Actes INRETS. INRETS, Bron.
- Gutiérrez, J., Monzón, A., Piñero, J.M., 1998. Accessibility, Network Efficiency, and Transport Infrastructure Planning. *Environ. Plan. Econ. Space* 30, 1337–1350. <https://doi.org/10.1068/a301337>
- Hammadou, H., Thomas, I., Verhetsel, A., Witlox, F., 2008. How to Incorporate the Spatial Dimension in Destination Choice Models: The Case of Antwerp. *Transp. Plan. Technol.* 31, 153–181. <https://doi.org/10.1080/03081060801948126>
- Handy, S.L., Niemeier, D.A., 1997. Measuring Accessibility: An Exploration of Issues and Alternatives. *Environ. Plan. Econ. Space* 29, 1175–1194. <https://doi.org/10.1068/a291175>
- Holguín-Veras, J., Kalahasthi, L., Campbell, S., González-Calderón, C.A., (Cara) Wang, X., 2021. Freight mode choice: Results from a nationwide qualitative and quantitative research effort. *Transp. Res. Part Policy Pract.* 143, 78–120. <https://doi.org/10.1016/j.tra.2020.11.016>
- Jourquin, B., 2021. Mode choice in strategic freight transportation models: a constrained Box–Cox meta-heuristic for multivariate utility functions. *Transp. Transp. Sci.* 1–21. <https://doi.org/10.1080/23249935.2021.1937375>
- 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.
- Jourquin, B., Beuthe, M., 2019. Cost, transit time and speed elasticity calculations for the European continental freight transport. *Transp. Policy* 83, 1–12. <https://doi.org/10.1016/j.tranpol.2019.08.009>
- Kim, H.-C., Nicholson, A., Kusumastuti, D., 2014. Freight Transport Mode Choice and Mode Shift in New Zealand: Findings of a Revealed Preference Survey, in: Macharis, C., Melo, S., Woxenius, J., Lier, T.V. (Eds.), *Transport and Sustainability*. Emerald Group Publishing Limited, pp. 165–192. <https://doi.org/10.1108/S2044-994120140000006007>
- Kvalseth, T.O., 1983. Note on the R² measure of goodness of fit for nonlinear models. *Bull. Psychon. Soc.* 21, 79–80. <https://doi.org/10.3758/BF03329960>
- Liedtke, G., Carrillo Murillo, D.G., 2012. Assessment of policy strategies to develop intermodal services: The case of inland terminals in Germany. *Transp. Policy* 24, 168–178. <https://doi.org/10.1016/j.tranpol.2012.06.002>
- Madaan, J., Dwivedi, A., Shardeo, V., Patil, A., 2021. Mode choice problem in context of freight transportation: A systematic review and bibliometric analysis. *Int. J. Logist. Syst. Manag.* 1, 1. <https://doi.org/10.1504/IJLSM.2021.10039381>
- McFadden, D., 1973. Conditional Logit Analysis of Qualitative Choice Behaviour, in:

- Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press New York, New York, NY, USA, pp. 105–142.
- Morris, J.M., Dumble, P.L., Wigan, M.R., 1979. Accessibility indicators for transport planning. *Transp. Res. Part Gen.* 13, 91–109. [https://doi.org/10.1016/0191-2607\(79\)90012-8](https://doi.org/10.1016/0191-2607(79)90012-8)
- Pooler, J.A., 1995. The use of spatial separation in the measurement of transportation accessibility. *Transp. Res. Part Policy Pract.* 29, 421–427. [https://doi.org/10.1016/0965-8564\(95\)00013-E](https://doi.org/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.
- Reggiani, A., Bucci, P., Russo, G., 2011. Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network. *Int. Reg. Sci. Rev.* 34, 230–252. <https://doi.org/10.1177/0160017610387296>
- Rich, J., Holmblad, P.M., Hansen, C.O., 2009. A weighted logit freight mode-choice model. *Transp. Res. Part E Logist. Transp. Rev.* 45, 1006–1019. <https://doi.org/10.1016/j.tre.2009.02.001>
- Rodrigue, J.-P., 2020. *The geography of transport systems*, Fifth edition. ed. Routledge, Abingdon, Oxon ; New York, NY.
- Rothengatter, W., 2019. Megaprojects in transportation networks. *Transp. Policy* 75, A1–A15. <https://doi.org/10.1016/j.tranpol.2018.08.002>
- Spiess, A.-N., Neumeyer, N., 2010. An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol.* 10, 6. <https://doi.org/10.1186/1471-2210-10-6>
- Szimba, E., Kraft, M., Ihrig, J., Schimke, A., Schnell, O., Kawabata, Y., Newton, S., Breemersch, T., Versteegh, R., Meijeren, J.V., Jin-Xue, H., Stasio, C.D., Fermi, F., 2012. ETISplus Database Content and Methodology. <https://doi.org/10.13140/RG.2.2.16768.25605>
- Thomas, I., Hermia, J.-P., Vanelslander, T., Verhetsel, A., 2003. Accessibility to freight transport networks in Belgium: a geographical approach. *Tijdschr. Voor Econ. En Soc. Geogr.* 94, 424–438. <https://doi.org/10.1111/1467-9663.00271>
- Yi, Y., Kim, E., 2018. Spatial economic impact of road and railroad accessibility on manufacturing output: Inter-modal relationship between road and railroad. *J. Transp. Geogr.* 66, 144–153. <https://doi.org/10.1016/j.jtrangeo.2017.11.017>

Appendix: The NST/R groups of commodities

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