Descriptive modeling of freight tour formation: a shipment-based approach

Sebastiaan Thoen[a,b,*],

Lóránt Tavasszy[a,c]

Michiel de Bok[a,b]

Goncalo Homem de Almeida Correia[a]

Ron van Duin[c,d]

[a] Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg1, Delft, 2628 CN, the Netherlands

[b] Significance, Grote Marktstraat 47, The Hague, 2511 BH, the Netherlands

[c] Department of Engineering Systems and Services, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5 Delft, 2628 BC, the Netherlands

[d] Knowledge Center Sustainable Port Cities, Rotterdam University of Applied Sciences, Heijplaatstraat 23 Rotterdam, 3089 JB, the Netherlands

[*] Corresponding author

E-mail address: thoen@significance.nl

Declarations of interest: none.

Abstract

An increasing amount of research is dedicated to the consideration of tour formation in freight transportation demand models. While empirical tour formation models so far have been starting from limiting assumptions about the resulting trips, we develop a generalized shipment-based model. We formulate a random utility model embedded in an iterative algorithm to construct tours through the incremental allocation of shipments. It considers different objectives and constraints and acknowledges the difference between commodity, vehicle and location types. Parameters are estimated on a large and comprehensive shipment database. The model reproduces observed tour patterns well for the given set of shipments.

Keywords: freight transportation modeling, tour formation, discrete choice model, road transport survey

1. Introduction

One of the planning activities that firms undertake to prepare freight movements is tour formation. Here, planners combine pick-up and delivery locations for several shipments into round trips. Tour formation is becoming a well-established component in descriptive freight simulation models (Hunt & Stefan, 2007; Sánchez-Díaz et al., 2015; de Bok et al., 2018). Incorporating tour formation in such models is necessary as vehicle trips and commodity flows do not necessarily have matching ODs (Roathanachonkun et al., 2007; Holguín-Veras et al., 2014).

Freight models can have a trip-based or a commodity-based architecture (see e.g. Holguín-Veras & Thorson, 2000). Until now, tour formation models have been mostly trip-based. An important issue with these models which limits their predictive capability is that assumptions need to made about the outcome of the tour building process, before the actual tours are built. These may concern the number of stops, the average payload, the routes followed, or the number of tours starting from a region. The few shipment-based models that do exist are limited in their empirical validity and have not provided results yet that are generalizable to broader freight markets.

The main focus and contribution of this paper is a shipment-based tour formation model, which is estimated on a large and comprehensive dataset of shipment and freight trip data. The dataset contains the company specific observations (also called microdata) of movements from the national road freight survey produced by Statistics Netherlands. It includes different types of commodities and logistical activities including manufacturing, retail, transshipment and distribution.

The paper is organized as follows. Section 2 provides the literature review of tour formation modeling and defines the knowledge gap that our study aims to contribute to fill. Section 3 describes the tour formation model, while Section 4 introduces the data that allowed to estimate the model. In Section 5 the results of the estimation are presented and interpreted. Section 6 reports on additional validation work and on a sensitivity analysis. Conclusions and recommendations for future work are presented in Section 7.

2. Literature review and research gap

Different descriptive modelling approaches are available in the literature that include tour formation. We distinguish between two lines of work: (1) mathematical optimization based approaches and (2) behavioral choice model based approaches.

Mathematical optimization based approaches apply the Vehicle Routing Problem (VRP), as also used by firms for tactical planning purposes (e.g. Boerkamps & van Binsbergen, 1999; Taniguchi & van der Heijden, 2000; Wisetjindawat et al., 2006; Polimeni et al., 2010; Anand et al., 2014). As VRPs are used to predict tours, one has to assume that the model sufficiently reproduces the decision-maker's behavior and that constraints can be specified adequately. To the best of our knowledge, the validation of this assumption has only been addressed in one study, albeit not in a shipment-based setting. You et al. (2016) apply inverse optimization based on GPS truck diary data of the San Pedro Bay Ports in California, USA. Validation is based on visual comparisons between modelled and observed tours. The authors do not report any quantitative measures of fit and parameters are not calibrated in a way in which statistical significance can be checked. Finally, the approach is computationally too heavy to be applied in a large scale urban freight model.

Choice modelling approaches build on random utility theory and provide a statistical framework for the estimation of behavioral parameters in models, in a way that these replicate real-life choices. Econometric techniques allow to test hypothesized behavioral rules empirically, generalize findings to a population and control for the correlation between predictors. The difficulty of applying choice models for the tour formation activity is that it is not possible to narrow down tour formation to a single choice, which can be easily observed in practice or reconstructed in choice surveys. Therefore, in the literature, different approaches have been proposed which model tour building with approximate choices. Hunt & Stefan (2007) pioneered an approach of stepwise descriptive tour formation modeling with an application to the city of Calgary, Canada. Firstly, in their method, the number of tours originating in each zone is estimated. Secondly, vehicle type and tour purpose are chosen. Thirdly, the tour is built up iteratively by choosing next stop locations until the choice is made to return to the home base. Since number of tours, vehicle types and tour purposes are chosen before the tour formation, the model cannot be classified as shipment-based, however. Raothanachonkun et al. (2007) propose a similar incremental tour building algorithm to convert aggregate commodity flows to vehicle tours. However, their approach lacks an empirical foundation based on firm-level data; tour decisions are modelled deterministically based on average payloads.

Nuzzolo et al. (2012) and Outwater et al. (2013) develop models with behavioral components that extend the above approach based on shipment data. Nuzzolo et al. (2012) propose a model for restocking tours for retail shipments. Nuzzolo & Comi (2014) present an application of this model in Rome, Italy. In their method, tour formation starts by deciding for each shipment the number of trips of the tour that it will be part of. After that, the tours are constructed with a 'next stop location' MNL choice model. Ruan et al. (2012) use commercial vehicle data from Texas, USA to estimate an MNL model for the number of stops and tour pattern, i.e. the number of tours required to deliver all shipments. Outwater et al. (2013) apply this model in a shipment-based context in their framework for Chicago, IL, USA. Geographically close shipments with the same tour pattern and number of stops are grouped into tours using a hierarchical clustering method, after which a nearest neighbor search is used to construct the sequence of locations. Only tours that distribute food and manufactured goods from a central warehouse are modeled in their study. The scope of these applications is limited to retailer replenishment tours and tours that distribute food and manufactured goods from a central warehouse. Additionally, the assumption that the number of stops is chosen before tours are constructed is questionable. In reality the number of stops is an outcome of the process of grouping shipments into tours. Therefore, these models are not strictly shipment-based.

To conclude, to the best of our knowledge, there is no descriptive shipment-based tour formation model, that has been validated for multiple goods types or location types and can thus be applied in a general urban freight model. Our contribution aims to help to fill this gap with a model built on a large dataset of carrier and shipment microdata in the Netherlands. We present the approach in the next section.

3. The tour formation model

The objective is to model the assignment of shipments to tours in a way that is both effective, i.e. reproducing the observed logistics patterns, and - given the context of a large scale urban freight model - efficient, in terms of calculation times. The choice problem is formulated as a sequence of tour-building choices by analogy to the approach of Hunt and Stefan (2007). Adding shipment selection, however, we now re-frame this approach as a shipment-based model. This allows us to take into consideration a number of logistical constraints, such as the size of shipment or vehicle, and the available set of shipments to build tours with. In our model, carriers build tours by repeatedly selecting shipments from a set and adding them to build a tour, until this tour is long enough. The two choices modeled are (1) whether a tour can be completed or not; adding, in the latter case, an additional shipment (the "End Tour" choice) and (2) which shipment to add to the tour from those not yet served ("Select Shipment" choice). Figure 1 shows the flow of the overall tour formation model (notations as listed in Table 1). We discuss the two submodels in more detail below.

Table 1. Notations

c	Carrier index
t	Tour index
i	Shipment index, denotes the iteration of the formation of a tour in which the *allocated shipment* was added to the tour
j	Shipment alternative index, indicates the position of a *yet to allocate shipment* in the choice set to add to a tour
C_i^{ET}	Alternative specific constant for ending the tour in the End Tour (ET) choice model
β_{ri}^{ET}	Estimated parameter for the *r*th attribute in the utility function of the ET choice model
β_r^{SS}	Estimated parameter for the *r*th attribute in the utility function of the SS choice model
n_i^{ET}	The number of attributes in the utility function of the ET choice model
n^{SS}	The number of attributes in the utility function of the Select Shipment (SS) choice model
s_{ctij}	*j*th shipment in the choice set of shipments which can be added to tour *t* by carrier *c* in iteration *i* of forming the tour
S'_{cti}	The shipment that was chosen from the choice set and added to tour *t* by carrier *c* in iteration *i* of forming the tour
U_{cti}^{ET}	Utility of ending tour *t* of carrier *c* in iteration *i* of forming the tour
$U_{s_{ctij}}^{SS}$	Utility of adding shipment s_{ctij} to tour *t* of carrier *c* in iteration *i* of forming the tour
x_{rcti}^{ET}	The value of the *r*th attribute in the utility function of the ET choice model in iteration *i* of forming tour *t* of carrier *c*
$x_{rs_{ctij}}^{SS}$	The value of the *r*th attribute in the utility function of the SS choice model for shipment s_{ctij}
α	The value of the proximity constraint [km]

γ	The number of shipment alternatives in the choice set of the SS choice model

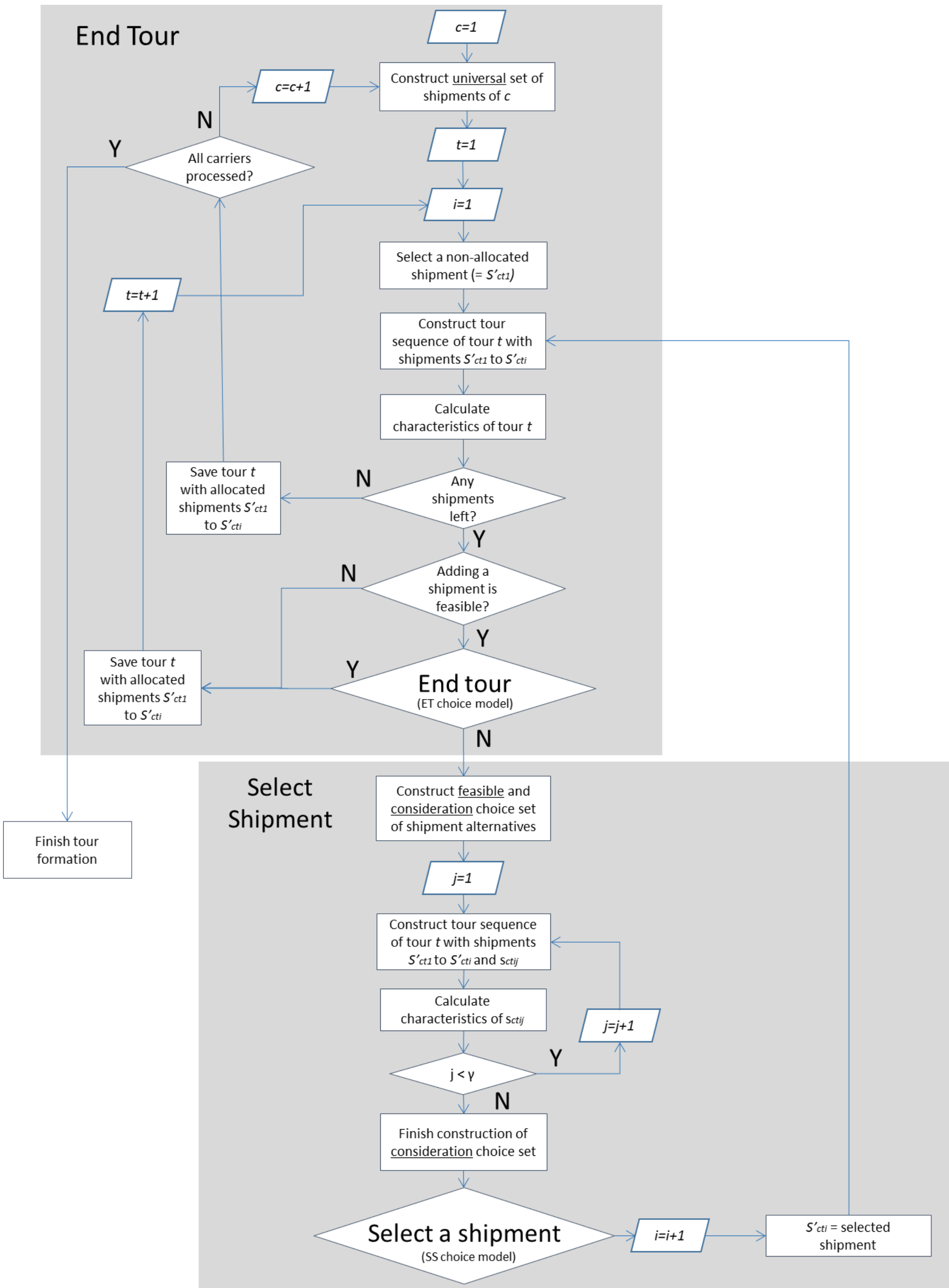

Figure 1. Flow diagram of the proposed tour formation model.

3.1 The End Tour model

The function of the End Tour model is to construct the sequence of locations to visit by adding shipments, until the tour can be completed. We consider shipments transported for a whole day for individual carriers, knowing the portfolio of shipments of each carrier as well as the delivery dates of shipments. The choice model involves a sequence of binary choices for next stops in a tour, considering whether to add a next stop to the tour or not. The ET choice model has a binary dependent variable with the categories '0 = continue adding shipments to tour' and '1 = end tour here'. The utility of ending the tour is calculated as follows:

$$U_{cti}^{ET} = C_i^{ET} + \sum_{r=1}^{n_i^{ET}} (\beta_{ri}^{ET} * x_{rcti}^{ET}) \quad (1)$$

The list of attributes (x_{rcti}^{ET}) in the utility function of the ET choice model includes characteristics of the tour (e.g. tour duration), vehicle (e.g. capacity utilization), visited locations (e.g. zones with transshipment or distribution activities) and goods to move (e.g. goods type). We take a data-driven econometric approach to determine the final list of explanatory variables, based on the rich dataset available about freight trip and shipment characteristics. Therefore, we report and interpret the attributes identified (C_i^{ET} and β_{ri}^{ET}) after having introduced the dataset, in section 5.

To calculate tour duration, we have to re-construct the sequence of visiting the loading and unloading locations of all shipments that have been allocated so far to the tour. A random shipment in the set is selected as the first shipment of a new tour. An alternative approach could be to add a third choice model which determines the starting shipment of choice. This will be the subject of later research. To identify subsequent stops, we use a nearest neighbor search approach - after each location, the nearest remaining location is visited. We developed two alternative search algorithms: the first visits all loading locations before unloading locations are visited, while the second visits alternately loading and unloading locations. Using more advanced algorithms to solve a Traveling Salesman Problem could lead to more efficient sequences (AlSalibi et al., 2013). However, the computational efficiency of the nearest neighbor search is of large importance in the context of a real scale urban freight model. With our nearest neighbor search the proposed tour formation algorithm takes three minutes to form tours for approximately 39,000 shipments, using a PC with an i7 processor and 16.0GB RAM.

When constraints are violated, the tour is ended regardless of the probability calculated with the ET choice model. Three types of logical constraints are specified, which may cause the tour to end: (1) proximity, (2) vehicle capacity, and (3) work shift constraints. Firstly, if there are no non-allocated shipments left within a radius of α km to the tour as constructed so far, then the tour is ended because all non-allocated shipments would require a long additional time. Secondly, because of regulations and physical limitations, the total transported weight may not exceed the vehicle capacity. Thirdly, the tour is ended after nine hours to acknowledge work shift constraints. Finally, as concrete and cement shipments only have direct tours (i.e. tours with one shipment) this commodity causes tours to end immediately.

3.2 The Select Shipment model

If the tour is not ended, the Select Shipment (SS) choice model is used to select which shipment is added to the tour. The SS choice model is a multinomial logit choice model with a choice set of γ shipments as candidates. The utility of selecting shipment s_{ctij} is calculated as follows:

$$U_{s_{ctij}}^{SS} = \sum_{r=1}^{n^{SS}} (\beta_r^{SS} * x_{rs_{ctij}}^{SS}) \tag{2}$$

The attributes ($x_{rs_{ctij}}^{SS}$) in the utility function of the SS choice model are related to the goods type of the shipment and the efficiency with which the shipment can be added to the tour. As with the End Tour model, the choice of parameters was driven by the available data and econometric analysis. We report and interpret the parameters (β_r^{SS}) of the SS choice model after the introduction of the dataset, in Section 5.

For constructing the set of candidate shipments, we distinguish between 3 types of choice sets: the universal choice set (UC), the feasible choice set (FC), and the consideration choice set (CC). This allows us to remove shipments that violate constraints and to ensure a reasonable choice set size for computational efficiency. The UC consists of all shipments of the same carrier and day. The FC is a subset of the UC that respects constraints (such as vehicle capacity), while the CC is a randomly sampled subset of the FC with a fixed number of alternatives γ. To be consistent with the constraints in the ET procedure, we define the following types of constraints that guide the formation of the FC: (1) proximity, (2) commodity type, and (3) vehicle capacity. Shipments are removed from the choice set when they are not located within a radius of α km of the tour locations, when the goods type has no tours (in our application, mostly concrete/cement), and when the shipment causes the total transported weight to exceed the vehicle capacity.

4. The carrier and shipment database

For the development of the model, we use the carrier survey data collected by Statistics Netherlands (CBS). A large amount of data is available, about 2.6 million shipments from 2013 to 2015. Carriers and own-account shippers are legally obliged to report transported shipments if they are part of the CBS sample and can do so digitally, using their Transport Management System (see for more detail e.g. de Bok et al., 2018). The data are listed as separate shipments and include an association between shipments and tours. The definition of a tour is unique compared to definitions found in other studies. In the data, a tour starts at the location where the first shipment is loaded into an empty vehicle, and a tour ends at the location where the vehicle turns empty or at the home base location. Consequently, empty trips are not reported, and when a vehicle turns empty before picking up its next shipment, a new tour record is started.

In addition to shipment data, we use land use data (CBS, 2015), employment data (CBS, 2017), and travel time data (off-peak travel times and distances from the Dutch NRM-West transportation model). Land use data is used to distinguish urban and retail zones, while employment data provides the information to determine which zones have transshipment and goods distribution activities. We distinguish zones at the very detailed postal code level of 'buurten', a Dutch administrative zonal classification with an average zone size of approximately 3.5 km^2. As not all attributes needed for our model are completed in the survey for all records, 515,810 valid shipment records of the approximately 2.6 million remain for our analyses.

Table 2 provides the descriptive statistics of the dataset. We can see that the largest portion of tours is direct (92%), i.e. with only a single destination per tour. There is little literature to compare our numbers with. Khan & Machemehl (2017) find only 34% of direct tours in their dataset of 338 trucks in Central Texas, USA. We expect that this is due to the aforementioned definition of a tour and the large share of bulk concrete/cement shipments in our dataset. Due to large shipment sizes and a high time-sensitivity, multiple-stop tours are often not feasible (Khan & Machemehl, 2017). Additionally, relatively short distances are observed because we only analyze tours within the Netherlands. Table 3 shows for which goods, vehicles, and locations, direct tours are observed most often. The analyses have guided our search for explanatory variables in the model during the estimation. We report these results in the next section.

Table 2. Descriptive tour statistics.

Tour characteristics	Frequency (tours)
Number of stops	
1-2 (direct)	365905 (92.4%)
3-5	18538 (4.7%)
6-10	10008 (2.5%)
>10	1361 (0.3%)
Tour distance bands [km]	
0-20	172341 (43.5%)
20-40	82995 (21.0%)
40-120	62614 (15.8%)
120-200	38019 (9.6%)
≥200	39843 (10.1%)
Concrete/cement	179468 (45.3%)
NSTR goods type	
0: agricultural	9541 (2.4%)
1: food & fodder	20617 (5.2%)
2-5: fuels, oils, metals	746 (0.2%)
6: construction materials	45279 (11.4%)
7: manure/fertilizers	457 (0.1%)
8: chemical products	210151 (53.1%)
9: machinery and other	109021 (27.5%)
Vehicle type	
Truck	194875 (49.4%)
Truck + trailer	37660 (9.5%)
Tractor + trailer	160049 (40,6%)
Other/special vehicle	2134 (0.5%)
Any location visited in tour	
Transshipment	102679 (25.9%)
DC	176249 (44.5%)

Table 3. Direct tour characteristics

Tour characteristics	Percentage of direct tours
Average	92.4%
Average (excl. concrete)	86.2%
Concrete/cement	100.0%
NSTR goods type (excl. concrete)	
0: agricultural	73.1%
1: food & fodder	64.1%
2-5: fuels, oils, metals	96.9%
6: construction materials	97.7%
7: manure/fertilizers	77.0%
8: chemical products	95.2%
9: machinery and other	84.1%
Vehicle type (excl. concrete)	
Truck	72.9%
Truck + trailer	96.4%
Tractor + trailer	85.3%
Other/special vehicle	97.3%
Any location in tour (excl. concrete)	
Transshipment loading	96.4%
Transshipment unloading	96.0%
DC loading	68.5%
DC unloading	70.4%
Urban zone	61.0%
Retail zone	72.0%

Urban zone	146098 (36.9%)
Retail zone	48164 (12.2%)

5. Estimation results

This section presents the estimated ET and SS choice models. We distinguish between three types of explanatory variables: (1) instrumental variables, (2) location variables, and (3) vehicle/goods type variables. Variables were added consecutively to the models and removed when the p-value was higher than 0.05, or when multicollinearity was found. Instrumental variables were added first; these reflect the decision-making process of a transportation planner and are most intuitive. We tested the square root, the natural logarithm, and the square of non-categorical variables in order to investigate non-linear effects. The non-linear specification was chosen if it lead to the highest pseudo-R^2 and if the non-linearity of the effect was clearly explicable. The ET choice model is estimated separately for the first shipment and for later shipments, because it was observed that the majority of the tours ended after the first shipment; different effects can explain the two ET choices.

For the purpose of estimating the ET choice model we assume that a complete tour (i.e. a tour including all its listed shipments) is ended (ET=1) and a sub tour (i.e. a tour with only a subset of its listed shipments) is not ended (ET=0). Four model variations (A to D) (Table 4) are used for model estimation. In Model A to C we vary the choice set size (γ=6 or γ=11) and the rigidity of the proximity constraint (α=100km or α=150km), because these model specifications are more difficult to define intuitively than operational constraints such as vehicle capacity utilization. The observations of fifty percent of the days in the data are used to estimate Model A to C. All carriers provide shipments for the estimation data sets of Model A to C. Model D tests how results differ when data of only 50% of the carriers are used for estimation.

Table 4. Specification of model variations A to D.

Model	Proximity constraint (α)	Choice set size (γ)	Data used for estimation
A	<100km	6	50% of days
B	<100km	11	50% of days
C	<150km	11	50% of days
D	<100km	6	50% of carriers

5.1 End Tour choice model

Tables 5 and 6 present respectively the estimation results of the ET first shipment model and the ET later shipments model. A positive parameter leads to higher utility and probability of ending the tour.

Table 5. Estimation results ET first shipment.

Table 6. Estimation results ET later shipments.

(cells present the estimated Beta and standard error)

ET first shipment	A, B	C	D
$R^2_{Nagelkerke}$	0.442	0.439	0.570
-2 LL	47315	55186	55866
Percentage correct	84.8	85.3	87.8
N	75255	90000	99273
$C_{i=1}^{ET}$ ***Constant***	1.684 (0.029)	1.473 (0.027)	1.681 (0.024)
$\beta_{r=1,i=1}^{ET}$	-1.698 (0.037)	-1.112 (0.030)	-2.403 (0.034)
$\beta_{r=2,i=1}^{ET}$ ***(W/C)²***	5.471 (0.102)	6.022 (0.098)	5.258 (0.088)
$\beta_{r=3,i=1}^{ET}$ ***anyTS***	1.588 (0.037)	1.484 (0.035)	2.354 (0.037)
$\beta_{r=4,i=1}^{ET}$ ***anyDCload***	-0.578 (0.026	-0.517 (0.024)	-0.942 (0.025)
$\beta_{r=5,i=1}^{ET}$ ***anyDCunload***	-0.475 (0.026)	-0.450 (0.024)	-0.765 (0.025)
$\beta_{r=6,i=1}^{ET}$ ***anyURB***	-0.461 (0.038)	-0.605 (0.036)	-0.499 (0.037)
$\beta_{r=7,i=1}^{ET}$ ***vehicle type*** *[0: truck]*	-1.295 (0.039)	-1.370 (0.037)	-1.684 (0.039)
$\beta_{r=8,i=1}^{ET}$ *[1: truck + trailer]*	1.850 (0.049)	1.980 (0.045)	2.508 (0.047)
[2: tractor + trailer]	-	-	-
[3: other/special]	-	-	-
$\beta_{r=9,i=1}^{ET}$ ***goods type*** *[0: agricultural]*	-0.736 (0.047)	-0.881 (0.044)	-0.271 (0.037)
$\beta_{r=10,i=1}^{ET}$ *[1: food & fodder]*	-0.659 (0.032)	-0.808 (0.029)	-0.672 (0.037)
$\beta_{r=11,i=1}^{ET}$ *[2-5: fuels, oils, metals]*	1.495 (0.337)	1.298 (0.324)	1.121 (0.311)
$\beta_{r=12,i=1}^{ET}$ *[6: construction materials]*	1.452 (0.058)	1.472 (0.051)	2.253 (0.048)

ET later shipments	A, B	C	D
$R^2_{Nagelkerke}$	0.292	0.293	0.186
-2 LL	37894	39933	62022
Percentage correct	81.8	81.6	75.1
N	44618	46336	59869
$C_{i>1}^{ET}$ ***Constant***	-2.526 (0.062)	-2.547 (0.060)	-2.516 (0.054)
$\beta_{r=1,i>1}^{ET}$ ***TD***	0.386 (0.014)	0.364 (0.014)	0.449 (0.012)
$\beta_{r=2,i>1}^{ET}$ ***W/C***	3.286 (0.057)	3.285 (0.055)	3.122 (0.048)
$\beta_{r=3,i>1}^{ET}$ ***prox***	0.009 (0.001)	0.008 (0.000)	0.008 (0.000)
$\beta_{r=4,i>1}^{ET}$ ***lnstops***	-0.911 (0.042)	-0.841 (0.041)	-0.828 (0.036)
$\beta_{r=5,i>1}^{ET}$ ***anyTS***	0.526 (0.047)	0.545 (0.046)	0.450 (0.040)
$\beta_{r=6,i>1}^{ET}$ ***anyDCload***	-0.191 (0.036)	-0.179 (0.035)	-0.281 (0.031)
$\beta_{r=7,i>1}^{ET}$ ***anyDCunload***	0.094 (0.036)	0.078 (0.035)	-0.150 (0.031)
$\beta_{r=8,i>1}^{ET}$ ***anyURB***	-0.145 (0.032)	-0.175 (0.032)	-0.036 (0.027)
$\beta_{r=9,i>1}^{ET}$ ***vehicle type*** *[0: truck]*	-1.968 (0.061)	-1.968 (0.058)	-2.354 (0.059)
$\beta_{r=10,i>1}^{ET}$ *[1: truck + trailer]*	-0.954 (0.088)	-1.003 (0.086)	-0.845 (0.085)
[2: tractor + trailer]	-	-	-
[3: other/special]	-	-	-
$\beta_{r=11,i>1}^{ET}$ ***goods type*** *[0: agricultural]*	2.226 (0.059)	2.203 (0.056)	2.182 (0.045)
$\beta_{r=12,i>1}^{ET}$ *[1: food & fodder]*	0.871 (0.035)	0.873 (0.033)	0.546 (0.031)

$\beta^{ET}_{r=13,i=1}$ *[7: manure/fertilizers]*	0.713 (0.253)	-	0.878 (0.237)
$\beta^{ET}_{r=14,i=1}$ *[8: chemical products]*	0.583 (0.045)	0.530 (0.042)	1.821 (0.053)
[9: machinery and other]	-	-	-

[2-5: fuels, oils, metals]	-	-	-
$\beta^{ET}_{r=13,i>1}$ *[6: construction materials]*	0.556 (0.081)	0.538 (0.078)	0.396 (0.068)
$\beta^{ET}_{r=14,i>1}$ *[7: manure/fertilizers]*	-1.105 (0.327)	-1.702 (0.289)	-0.888 (0.244)
$\beta^{ET}_{r=15,i>1}$ *[8: chemical products]*	1.517 (0.063)	1.468 (0.060)	1.168 (0.062)
[9: machinery and other]	-	-	-

If the first shipment of a tour requires a longer tour duration (***TD***) in hours from loading to unloading, the probability of ending the tour is lower (Table 5). The square root ($\sqrt{\boldsymbol{TD}}$) indicates a stronger effect for lower tour durations; the attractiveness of a direct tour does not decrease as strongly for longer tour durations. A direct tour is more likely to be chosen for a shipment within short reach. Nuzzolo et al. (2012) found similar effects and reasoned that carriers prefer constructing direct tours to reduce the complexity of planning. Additionally, the travel time savings of grouping shipments might be smaller for these nearby shipments.

The capacity utilization (***W/C***) is calculated as the ratio between the total transported weight of the tour and the carrying capacity of the truck. The probability of ending the tour increases with a larger share of the vehicle capacity used. This reflects the strategy of transportation planners to fill vehicles optimally to save transportation costs. The quadratic component ($(\boldsymbol{W/C})^2$) implies a stronger effect for higher utilization rates; the transportation planner prefers not to end the tour until the capacity is nearly reached. As capacity utilization could only be obtained with respect to weight, many other parameters are expected to reflect differences in volume.

AnyTS is a binary variable stating whether a transshipment zone is visited in the tour. Tours that visit a transshipment zone (***anyTS*** = 1), e.g. a port, are more likely to be ended after the first shipment. The transported shipment is likely to be a producer flow as part of an international logistics chain. These shipments tend to have larger volumes (Friedrich et al., 2014). Consequently, it is usually not feasible to transport multiple shipments in a single tour.

AnyDC**$_{load}$* and ***anyDC**$_{unload}$* are binary variables stating whether any of the locations visited for loading/unloading goods is a distribution center. When a distribution center is visited (anyDC**$_{load}$* or ***anyDC**$_{unload}$* = 1), the probability of ending the tour decreases. The transported shipments are more likely to be transported to a place of consumption and to have a smaller volume (Friedrich et al., 2014). In addition, distribution centers organize their (un)loading activities in such a way that more customer visits can be made (Khan & Machemehl, 2017) and tend to use larger vehicles (van Duin et al., 2012). The effect is stronger when shipments are loaded at a distribution center (***anyDC**$_{load}$* = 1) than when they are unloaded (***anyDC**$_{unload}$* = 1). Shipments unloaded at a distribution center correspond more often to flows originating from a producer.

The probability of ending the tour after the first shipment is lower when an urbanized zone is visited (***anyURB*** = 1). The demand is more concentrated in cities, efficient tours serving multiple customers might be possible more often. Especially if the driver has to enter a large city from a rural location it saves a lot of time to reduce the number of trips in and out of the city.

The variables ***vehicle type [0-1]*** and ***goods type [0-8]*** are binary variables stating the vehicle type used for the tour and the NSTR category of the transported goods. Differences between vehicle types can be explained through differences in volumes and ease of (un)loading. Truck + trailers are less practical for transportation of shipments to multiple customers, as the trailer needs to be uncoupled to unload goods from the truck. Differences in goods types can be related to differences in volume, ease of (un)loading, stricter restrictions in combination with other goods, and dispersion of supply/demand. Restaurants with a demand for food products (***goods type [1]***) might be concentrated in a food district, while gas stations (***goods type [2-5]***) might be more dispersed. The estimated parameters for vehicle, goods, and location types in the ET first shipment model show effects similar to the descriptive statistics in Table 3; the categories with a positive parameter have a higher percentage of direct tours in Table 3.

Most effects are similar between the ET first shipment model and the ET later shipments model, but three key differences are found: (1) the sign of tour duration (***TD***) switches from negative to positive, and (2) ***prox*** and (3) ***lnstops*** are not included in the ET first shipment model.

The probability of ending the tour increases with a higher tour duration (***TD***) in the ET later shipments model. Tours with multiple shipments are more likely to cover a full working shift than tours with one shipment. The transportation planner prefers not to construct tours that last close to a maximum work shift duration. If the tour lasts longer than expected due to congestion, then customers might experience a delay of a day or the driver must work overtime.

Prox is the distance [km] of the nearest non-allocated shipment to the tour. If the nearest non-allocated shipment is closer to the tour as constructed so far (lower value of ***prox***), then the probability of ending the tour is lower. If there are shipments that can be added with little additional time, then the transportation planner prefers to add more shipments to the tour. The variable ***lnstops*** is the natural logarithm of the number of stops in the tour as constructed so far. The parameter shows a negative sign: when the tour has more stops, the probability of ending the tour is lower. An additional shipment is not as unattractive when the tour visits many stops, the tour is already long and complex. The natural logarithm indicates a stronger effect for lower values; tours with fourteen or fifteen stops are considered more similar than tours with three or four stops.

Models A and B are identical in the End Tour process, the choice set size (γ) only impacts the shipment selection, it does not influence the choice to end the tour. A more lenient proximity constraint (α=150km) has a minor impact on the estimated parameters. Model D, estimated on a subset of the carriers, leads to larger differences with Model A to C. The only sign that changes direction with different model specifications is that of the $\boldsymbol{anyDC_{load}}$ parameter in the ET later shipments model; however, in accordance to the parameters for Model A to C, the $\boldsymbol{anyDC_{unload}}$ parameter is still lower than that of $\boldsymbol{anyDC_{load}}$ in Model D.

The high Nagelkerke pseudo-R^2 values of the ET first shipment model (0.570 for Model D) indicate a good model fit. For the ET later shipments model the Nagelkerke pseudo-R^2 values are relatively low (0.186 for Model D). The data used for estimation of the ET later shipments model covers a broader range of choices, which makes it more difficult to fit the data. The ET later shipments model predicts whether a tour ended after the second shipment but also after each consecutive shipment, while the ET first shipment model only predicts whether a tour ended after the first shipment.

5.2 Select Shipment choice model

The choice sets for estimating the SS choice model are generated by sampling a shipment that is part of the same tour as the observed chosen shipment and sampling γ-1 shipments of other tours by the same carrier on the same date as the unchosen shipments. Table 7 presents the estimation results of the SS model. A positive parameter increases the probability that an alternative (i.e. s_{ctij}, a non-allocated shipment) is selected as the additional shipment to a tour. All three variables can be considered instrumental, they reflect the decision-making process of the transportation planner.

Table 7. Estimation results Select Shipment choice model. Cells present the Beta and standard error.

	Specification	A	B	C	D
	$R^2_{McFadden}$	0.187	0.169	0.249	0.156
	LL	-63256	-73929	-73834	-101620
	N	43409	37112	41001	67181
$\beta_{r=1}^{SS}$	***addcost***	-0.005 (0.000)	-0.005 (0.010)	-0.010 (0.000)	-0.006 (0.000)
$\beta_{r=2}^{SS}$	***addstops***	-1.039 (0.010)	-1.088 (0.010)	-1.176 (0.010)	-0.918 (0.008)
$\beta_{r=3}^{SS}$	***sameNSTR***	2.313 (0.038)	2.712 (0.042)	2.627 (0.041)	2.176 (0.031)

The additional generalized cost (***addcost***) is a weighted sum of the travel time (€45.12/h) and the distance (€0.45/km) a shipment adds to a tour. These weights have been used in the Dutch national freight model BasGoed and reflect the costs (e.g. labor and fuel) that carriers spend for each driven hour or kilometer

(Significance, 2018). A shipment with a higher additional cost has a lower probability of being selected, as carriers wish to minimize transportation costs by constructing efficient tours.

As each shipment requires only two stops, one for loading and one for unloading, the additional number of stops (***addstops***) of a shipment can be zero, one, or two. A shipment that adds more stops to the tour (i.e. a shipment with fewer stops in common with the tour) has a lower probability of being selected. Shipments that have more stops in common with the tour add less complexity to the tour and might require less additional dwelling time (e.g. parking, (un)loading).

SameNSTR is a binary variable stating whether a shipment alternative has the same NSTR goods type as the NSTR goods type of which the highest weight is transported in the constructed tour. In 93% of the cases in which multiple shipments are transported in a tour, we observe that all shipments have the same NSTR goods classification. Consequently, in the SS choice model the probability of selecting a shipment is higher if it has the same goods type as the other shipments in the tour (***sameNSTR*** = 1). This can be explained with restricted goods combinations.

Estimation results are relatively stable for Model A to D. The McFadden pseudo-R^2 of Model C is higher and the ***addcost*** parameter of Model C is twice as low compared to Models A and B. In Model C, α is increased from 100km to 150km. Consequently, the choice set includes more distant, less attractive, shipments. Correctly predicting the observed choice is easier in such a choice set, which improves the McFadden pseudo-R^2. Distant shipments have a higher additional cost, these higher values influence the ***addcost*** parameter.

6. Validation and sensitivity analysis

The estimation of the ET and SS choice models in itself does not provide sufficient information to judge the performance of the tour formation algorithm. Other aspects, such as assumed constraints and choice set formation approach, influence how tours are constructed. For this reason, we test the model's performance in two ways: by constructing tours with the shipments in an out-of-sample validation data set (i.e. 50% of the data, which we do not use for estimation), and by testing the sensitivity of the model outcomes to variations in travel times.

6.1 Validation

The model performance is assessed by comparing the observed tours in the validation set with a prediction of tours by our model. For this purpose, we calculate the coincidence ratio between the observed and predicted frequency distribution of tours by number of stops and by tour distance. A coincidence ratio higher than 80% is generally considered good in validating zonal freight trip distance distributions (National Cooperative Highway Research Program, 2008). As the coincidence ratio is above 80% for both the number of stops and tour distance (Table 8), we conclude that our model reproduces aggregate tour statistics well for a given set of shipments. In addition, the distribution of the number of stops is reproduced sufficiently for different location and goods types (Table 9). As expected, the model shows that tours that visit a distribution center tend to have more stops. Concrete/cement shipments, for which we only construct direct tours, are listed as NSTR8 in the data, which is why the coincidence ratio is very high in this category. However, also for other goods categories we find high coincidence ratios, indicating a high explanatory power of the End Tour choice model. For unknown reasons, though, too many direct tours are predicted for foodstuffs (*NSTR1*).

Table 8. Coincidence ratio between observed and predicted distributions of number of stops and distance

Model	Coincidence ratio: number of stops (averaged over three models runs for A-C and two model runs for D)	
	Number of stops	Tour distance
A	98.8%	89.3%
B	99.0%	89.4%
C	98.6%	89.5%
D	96.9%	84.2%

Table 9. Coincidence ratio between observed and predicted distributions of number of stops by location and goods type.

Model	Coincidence ratio: number of stops *(averaged over three models runs for A-C and two model runs for D)*								
	DC visited	no DC visited	NSTR0	NSTR1	NSTR2-5	NSTR6	NSTR7	NSTR8	NSTR9
A	99.1%	96.6%	92.7%	69.6%	95.6%	96.4%	77.9%	99.5%	92.5%
B	99.0%	97.0%	93.7%	68.8%	95.2%	96.6%	78.4%	99.6%	93.1%
C	98.8%	97.3%	91.5%	70.6%	95.5%	98.0%	80.6%	99.5%	95.3%
D	98.6%	90.4%	95.8%	85.1%	94.4%	94.9%	89.0%	96.2%	90.8%

The differences between the coincidence ratios of Models A to C are negligibly small. Consequently, we can conclude that the model performance is robust to differences in the choice set size γ (A to B) and the rigidity of the proximity constraint α (B to C). Model D shows lower coincidence ratios overall when compared to Models A to C, indicating a worse performance. Model D was estimated with less diverse information (only a subset of carriers instead of a subset of days), and applied to a more dissimilar validation data set (data of other carriers instead of other days). However, the coincidence ratios of Model D are still highly satisfactory. This indicates that model parameters estimated for one set of carriers are applicable to another set of carriers. Because the data shows a strong self-selection of large third-party carriers, the estimated model is considered not representative for own-account carriers.

While observed distributions are reproduced well, models A to C slightly overestimate the percentage of tours with three or four stops and underestimate the percentage of tours with six or seven stops (Table 10). This is caused by the fact that the ET later shipments model is estimated on all observations with multiple shipments. A separate model for each iteration (i.e. ET third shipment, ET fourth shipment) is expected to lead to better results. Additionally, too many tours with more than fifteen stops are predicted; the process of adding shipments can linger on too long in our probabilistic iterative approach.

Models A to D overestimate the percentage of tours with a short distance (Table 11). We expect this to be caused by measurement differences between observed and predicted tour distances. The companies fill out observed tour distances in the survey while the predicted distances are calculated with our tour sequence algorithm and off-peak skim matrices. Consequently, the observed tour distances may include kilometers driven to refuel, to have lunch, or to evade a congested AM or PM peak highway; kilometers that our predicted tour distance does not include.

Table 10. Observed and predicted distribution of number of stops.

Number of stops	Percentage of tours (*averaged over three models runs for A-C and two model runs for D*)					
	50% of *days* for estimation				50% of *carriers* for estimation	
	Observed	Predicted (A)	Predicted (B)	Predicted (C)	Observed	Predicted (D)
1-2 (direct)	92.5%	92.5%	92.6%	93.0%	90.8%	89.2%
3	2.0%	2.2%	2.1%	1.9%	3.3%	3.5%
4	1.5%	1.8%	1.7%	1.6%	2.2%	2.8%
5	1.2%	1.2%	1.2%	1.1%	1.4%	1.5%
6	1.1%	0.8%	0.8%	0.8%	0.8%	0.9%
7	0.7%	0.5%	0.6%	0.5%	0.5%	0.5%
8	0.3%	0.3%	0.3%	0.3%	0.3%	0.4%
9	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%
10	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%
11	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
12	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
13	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
≥15	0.1%	0.1%	0.1%	0.1%	0.0%	0.4%

Table 11. The observed and predicted distribution of tour distance.

Tour distance [km]	Percentage of tours (*averaged over three models runs for A-C and two model runs for D*)					
	50% of *days* for estimation				50% of *carriers* for estimation	
	Observed	Predicted (A)	Predicted (B)	Predicted (C)	Observed	Predicted (D)
<50	67.8%	72.5%	72.4%	71.9%	49.5%	58.1%
50-100	9.1%	10.1%	10.1%	10.5%	19.8%	16.5%
100-150	8.0%	7.6%	7.7%	7.9%	12.6%	11.4%
150-200	5.1%	4.3%	4.4%	4.5%	7.7%	7.2%
200-250	3.1%	2.6%	2.5%	2.5%	3.8%	2.9%
250-300	2.2%	1.1%	1.1%	1.0%	2.1%	1.4%
300-350	1.6%	0.6%	0.6%	0.6%	1.5%	0.8%
350-400	1.1%	0.4%	0.4%	0.4%	1.2%	0.5%
400-450	0.7%	0.3%	0.3%	0.2%	0.7%	0.3%
450-500	0.4%	0.2%	0.2%	0.1%	0.4%	0.2%
500-550	0.3%	0.1%	0.1%	0.1%	0.2%	0.2%
550-600	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%
600-650	0.1%	0.1%	0.1%	0.0%	0.1%	0.1%
650-700	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
700-750	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
750-800	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
800-850	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
850-900	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
900-950	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
950-1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
≥1000	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%

We should note that the observed and predicted tours are constructed with the same set of observed shipments. This explains at least partially why observed tour statistics are reproduced well. Solid statements about the extent to which this tour formation model can improve traffic forecasts can be made only when the model is applied to a synthesized set of shipments and when assigned vehicle trips are compared with traffic counts.

6.2 Sensitivity analysis

To further validate our model and understand its behavior, we analyze its sensitivity to travel time changes. Four simple scenarios are defined in which all OD pairs experience the same increase or decrease in travel time. When travel times in the network increase, fewer direct tours (Figure 2) and fewer tours with 15+ stops are predicted (Figure 3). Longer travel times lead to higher transportation costs; therefore, carriers have a stronger focus on travel time savings, which may be achieved by combining multiple shipments efficiently more often. In addition, a tour with the same set of shipments requires a longer travel time in this scenario; regulated maximum driver shifts are reached with fewer shipments, which limits the construction of tours with many shipments. Both impacts are interpretable and plausible, and are found repeatedly over model runs.

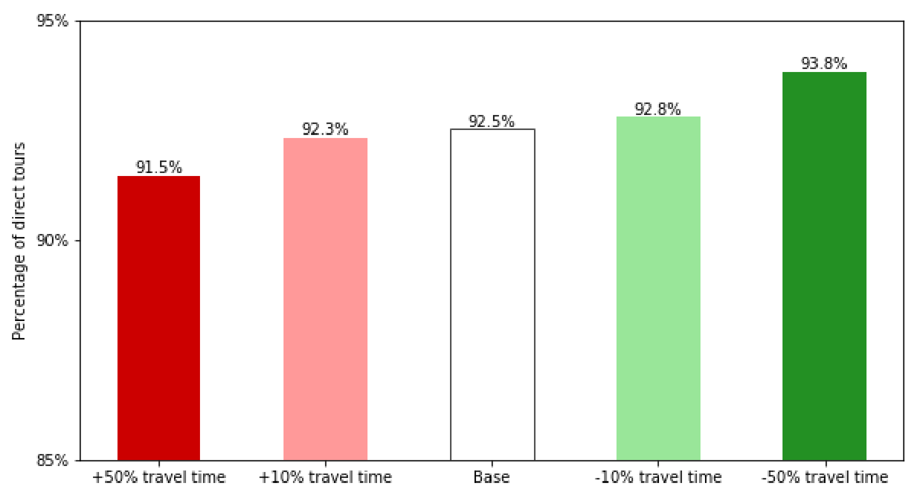

Figure 2. The percentage of direct tours under varying travel time scenarios. The results are averaged over two runs with Model A.

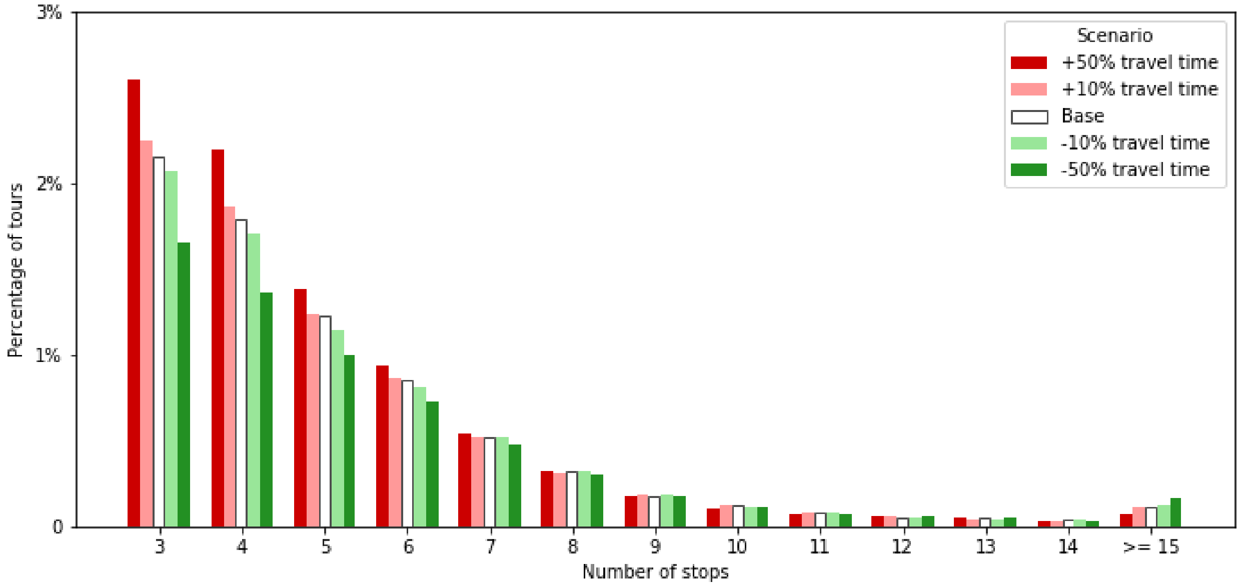

Figure 3. The percentage of tours with multiple stops under varying travel time scenarios. The results are averaged over two runs with Model A.

7. Conclusions

In this research, we developed a descriptive tour formation model in which tours are grown iteratively by allocating one additional shipment until the choice is made to end the tour. Typical for the approach is that it is shipment-based and the parameters of the choice models are estimated using a large and comprehensive database that initially covers over two million shipments for all goods types transported by third-party road freight carriers in the Netherlands. The tour patterns constructed with the model take realistic considerations into account, for instance transportation cost minimization and constraints related to vehicle capacity or working shift regulations. The estimations also indicate a preference for the formation of multiple stop tours when distribution centers and urbanized areas are visited.

An out-of-sample validation study showed a close reproduction of observed statistics regarding tour distance and number of stops, with coincidence ratios exceeding 90%. Both the model estimates and performance are robust for varying choice set sizes and shipment selection rules. Consequently, we conclude that this model can also be applied in a shipment-based freight simulation framework.

Several features that might be added or improved about the model include the following. Firstly, a model that predicts empty trips is of large importance. While empty trips constitute a large portion of all freight trips (Sánchez-Díaz et al., 2015), these empty trips are not reported in the data and, therefore, we do not model them. A possible strategy forward could be to infer the empty trips from the current data, using an empty trip production model. Additional data is required for validation. Secondly, a departure time choice model would allow us to consider that traffic flows and travel times vary throughout the day. As routing and scheduling decisions are often made together because of variations in required delivery times, we expect that the combined treatment will improve the predictive capabilities of the tour formation model.

Acknowledgements

The work reported in this article follows from a collaboration project with Statistics Netherlands. The authors are grateful for access to their transport statistics. Any interpretation or opinion about the data expressed in this article is solely those of the authors.

Bibliography

AlSalibi, B., Jelodar, M., & Venkat, I. (2013). A comparative study between the nearest neighbor and genetic algorithms: A revisit to the traveling salesman problem. *International Journal of Computer Science and Electronics Engineering 1(1)*, 34-38.

Anand, N., van Duin, J., & Tavasszy, L. (2014). Ontology based multi-agent system for urban freight transportation. *International Journal of Urban Sciences 18(2)*, 133-153.

Boerkamps, J., & van Binsbergen, A. (1999). GoodTrip - A new approach for modelling and evaluation of urban goods distribution. *International Conference on City Logistics 1st*, (pp. 175-186). Cairns, Australia.

CBS. (2015). Kerncijfers Wijken en Buurten 2015.

CBS. (2017). Algemeen Bedrijven Register.

de Bok, M., Tavasszy, L., Bal, I., & Thoen, S. (2018). The incremental development path of an empirical agent-based simulation system for urban goods transport (MASS-GT). *World Conference on Transport Research.* Mumbai, India.

Friedrich, H., Tavasszy, L., & Davydenko, I. (2014). Distribution Structures. In L. Tavasszy, & G. de Jong, *Modeling Freight Transport* (pp. 65-87). London, UK: Elsevier.

Holguín-Veras, J., & Thorson, E. (2000). Trip length distributions in commodity-based and trip-based freight demand modeling: investigation of relationships. *Transportation Research Record 1707(1)*, 37-48.

Holguín-Veras, J., González-Calderón, C., Sánchez-Díaz, I., Jaller, M., & Campbell, S. (2014). Vehicle-Trip Estimation Models. In L. Tavasszy, & G. de Jong, *Modeling Freight Transport* (pp. 143-162). London, UK: Elsevier.

Hunt, J., & Stefan, K. (2007). Tour-based microsimulation of urban commercial movements. *Transportation Research Part B 41*, 981–1013.

Khan, M., & Machemehl, R. (2017). Analyzing tour chaining patterns of urban commercial vehicles. *Transportation Research Part A 102*, 84-97.

National Cooperative Highway Research Program. (2008). *Synthesis 384: Forecasting Metropolitan Commercial and Freight Travel.* Washington, DC, USA: Transportation Research Board.

Nuzzolo, A., & Comi, A. (2014). Urban freight demand forecasting: A mixed quantity/delivery/vehicle-based model. *Transportation Research Part E 65*, 84-98.

Nuzzolo, A., Crisalli, U., & Comi, A. (2012). A system of models for the simulation of urban freight restocking tours. *Procedia - Social and Behavioral Sciences 39*, 664 – 676.

Outwater, M., Smith, C., Wies, K., Yoder, S., Sana, B., & Chen, J. (2013). Tour based and supply chain modeling for freight: integrated model demonstration in Chicago. *Transportation Letters 5(2)*, 55-66.

Polimeni, A., Russo, F., & Vitetta, A. (2010). Demand and routing models for urban goods movement simulation. *European Transport 46*, 3-23.

Raothanachonkun, P., Sano, K., Wisetjindawat, W., & Matsumoto, S. (2007). Truck trips origin destination using commodity based model combined with an empty trip model. *Transportation Research Board 86th Annual Meeting.* Washington DC, USA.

Ruan, M., Lin, J., & Kawamura, K. (2012). Modeling urban commercial vehicle daily tour chaining. *Transportation Research Part E 48*, 1169-1184.

Sánchez-Díaz, I., Holguín-Veras, J., & Ban, X. (2015). A time-dependent freight tour synthesis model. *Transportation Research Part B 78*, 144–168.

Significance. (2018). *Schattingsrapport Basgoed 2018 (Werkversie 23 april 2018).* Den Haag, the Netherlands: Significance.

Taniguchi, E., & van der Heijden, R. (2000). An evaluation methodology of city logistics. *Transport Reviews 20(1)*, 65-90.

van Duin, J., van Kolck, A., Anand, N., Tavasszy, L., & Taniguchi, E. (2012). Towards an agent-based modelling approach for the evaluation of dynamic usage of urban distribution centres. *Procedia - Social and Behavioral Sciences 39*, 333-348.

Wisetjindawat, W., Sano, K., Matsumoto, S., & Raothanachonkun, P. (2006). Micro-simulation model for modeling freight agents interactions in urban freight movement. *Transportation Research Board 86th Annual Meeting.* Washington, DC, USA.

You, S. I., Chow, J. Y., & Ritchie, S. G. (2016). Inverse vehicle routing for activity-based urban freight forecast modeling and city logistics. *Transportmetrica A: Transport Science 12(7)*, 650-673.