A behavioral shipment-based model of freight tour formation

Master Thesis

Transport, Infrastructure & Logistics

Delft University of Technology

October 1st, 2018

S. Thoen

A behavioral shipment-based model of freight tour formation

Calibration of a statistical algorithm using big data on Dutch road carriers

MSc Thesis Report

Author:	Sebastiaan Thoen
Student number:	4304942
University:	TU Delft
MSc program:	Transport, Infrastructure & Logistics
Graduation presentation:	October 15th, 2018 – 15:30

Thesis committee:	Prof.dr.ir. L.A. Tavasszy (Chairman)	TU Delft (ESS, T&P)
	Dr.ir. G. Homem de Almeida Correia	TU Delft (T&P)
	Dr. J.H.R. van Duin	TU Delft (T&L)
	M.A. de Bok	TU Delft (T&P), Significance

CONTENTS

PREFACE

This report is the final product of the research performed for my MSc Thesis of the TU Delft program Transport, Infrastructure & Logistics. I have developed and estimated a behavioral model that allocates shipments to tours. For this purpose, I had access to a large data set that the Dutch Central Bureau for Statistics (CBS) collects automatically from the planning systems of Dutch freight carriers. The model shows promising results and can be applied as a module in a larger freight simulation framework to predict truck flows on the Dutch road network. I hope my research reaches an audience, so it can contribute to better freight models and to well underpinned policies that make this world a better place.

I started out this project with an interest in modeling passenger transportation and spatial effects. Already as a child, I loved looking at and drawing maps. Predicting and visualizing traffic flows on these maps in the course CIE4801 reinvigorated that child-like excitement. When a research proposal about modeling freight tour formation came across, it seemed like an interesting challenge with a different perspective. Upfront, I knew very little about freight transportation, but I was right; looking back at my research I can confirm that it has been highly interesting and a big challenge. I have learned so much about freight transportation, but also about data analysis, writing code, structuring ideas, scientific writing, and about myself.

I would like to thank several people for their help in elevating this project and getting more out of myself. Thank you, Lóri and Michiel, for giving me the opportunity to work on this project with this unique data set. And the CBS, for their quick and apt answers to questions regarding the data and Remote Access environment. Thank you, Harrie Tissen, for taking the time of your day to answer my questions about the work of a transportation planner. Thank you, to the people at Significance, for allowing me to conduct my research at your office and for being an inspiring environment to me. And Ivar, I am glad to have had a co-student on the same boat as me, to be able to share our process of trying to get things out of this data has had a positive impact on me. Thank you to Lóri, Goncalo, Ron, and Michiel for the large amount of constructive feedback. Ron, our conversations stimulated my enthusiasm about logistics, and facilitation of an interview with a transportation planner added a useful facet to this research. Goncalo, you have helped me a lot in structuring my method and storyline, and in being critical towards myself and others. And Michiel, our weekly discussion meetings have not only been of great use to the content of my research, but also helped me preserve my confidence and fun in this project.

Finally, the support of my parents has been of immeasurable importance to me during this project, I cannot thank you enough. Thank you to my sister, Rozemarijn, for taking the time to make the beautiful picture on the front cover. Scarlett, I'd like to thank you for your support and your assistance in further improving my English. And a final thank you to Phoebe, Yvette, and all other people who have been there for me in some way during this project.

Sebastiaan Thoen

Den Haag, October 1st, 2018

SUMMARY

Freight transportation is a crucial pillar of our economy, yet freight trucks place disproportionally large burdens on society compared to passenger cars, especially relating to air pollution, traffic safety, congestion, and pavement wear (Hunt & Stefan, 2007; Quak, 2008; Kim et al., 2014; Sánchez-Díaz et al., 2015). As both national and international freight flows continue to grow (Tavasszy, 2008; ITF-OECD, 2015; CBS, 2017a), policies are developed that stimulate more sustainable and efficient transportation of goods. Simulation models are a common tool for development and evaluation of these policies (de Bok & Tavasszy, 2018).

Freight modelers increasingly realize the importance of tour formation for accurate prediction of traffic flows with these simulation models. Tour formation is the construction of a vehicle journey to load and unload shipments. As multiple shipments can be transported in a tour, simulation models that omit tour formation fail to recognize that a truck does not need to drive directly from the loading location to the unloading location of a shipment (Holguín-Veras et al., 2014; Sánchez-Díaz et al., 2015).

Many tour formation models are not shipment-based (e.g. Hunt & Stefan, 2007), lack parameters that are statistically calibrated on empirical data (e.g. Wisetjindawat et al., 2006), or focus on a narrow segment of freight transportation (e.g. Nuzzolo et al., 2012). Modeling shipments allows us to consider that shipments determine the possibilities and constraints for tour formation, statistical calibration provides the empirical foundation to test hypothesized behavioral effects, and having a large and inclusive scope of freight transportation allows us to develop a tour formation model that is applicable to a large variety of shipments in a regional (or national) freight simulation framework.

The objective of this research is to develop a tour formation model that can allocate shipments to tours in a way that is similar to observed tour formation behavior. For this purpose, we have access to a data set, the XML microdata, that lists approximately 2.6 million shipments transported over the Dutch roads. This unique data is collected by the Dutch Central Bureau of Statistics (CBS) and tapped automatically from the transport management systems of Dutch freight carriers. This data is highly inclusive, it makes no selections regarding goods types or regions within the Netherlands, although a self-selection of large third-party carriers (i.e. transporting goods for other parties) can be identified.

Analysis of tour statistics has revealed many insights that help us understand freight tour formation. For short-distance shipments, more often direct tours (one loading point, one unloading point) are found, indicating a preference for simple tours when the efficiency gains of grouping shipments are smaller. Tours that load shipments at a distribution center make more stops. These shipments are likely to be transported to consumers instead of producers and have smaller sizes (Friedrich et al., 2014), which allows more shipments to be transported in the same vehicle. Tours that visit a port transshipment node make fewer stops, for its shipments are more likely to originate from producers and have larger sizes. Due to differences in shipment size, dispersion of demand, goods type restrictions, and ease of loading/unloading, we observe more stops in tours transporting agricultural products, foodstuffs and manure, and few stops in tours transporting oils, metals, construction materials, and chemical products. In addition, cement/concrete shipments are virtually never transported in tours with multiple stops due to their high time-sensitivity and volumes (Khan & Machemehl, 2017).

To be able to calibrate a tour formation model with the data, we have developed an iterative shipment allocation algorithm. A tour is grown iteratively through allocation of an additional shipment until the decision to end the tour is made. After each shipment allocation, the sequence of visiting loading and unloading locations is reconstructed. This process is repeated until the shipments of all carriers have been allocated to a tour. Two probabilistic choice steps are present in this algorithm, which we call the End Tour (ET) and the Select Shipment (SS) choice models. The utility functions of these two choice models are estimated on the data. We do not assume that the two choice models represent an actual choice, though. They should be seen as

statistical models that allow us to use empirical data to consider behavioral effects in an algorithm that simplifies the elaborate and complex process of allocating shipments to tours.

The ET choice model is a Binary Logistic Regression. The binary dependent variable can have a value of 0, which means that the choice is made to allocate an additional shipment to the current tour, and a value of 1, which means that the current tour is ended and a new one is started. Its utility function is estimated separately for the first shipment and for later shipments in a tour, because the majority of tours transports only one shipment. Its explanatory variables are the following: tour duration, capacity utilization, proximity of the nearest remaining shipment, number of visited stops, location type of visited stops, goods type, and vehicle type.

The SS choice model is a Multinomial Logit, a model which chooses which shipment to add to a tour, if the decision to add another shipment to the tour is made. For this purpose, we sample a choice set of six shipments that a carrier has to transport on a day, based on selection criteria such as correspondence to vehicle capacity. In the SS choice model, the probability of selecting a shipment increases when the shipment adds a lower generalized cost to the tour, has more stop locations in common with the tour, and has the same goods type as the other shipments in the tour.

While our iterative model structure simplifies the complex tour formation process, many objectives and constraints that influence freight tour formation are considered by the model. For example, carriers wish to construct efficient tours to minimize transportation costs, but at the same time might prefer to construct tours with few stops in order to keep the planning simple. Important constraints, such as vehicle capacity, availability of shipments, and maximum work shifts, are respected. In addition, the model acknowledges the differences in tour formation for various types of goods, vehicles, and locations. For example, direct tours are constructed for cement/concrete shipments, while a tour with multiple stops is more likely when goods are loaded at a distribution center. Several features of freight tour formation are not included, though, due to data availability. Most importantly, we do not consider empty trips and time window constraints.

To validate the model, we use it to construct tours with the shipments listed in data that we separate from the data used for estimation. The validation shows that the model can reproduce observed distributions of tour statistics such as number of stops and tour distance very satisfactorily for a given set of shipments, even for different location and goods types. Additionally, a sensitivity analysis with varying travel times in the network shows plausible results. When travel times increase, fewer direct tours are made because of a stronger focus on travel time savings and fewer tours with a very large number of stops are made because working hour constraints are violated more quickly with longer travel times.

As highly promising validation results are obtained when we apply the model to the shipments of a different set of carriers than the carriers that provide data for model estimation, we conclude that the model can be used to construct tours for other carriers in a shipment-based freight simulation framework. However, several conditions must be fulfilled. Firstly, the geographical scope of the framework should be road freight transportation within the Netherlands. Because the Netherlands is a particularly dense and small country, and freight patterns differ between regions (Zhou et al., 2014), other constraints and parameters might be more appropriate in other countries. Secondly, the model should be used to construct tours for third-party carriers. Due to requirements of an XML-interface, our data shows a strong self-selection of third-party carriers with advanced transport management systems. Thirdly, off-peak travel times should be used in the framework, as we used these in our estimations. Finally, a vehicle type choice model and a synthesized set of shipments that are assigned to carriers are needed before tours can be constructed with our model.

The application of our tour formation model in a freight simulation framework provides the most interesting directions for future research. Relevant research relates to (1) shipment synthesis, (2) carrier assignment, (3) integration between tour formation and vehicle type choice, and (4) empty trips. To synthesize a realistic set of shipments between firms or zones, we recommend to analyze the spatial distribution of shipments and the relationships between shipment attributes. In addition, a good decision rule should be developed that assigns shipments to carriers. A vehicle type choice model is currently under development for the MASS-GT framework of the agglomeration of Rotterdam, the Netherlands (de Bok et al., 2018). An appropriate integration of tour formation and vehicle type choice is a challenging task that requires further research; the vehicle type sets constraints prior to tour formation but a larger vehicle may be chosen if the vehicle capacity is reached with a tour. Additionally, a model that predicts empty trips is of large importance. In combination with a traffic assignment module, it is possible to compare observed and predicted link flows, which would provide solid insights into the extent to which our tour formation model improves the predictive performance of a freight simulation framework.

Several priorities are identified to improve the data set of the CBS, although these priorities are relevant for any new freight data collection effort. Firstly, empty trips constitute a large part of all freight trips (Sánchez-Diaz et al., 2015), which is why inclusion of these empty trips is of large importance to understand and predict truck flows. Secondly, we recommend to include consistent variables that measure the volume of shipments and the volume capacity of trucks, because an interview with a transportation planner and our estimation results indicate that the volume capacity of the truck is an important constraint for tour formation. Thirdly, listing intermediate arrival and departure times at stops allows us to understand more aspects of freight tour formation, such as dwelling times and tour sequences. Fourthly, more carrier characteristics, such as the vehicle fleet size, can assist in exposing the heterogeneity of carrier behavior. Finally, adapting the XML-interface to the trucks and planning systems of smaller and own-account carriers would allow us to expose the behavioral differences of third-party and own-account carriers, and to obtain a more representative data set.

List of tables

LIST OF FIGURES

ABBREVIATIONS

BGW	Basisbestanden Goederenwegvervoer
BLR	Binary Logistic Regression
CBS	Central Bureau of Statistics of the Netherlands
CC	Consideration Choice set (a randomly sampled subset of the Feasible Choice set)
DC	Distribution center
ET	End Tour (estimated choice model)
FC	Feasible Choice set (a subset of the Universal Choice set, shipments that violate specified constraints are removed)
MNL	Multinomial Logit
RUM	Random Utility Modeling
TSP	Traveling Salesman Problem
SS	Select Shipment (estimated choice model)
VKT	Vehicle Kilometers Traveled
VRP	Vehicle Routing Problem
UC	Universal Choice set (all the shipments that a carrier has to transport on a day)
XML	Extensive Markup Language

GLOSSARY

Actor	"Individual people, organizations such as firms, or bodies such as nation-states." (Gilbert, 2008:5).
Agent	"Either separate computer programs or, more commonly, distinct parts of a program that are used to represent social actors." (Gilbert, 2008:5),
Behavior	The observable outcome of a set of choices made an actor.
Behavioral	Describing how objectives and constraints lead to the behavior of an actor. I.e. descriptive, but with a focus on behavior.
Buurt	A Dutch administrative zonal unit.
Calibration	Tweaking model parameters to optimally reproduce observed data.
Constraint	A restriction on what a decision maker can choose from (Hillier & Lieberman, 2001).
Direct tour	A tour that visits only one loading and one unloading location.
Descriptive	Describing how things are rather than prescribing how things should be.
Empirical	Based on observed data.
Estimation	Calibration using statistical methods such linear regression and multinomial logistic regression.
Home base	The location where a vehicle is returned to after a work shift.
Objective	A measure of performance that reflects what is a desirable situation to a decision maker (Hillier & Lieberman, 2001).
Own-account carrier	A firm that transports its own goods using its own vehicle fleet.
Shipment	A physical object with a unique combination of loading location, unloading location, goods type, and tour that it is allocated to.
Stop	A unique location (buurt) visited in a tour.
Sub tour	A tour with only a subset of its shipments.
Third-party carrier	A firm that transports goods of other parties.
Tour	A sequence of visiting locations to load and unload shipments. *In the XML microdata (and in our model):* A sequence of visiting locations to load and unload shipments, starting at the location where the first shipment was loaded into an empty vehicle, and ending at the location where the last shipment was unloaded or at the home base.
Trip	A drive between two locations. A tour can consist of multiple trips.

1. INTRODUCTION

Freight transportation facilitates trade of goods between regions. Therefore, it is crucial for economic development and provision of regions with the resources they have a shortage of (Sánchez-Díaz et al., 2015). However, it also carries many negative impacts on society. For example, compared to passenger cars, freight trucks have a disproportionally large impact on congestion, air pollution, traffic safety, and pavement wear (Hunt & Stefan, 2007; Quak, 2008; Kim et al., 2014). Furthermore, both the amount and share of freight traffic on the roads is expected to increase (Tavasszy, 2008). Projections are as high as a quadrupling of international freight volumes by 2050 (ITF-OECD, 2015). In the Netherlands, the total volume of freight transported increased with 2.5% in 2016 (CBS, 2017a). Due to its large and ever increasing impacts on society, the public sector wishes to develop policies that steer freight transportation towards more sustainability and efficiency.

1.1 PROBLEM STATEMENT

Simulation models are an important tool to support the development and evaluation of freight transportation policies (de Bok & Tavasszy, 2018). Such models aim to represent the complex system that leads to freight flows. Expected impacts of future scenarios and policy measures can be calculated with these models.

The freight system is highly complex and heterogeneous. It features many different types of goods, vehicles, actors, and interactions (Alho et al., 2017; Khan & Machemehl, 2017). Exclusion of these details causes simulation models to have a low behavioral foundation. It also makes these models unsuitable for calculation of impacts of low-level tailor-made policies. Therefore, freight simulation models “are becoming increasingly disaggregate” (de Bok & Tavasszy, 2018:127).

One of these disaggregate directions that freight modeling has taken, is the field of agent-based modeling. In agent-based modeling, decisions made by individual actors are represented with agents. Agents can be defined as “either separate computer programs or, more commonly, distinct parts of a program that are used to represent social actors” (Gilbert, 2008:5), whereas actors can be “individual people, organizations such as firms, or bodies such as nation-states” (Gilbert, 2008:5). Agent-based modeling is a technique especially suitable for understanding and representing systems where decisions and interactions at the individual level play a key role (Gilbert, 2008), which is the case in freight transportation.

When the focus is on freight transportation by road, one of the key aspects that should be incorporated in an agent-based simulation model is tour formation. Tour formation is the construction of a set of tours from a set of shipments. A tour is a sequence of visiting locations, while a shipment is defined in this research as a physical object with a unique origin (loading location), destination (unloading location), and goods type. Many researchers have different definitions of a tour; therefore, we leave this definition rather broad in this introduction.

When a tour contains multiple shipments, the origin and destination of the shipment do not need to match the origin and destination of the vehicle, as other locations may be visited in the tour too (Holguín-Veras et al., 2014; Sánchez-Díaz et al., 2015). For this reason, a good tour formation model is expected to lead to more accurate predictions of vehicle flows (Figure 1.1). Furthermore, different tours lead to different tour distances. Consequently, tour formation is an important phenomenon that should be considered in prediction of the total Vehicle Kilometers Traveled (VKT) in a network (Figliozzi, 2007; Khan & Machemehl, 2017).

Tour formation has increasingly gained attention in the scientific literature on freight modeling. Two main directions can be identified in tour formation modeling: (1) tour construction algorithms that use discrete choice models (e.g. Hunt & Stefan, 2007; Nuzzolo et al., 2012) and (2) vehicle routing optimization techniques from the field of operations research (e.g. Boerkamps & van Binsbergen, 1999; You et al., 2016).

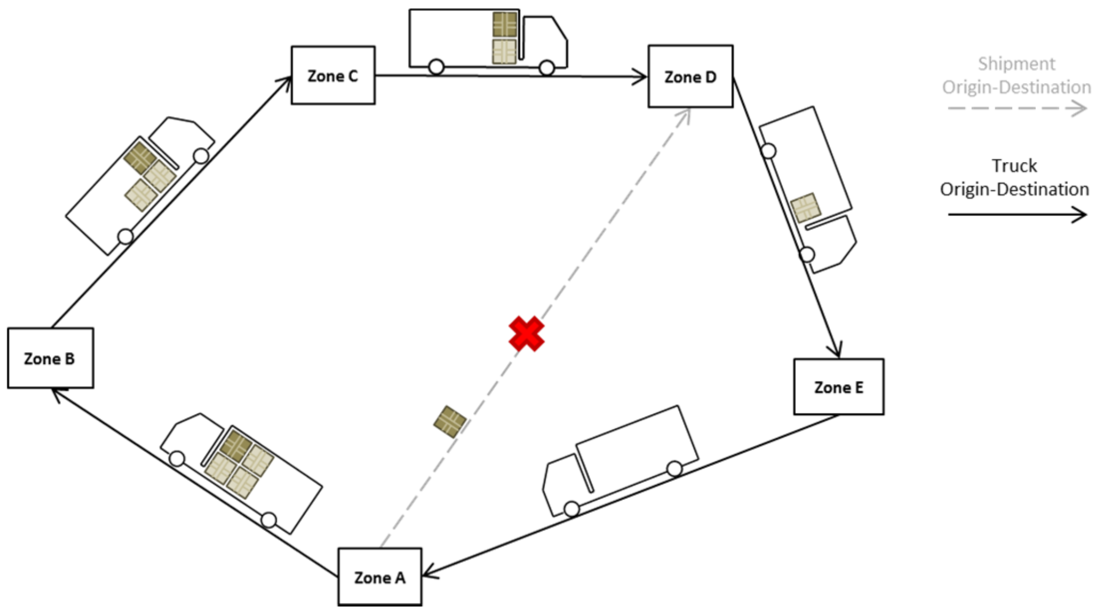

Figure 1.1. A good tour formation model allows us to consider that a truck does not have to drive directly from the origin to the destination of a shipment.

Very few tour formation models are both shipment-based and calibrated in a statistical way. Shipment-based models provide a more accurate representation of behavior because many decisions in freight transportation, including tour formation, are made at the level of shipments (de Bok et al., 2018). Modeling shipments also allows us to consider the different economic characteristics, constraints, and geographical distribution of goods types (Holguín-Véras et al., 2014) and to analyze the effects of more detailed policies and scenarios (Boerkamps & van Binsbergen, 1999). Statistical calibration allows us to understand and reproduce observed tour patterns better, as hypothesized influences on tour formation can be tested empirically. The tour formation models of Nuzzolo et al. (2012) and Outwater et al. (2013) are shipment-based and statistically calibrated. However, they model only a small facet of freight transportation: Nuzzolo et al. (2012) only model the restocking tours of retailers while Outwater et al. (2013) only consider tours that distribute food and manufactured goods from a warehouse.

To improve the current state of tour formation modeling, we develop and estimate a new tour formation model using an innovative and enormous data set of observed tours on the Dutch road network: the XML microdata. This data is collected by the Dutch Central Bureau of Statistics (CBS) and lists approximately 2.6 million shipments with information such as: the tour that the shipment was transported in, gross weight, and goods type. Innovative data collection methods are used, carriers can install an XML-interface that allows survey data to be delivered automatically from the carriers' transport management systems (de Bok et al., 2018). This data set is not a representative sample of all Dutch freight carriers, though, a self-selection of third-party carriers (i.e. firms transporting goods for other parties) with advanced planning systems has taken place. However, from a methodological point of view, this large data set provides a unique and valuable opportunity to calibrate a behavioral tour formation model.

1.2 Research objective and scope

The objective of this research is to develop a tour formation model that can allocate shipments to tours in a way that is similar to observed tour formation.

Due to data availability, the scope of this research is road freight transportation that takes place within the Netherlands. For the same reason, no complex logistics chains with multiple legs are included, but only shipments with one loading and one unloading location. We will focus on the formation of tours out of shipments. The generation of shipments between firms and the vehicle type choice are outside of the scope and assumed to be given.

1.3 Research questions

To reach the objective of this research, the following research question is formulated.

Can we develop a behavioral shipment-based tour formation model that reproduces observed tour patterns?

The following sub questions guide us in answering the main research question.

- Which objectives, constraints, and other factors influence freight tour formation?
- To what extent is the XML microdata useful for calibration of a freight tour formation model?
- How can we structure the allocation of shipments to tours in such a model?
- Which aspects of freight tour formation can we include in the model?
- How well does the model reproduce observed tour patterns?

The term behavioral is used in the research question to underline that our model describes the process that leads to behavior, in contrast to normative models that prescribe optimal behavior. Behavior is the observable outcome of a set of choices made by an actor.

1.4 Research approach

To answer the research questions, we follow the approach summarized in Figure 1.2.

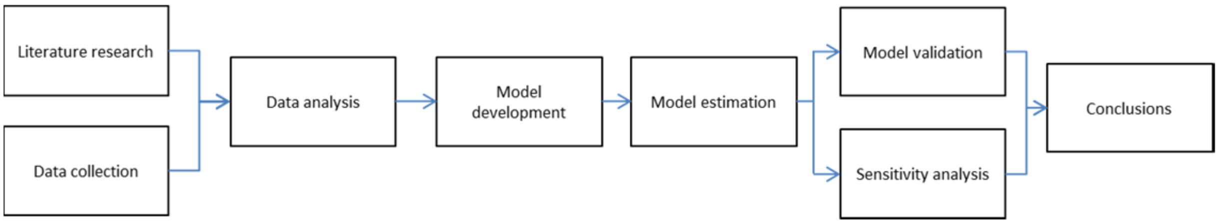

Figure 1.2. The research approach.

An extensive literature research is performed with two main objectives: (1) understanding freight tour formation and (2) discovering gaps in behavioral tour formation models/preventing reinvention of the wheel. For the first objective, we scan the scientific literature on the actors, objectives, and constraints involved in freight tour formation. For the second objective, we do the same with respect to freight simulation models that incorporate tour formation.

Several sources of data are collected for this research. Firstly, we have access to the aforementioned XML microdata. Secondly, skim matrices with travel times and distances between 'buurten' (a Dutch administrative zone unit), and information about these buurten (Kerncijfers wijken en buurten 2015) are used to enrich the XML microdata. Finally, qualitative data is collected in the form of an interview with a transportation planner of Rensa BV, a Dutch wholesaler of heating and ventilation products, in order to further improve our understanding of tour formation.

Next, statistics about variables such as tour duration and number of stops are obtained to improve our understanding of freight tour formation. Analyzing the data also assists in discovering its possibilities and limitations for development of a tour formation model.

Based on our understanding of tour formation and the available data, a model structure is developed that forms tours out of shipments. Parameters of this model are estimated with the available data. Next, we validate this model by applying it to a separate part of the data, and comparing the observed and predicted tours. In addition, a sensitivity analysis is performed with scenarios with varying travel times between zones. Finally, conclusions and further recommendations are formulated based on the findings.

Note that the flow diagram in Figure 1.2 does not have to be purely directional, feedback loops can occur. For example, findings of the data analysis influence the hypotheses tested in the interview (data collection) and the validation results can guide the choice for model specifications (development and estimation).

1.5 CONTRIBUTION

The contribution of this research can be summarized as follows. We have developed a behavioral tour formation model that can form tours for a given set of shipments. This model improves on previous studies by satisfying the following four criteria: (1) shipment-based; (2) statistical calibration on empirical data; (3) no limitation to specific goods types or (un)loading locations; (4) the number of stops per tour and number of tours per day is an outcome of shipment allocation decisions. Estimation of model parameters only requires data that reports: (1) which shipments are part of the same tour and (2) the loading and unloading locations of these shipments.

1.6 OUTLINE OF THE REPORT

This report has the following structure consisting of eight chapters. In Chapter 2, we give an overview of the actors, objectives, and constraints in freight tour formation. In Chapter 3, previous freight modeling efforts that incorporate tour formation are discussed. Chapter 4 explains the structure of the XML microdata and provides descriptive statistics of tour variables. In Chapter 5, we discuss our developed tour formation model structure and explain how we can estimate its parameters, which are reported and interpreted in Chapter 6. In Chapter 7, the validation of the model and the sensitivity analysis are reported respectively. Finally, in Chapter 8, answers to the research questions and recommendations for model improvement, future research, freight policies, and data collection are formulated.

2. UNDERSTANDING FREIGHT TOUR FORMATION

In this chapter we use the literature on freight tour formation and the interview with a transportation planner in order to understand why one tour differs from the other. Firstly, we provide a framework of the actors involved in urban freight tour formation. Secondly, we discuss which objectives and constraints influence the tours that are formed. Finally, we discuss the ways in which we can operationalize differences in tours, and factors that have been found or hypothesized to be able to explain these differences. It should be noted again that the scope of this research, and also of this chapter, is road transportation.

2.1 ACTORS INVOLVED IN FREIGHT TRANSPORTATION

In order to understand freight tour formation and to be able to accurately model it and evaluate policies aimed at influencing it, knowing which actors are involved in which decisions is crucial (Taniguchi & Tamagawa, 2005; Roorda et al., 2010; de Oliveira & de Oliveira, 2017). Observed behavior in freight transportation is rarely the result of decisions made by a single actor (Abate & Kveiborg, 2010). Instead, emergent behavior occurs due to complex interactions between heterogeneous actors (Anand et al., 2014). Additionally, these interactions are heterogeneous (e.g. different types of supply chains, long-term and short-term contracts), and one actor can take on different roles (Roorda et al., 2010). Due to this complexity and heterogeneity, a one-size-fits-all framework of actors and their interactions is not achievable. Subsequently, simplifications and assumptions have to be made, while keeping in mind that situations different from the framework are likely to occur (Roorda et al., 2010).

To arrive at a framework of actors that influence tour formation, distinguishing the actors and roles they play is useful. In this case, an actor can be a single firm, whereas a role describes the actions and decisions an actor takes in a certain setting (de Bok & Tavasszy, 2018). For example, a specific retail shop may take on the role of receiver for one shipment, whereas this shop is the shipper of another shipment.

Many different categorizations of roles with different wordings have been proposed. Three roles that are often described in these categorizations are related to the actions an actor takes with respect to a shipment: (1) the shipper, (2) the receiver, and (3) the carrier (i.e. transporter) of a shipment (Taniguchi & Tamagawa, 2005; Roorda et al., 2010; Stathapoulos et al., 2012; Anand et al; 2014; Zhou et al., 2014; de Oliveira & de Oliveira, 2017). Roorda et al. (2010) also mention the driver of the truck as another actor (role) that may be able to influence the tour formation. Two other actor (roles) that are sometimes identified, are city administrators and residents (e.g. Taniguchi & Tamagawa, 2005; de Oliveira & de Oliveira, 2017). For our purposes, we do not take these last two into account in our framework. City administrators are implementers of policies and are, therefore, not explicitly represented in a tour formation model, but are rather represented through manual implementation of city logistics policies in such a model. Residents experience the negative side effects, and their interests may be represented by looking at KPIs of the model, such as the total Vehicle Kilometers Traveled (VKT). In the remainder of this section, we discuss the actions and decisions of each role with respect to tour formation, to arrive finally at a conceptual diagram of actors/roles involved in urban freight transportation.

Shippers are firms that distribute goods for others to receive (McCabe et al., 2006). The shipper decides between performing transportation by itself (in which case it is also the carrier of the shipment, called an own-account carrier), or outsourcing transportation to a third party (Wisetjindawat et al., 2006; Roorda et al., 2010). Shipments can be transported by a third-party carrier on a contract-to-contract basis, but also long-term contracts with third-party carriers may be made (Roorda et al., 2010). Shippers may bundle shipments under the same contract to reduce transportation costs (Irannezhad & Hickman, 2016). The shipper is also usually the actor that pays for the transportation and sometimes chooses the mode/vehicle type (McCabe et al., 2006). Furthermore, shippers may decide to send the shipment directly or to use an intermediate distribution center (Irannezhad & Hickman, 2016).

The receivers of the goods are generally considered to make decisions regarding the following: order quantity and frequency (Anand et al., 2014; Irannezhad & Hickman, 2016), shipper selection (Wisetjindawat et al., 2006; Anand et al., 2014; Alho et al, 2017), and time windows (Taniguchi & van der Heijden, 2000; Poot et al., 2002; Quak, 2008; Zhou et al., 2014; de Jong et al., 2016). McCabe et al. (2006) mention that receivers may sometimes influence the mode choice. Receivers may also specify that they need to be served first or last during a tour, that certain goods types may not be shipped within the same tour, or only with a special vehicle (e.g. refrigerated transport) (Poot et al., 2002). Together with the shippers, the receivers constitute the demand for goods movement between locations. In general, receivers tend to set conditions for the transport as performed by the carriers (van Duin et al., 2012).

The carrier is the actor that performs and plans the transportation and is, therefore, the most notable actor that influences the way tours are formed. Shippers and receivers constitute the demand for goods movement, and may set constraints, but the carrier is the main actor that decides how tours are formed with the given set of shipments and constraints. The carrier decides which shipments are sent in the same tour, which vehicle and driver are used for the tour, the stop sequence of a tour, the tour departure time, and sometimes even the route between the stops (McCabe et al., 2006; Roorda et al., 2010). Another more strategic decision is the procurement of new vehicles (Alho et al., 2017).

Within a firm that acts as a carrier, the transportation planners make these tour formation decisions. This is usually done at a tactical level using optimization software (You, 2012). For Rensa BV, this software only serves as a support to the transportation planners. The transportation planners decide which shipments are allocated to the same tour, while the software checks for constraints such as time windows and provides the optimal route with given stop locations.

The drivers of trucks can influence tours at a more operational level. They may decide to change routes, or even the stop sequence, for example, because of congestion or lunch breaks (McCabe et al., 2006). Their work shift patterns also influence which tours the carrier can construct (Hunt & Stefan, 2007; Figliozzi, 2007), which was underlined in the interview. The driver is the actor that actually transports the goods and may also assist with (un)loading at stop locations (Quak, 2008).

A single actor can employ different roles, even for the same shipment. One example is an own-account carrier, a shipper that transports goods with its own vehicle fleet (McCabe et al., 2006; Alho et al., 2017). At a more tactical level, a shipper can decide for each shipment whether it uses its own vehicle fleet for transportation or outsources to a third party carrier (Alho et al., 2017). An actor that acts as a shipper for one shipment, may also very likely be the receiver of other shipments. In more complex supply chains, many individual actors may act as the receiver and shipper of the same shipment. Because of these complex linkages between actors and their roles, it is important to distinguish the two. More concretely, in an agent-based microsimulation framework, it is important to assign which agents act as the shipper, carrier, and receiver of a shipment.

The most notable conclusion that arises from this short literature overview of actors that influence tour formation is that carriers tend to be the ones that construct the tours at a tactical level, although the shippers and receivers of goods influence the tour formation indirectly by deciding who transports the shipments, which shipments are sent between which locations, and by specifying constraints such as time windows. Furthermore, in an agent-based model it is important to distinguish between agents and the roles they fulfill for each shipment. In Figure 2.1, the findings of this section are graphically summarized in a conceptual diagram.

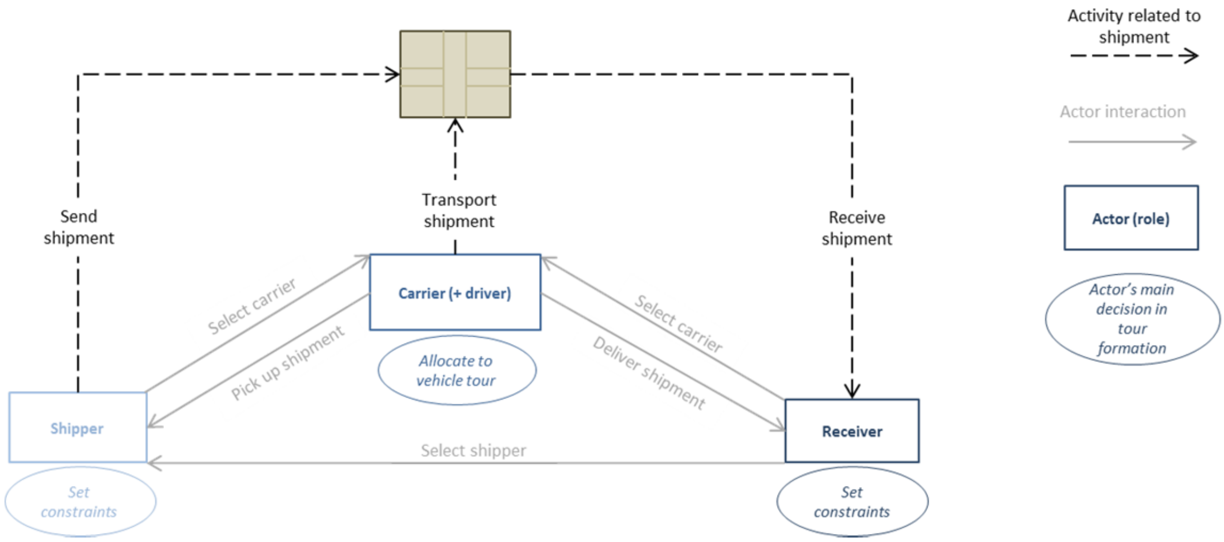

Figure 2.1. The actors/roles involved in freight tour formation and their key decisions and interactions.

2.2 OBJECTIVES AND CONSTRAINTS THAT INFLUENCE TOUR FORMATION

In the previous section, we identified that the transportation planner of a carrier constructs tours while respecting constraints. In doing so, the carrier has specific objectives in mind. In this section, we provide an overview of the objectives and constraints in freight tour formation. We first report objectives as found in the literature, and then relate these to findings from the interview with Harrie Tissen (transportation planner at Rensa BV) that is reported in Appendix A. Next, we do the same with respect to constraints.

Let us begin by defining what objectives and constraints are. Objectives provide a measure of performance to a decision maker, and reflect what is a desirable situation to this decision maker (Hillier & Lieberman, 2001). For objectives, trade-offs are possible, but constraints are characterized by their non-compensatory character (Martínez et al., 2009). They provide restrictions on what the decision maker can choose from (Hillier & Lieberman, 2001).

2.2.1 OBJECTIVES OF THE CARRIER

Carriers are (usually) private firms, and their objective can be considered to be simply profit maximization or cost minimization (You, 2012). However, it is all but straightforward to predict decisions based on profit maximization. Carriers may have particular strategies, tactics, and knowledge to operationalize profit maximization. Furthermore, inter-carrier heterogeneity is likely to exist, different carriers may operationalize profit maximization with different strategies and tactics, and may have different perceptions of costs imposed by certain decisions. Not all decisions made by carriers have to be driven merely by profit maximization, as more subjective preferences can also play a role. Therefore, it is important to dig deeper into what may drive the decisions of carriers.

You et al. (2016) distinguish a set of objectives that represent the strategies employed by carriers of drayage trucks in the San Pedro Bay Ports (SPBPs) in Southern California, USA. One of these objectives is to maximize the number of visits to the same stop location in a tour. This can be understood as a particular strategy employed by these carriers to maximize the profit of each tour.

Furthermore, carriers may minimize the total travel time of their truck operations (You et al., 2016). Longer travel times for the same set of shipments lead to increased costs related to fuel and driver wages and a subsequent decrease in profits (Anand et al., 2014). This is a key factor explaining tour formation behavior. In their model, Wisetjindawat et al. (2006) assumed time minimization to reflect fully the tour formation objectives of carriers. Truck access restrictions and congestion increase travel times and, therefore, influence tour formation decisions (Hunt & Stefan, 2007; Roorda et al., 2010). Instead of the total travel time, the total travel distance (Poot et al., 2002), or a generalized travel disutility based on both distance and time, can be used as an objective (Hunt & Stefan, 2007).

The other objectives You et al. (2016) distinguish are the following: (1) minimizing total truck operating hours, alternatively called the makespan of a day, (2) reducing total emissions, which can be seen as a form of Corporate Social Responsibility, but also a consequence of policies penalizing these carriers based on emissions, and (3) reducing early and late deviations from goal arrival times. The latter can be related to improving predictability of the carrier's own operations and also to improving level of service to keep customers satisfied (You et al., 2016). In some cases, early and late deviations can be treated as a constraint in the form of hard time windows, in which delivery or pick-up outside the time window is not allowed. This will be discussed in the next section.

Other objectives than those identified by You et al. (2016) have been found in the literature. These mostly relate to the coordination of the different tours of a carrier. Based on years of experience in adjusting their VRP software to the desires of clients, Poot et al. (2002) identified that carriers wish to generate a 'visually attractive' set of tours. Visual attractiveness is subjective and can, therefore, be operationalized in many ways, but Poot et al. (2002) mostly relate this to the extent to which tours can be clearly identified on a map. Factors that Poot et al. (2002) use to represent visual attractiveness include: the number of crossings between and within tours, the average number of nodes that is closer to the center of gravity of another tour than of its own tour, the average distance of nodes to the center of gravity of its tour, and the average number of nodes per tour that is captured in the convex hull of another tour. Visually attractive tours are more tractable for planners, and seem more intuitively logical and are therefore more easily accepted by drivers (Poot et al., 2002). Figliozzi (2007) endorses that it is desirable for carriers to have tours that do not cross each other, also to reduce the total VKT. Furthermore, Hunt & Stefan (2007) speculate that in industries where tours tend to have a looping shape instead of a zig-zag shape, there may be more time to plan tours in advance, such that tours can be constructed that look more orderly.

Related to the ease of following and distinguishing the tours, a transportation planner may also wish to reduce the complexity of the operations. More concretely, if it is not very beneficial to construct tours with many stops, a preference is sometimes present for simple tours that serve only one customer. Nuzzolo et al. (2012) found that when many customers are in short reach of the home base, the constructed freight restocking tours in Rome, Italy tend to include fewer stops. They hypothesized that this is due to the desire to reduce the complexity of the tour planning.

Another objective related to coordination of tours is to have a balanced tour set. This means that there is little deviation in the duration of different tours (Bodin et al., 2003). This leads to cost savings, less overtime hours for drivers, and a higher sense of equal and fair treatment of drivers (Bodin et al., 2003). It may also reduce the total truck operating hours of the carrier, one of the objectives identified by You et al. (2016).

Findings from interview with transportation planner

Tissen endorses that, for a transportation planner, the total travel time and distance are key aspects that reflect the attractiveness of the planning. More generally, the objective is to transport as many goods with the use of as few resources as possible. Closely related to time and distance, Rensa BV also tries to minimize their CO_2 emissions.

Another important objective is to construct a planning with a high probability that all orders are indeed delivered in the planned tours. If too many customers need to be visited in the nine hour shift of a driver, the driver may not be able to visit all customers, which leads to extra waiting time for these customers. The desire to construct a balanced tour set also resonates. Preferably, tours that last a full shift of nine hours are constructed in order to reduce problems relating to drivers that need to work more or fewer hours the next day. The same customer is usually served by the same driver, this driver knows about specific instructions of the customer and is more familiar with the routes leading to this customer.

Another objective, which has not yet been mentioned, is to construct tours with a high capacity utilization. This relates to the goal of transporting as many goods with as little resources possible, and can be seen as a particular strategy employed to minimize transportation costs.

Finally, Tissen does not confirm that there is a preference for tours with few stops at Rensa BV. If this is possible within a nine hour shift, 30 to 35 stops in a tour are not considered undesirable. This does not mean, however, that such a preference does not exist for other carriers.

2.2.2 Constraints

The demand for freight transportation is derived from the demand for goods (Figliozzi, 2007). The resulting shipments between shippers and receivers form the basis from which the carriers can start forming tours. The carrier is constrained in its construction of tours to those shipments that shippers and receivers want to be transported (Figliozzi et al., 2007).

Incompatibility of goods types (Poot et al., 2002; Alho et al., 2017) and special vehicle requirements for certain goods types (Beziat et al., n.d.) also constrain how tours can be formed. Some combinations of products may be forbidden (Poot et al., 2002). This restriction can be imposed by policies, by the receiver, or by simple common sense. For example, live animals should not be transported simultaneously with filled oil barrels (Robroeks, 2016). Some shipments require a dedicated vehicle, such as concrete shipments, which cannot be used to transport other goods types (Beziat et al., n.d.). Specific regulations apply to the transportation of food products, which restricts their compatibility with other goods types (Beziat et al., n.d.).

Hard time windows are constraints that can greatly reduce the number of customers that can be served in the same tour (Figliozzi, 2007) and may also impact the stop sequence and departure time (Quak, 2008). Time windows specify a period of time during which serving a customer is allowed. Such constraints can be imposed by the municipality but also by the receiver (Quak, 2008). In the Netherlands, approximately half of all municipalities have areas with time windows for freight (un)loading activities, usually from 7AM-11AM (CBS, 2015a). Time windows may be imposed on different parts of the day and with different time widths (Quak, 2008; de Jong et al., 2016). For municipalities, the reason to impose a time window on an area is usually to reduce safety risks and other negative impacts such as noise to the shopping public and nearby inhabitants, while for receivers this may allow for better planning of activities and ensuring that staff is available to receive the goods (Quak, 2008).

Another constraint in planning tours relates to the available vehicle fleet of the carrier and the capacity of these vehicles. A carrier simply cannot have more vehicles simultaneously transporting goods than the number of vehicles it owns. However, at a more long-term planning level, the carrier may choose to buy new vehicles, although the carrier will try to minimize the size of its vehicle fleet due to accompanying procurement and depreciation costs (Bodin et al., 2003; Alho et al., 2017). Different types of vehicles exist, with different permissible carrying capacities (Poot et al., 2002), which can be defined as maximum weight, volume, or both (Wisetjindawat et al., 2006). The vehicle capacity and the size of the shipments restrict which shipments can be delivered and/or picked up in the same tour (Poot et al., 2002; Figliozzi, 2007).

A very straightforward constraint is that shipments need to be loaded before they can be unloaded (Robroeks, 2016). A shipment may have a loading and unloading location both found elsewhere than the tour starting point. In this case the loading location has to be visited before the unloading location is visited, which can further constrain which tours can be constructed. The type of activity performed at a stop (loading or unloading) also closely relates to capacity constraints. Not all shipments need to be present in the vehicle at the same time. For example, in the same tour a set of shipments may first be delivered and afterwards another set of shipments may be picked up (Kim et al., 2014). The total volume or weight of the shipments transported in this tour may exceed capacity, but at no point in time does the set of shipments present in the vehicle exceed its capacity.

Finally, other often mentioned constraints in tour formation decisions are related to the interests of the drivers. Most notably, work shift patterns of the drivers dictate which tours can be formed (Bodin et al., 2003; Wisetjindawat et al., 2006; Hunt & Stefan, 2007; Figliozzi, 2007; Figliozzi et al., 2007). This can be in the form of a maximum tour duration (Poot et al., 2002; Bodin et al., 2003; Wisetjindawat et al., 2006), a maximum number of customers/stops per tour (Bodin et al., 2003), or a scheduled time for lunch breaks (Hunt & Stefan, 2007). The tour duration does not only include time spent traveling on the road but also dwelling time at stops. Given the maximum work shifts, a longer dwelling time negatively influences the number of customers that can be served in the same tour, which leads to increased transportation costs (Figliozzi, 2007).

FINDINGS FROM INTERVIEW WITH TRANSPORTATION PLANNER

The three main constraints mentioned in the interview are time windows, volume restrictions, and tour duration. Volume is much more often the limiting factor than weight, especially so in the case of heating and ventilation products. The duration of the work shift of the driver is a crucial constraint guiding the tour formation too, as maximum working hours are defined and regulated by the Dutch government.

For Rensa BV, the daily fluctuation of the number of shipments is high. On days with more shipments, usually more stops are found per tour, as there is a higher potential to combine nearby shipments. This further underlines the fact that the available set of shipments is a key constraint in tour formation.

2.2.3 CONCLUDING REMARKS

To conclude this section, there are many different objectives and constraints that can influence which tours are formed. In Table 2.1 and 2.2 the ones we have found in the literature and in the interview are summarized. Different carriers may have slightly different objectives, and for different carriers, days, and tours there may be different constraints guiding the tour formation. In Table 2.1 and 2.2, some similar factors are found both as an objective and as a constraint. Examples of this include the number of stops in the tour and the vehicle capacity utilization. For the purpose of structuring this section, we distinguished between objectives and constraints, but as it shows, the line of demarcation is not always straightforward.

Table 2.1. Objectives in tour formation.

Objectives	Operationalization	References	Interview[1]
Low travel costs / high profit	MIN travel time MIN travel distance MIN generalized travel cost MAX # visits to same node in tour	Poot et al. (2002; Wisetjindawat et al. (2006); Hunt & Stefan (2007); Roorda et al. (2010); Anand et al. (2014); You et al. (2016)	V
Makespan	MIN total operating hours	You et al. (2016)	0
Emissions	MIN emissions	You et al. (2016)	V
Punctuality	MIN early arrival times MIN late arrival times MIN probability of not visiting planned customer	You et al. (2016)	V
Visual attractiveness	MIN # crossing between tours MIN # nodes closer to center of gravity of other tour MIN Distance of stops to center of gravity	Poot et al. (2002); Figliozzi (2007); Hunt & Stefan (2007)	0

[1]V = confirmed in the interview
X = refuted in the interview
0 = not mentioned in the interview

	MIN # nodes that is captured in convex hull of other tour MAX enclosed angle between stops (looping shape)		
Simple tours	MIN # stops per tour	Nuzzolo et al. (2012)	X
Level of balance	MIN deviation tour durations	Bodin et al. (2003)	V
Capacity utilization	MAX used vehicle volume capacity		V

Table 2.2. Constraints in tour formation.

Constraint type	Examples / clarification	References	Interview
Available shipments	Only those shipments that customers wish to be transported can form the basis of a tour	Figliozzi (2007); Figliozzi et al. (2007)	V
Goods type	Compatibility goods Special vehicle requirements	Poot et al. (2002); Robroeks (2016); Alho et al. (2017); Beziat et al. (n.d.)	0
Time window	Municipal time window zone Customer imposed time window	Figliozzi (2007); Quak (2008); de Jong et al. (2016)	V
Vehicle capacity & shipment size	A vehicle has a maximum permissible weight & volume, which the shipments together may not exceed. The carrier is also constrained to which vehicles it has available	Poot et al. (2002); Bodin et al. (2003); Figliozzi (2007); Wisetjindawat et al. (2006); Alho et al. (2017)	V
Precedence stop locations	The loading location of a shipment must be visited before its unloading location	Robroeks (2016)	0
Driver work hours	Maximum tour duration (incl. dwelling) Maximum travel duration Maximum tour distance Maximum number of stops/customers	Poot et al. (2002); Bodin et al. (2003); Wisetjindawat et al. (2006); Hunt & Stefan (2007); Figliozzi (2007); Figliozzi et al. (2007)	V

2.3 Differences in tours

In the previous section, we identified that tours differ from each other due to different objectives and constraints acting upon them. Several ways exist to operationalize these differences between tours, which we discuss in Section 2.3.1. Next, we discuss some of the explanations for these differences, based on the constraints discussed in the previous section and empirical findings in the literature.

2.3.1 Characterizing tours

A first and often used factor to distinguish different types of tours is the number of different locations visited during a tour, alternatively called the number of stops or activities per tour (Figliozzi, 2007; Hunt & Stefan, 2007; Pluvinet et al., 2012; You, 2012; Mohammadian et al., 2013; Kuppam et al., 2014; Zhou et al., 2014; Beziat et al., n.d.). Variations in the definition of a stop are found throughout the literature, most notably Figliozzi (2007) distinguishes each customer served as a separate stop, whereas Kuppam et al. (2014) aggregate all customers served in the same zone as a stop. Using raw GPS data, Pluvinet et al. (2012) distinguish each stop as a location where the vehicle has a speed lower than 3 km/h for more than 150 seconds. Instead of distinguishing each visited stop location, the total number of shipments transported in a tour can be distinguished (Mohammadian et al., 2013).

Secondly, not only the number of stop locations may be distinguished, but also the type of activity performed at these locations. In freight transportation, most importantly we can distinguish loading (or pick-up) and unloading (or delivery) of goods (Kim et al., 2014; Beziat et al., n.d.). Based on these activities, four types of tours can be distinguished: (1) distribution tours, (2) collection tours, (3) mixed tours, and (4) direct tours (CBS, 2017b; Beziat et al., n.d.). In distribution tours, there is one loading location and several unloading locations. In collection tours, there are several loading locations and one unloading location. Mixed tours have several different loading and unloading locations, and direct tours have only one loading and one unloading location (Figure 2.2).

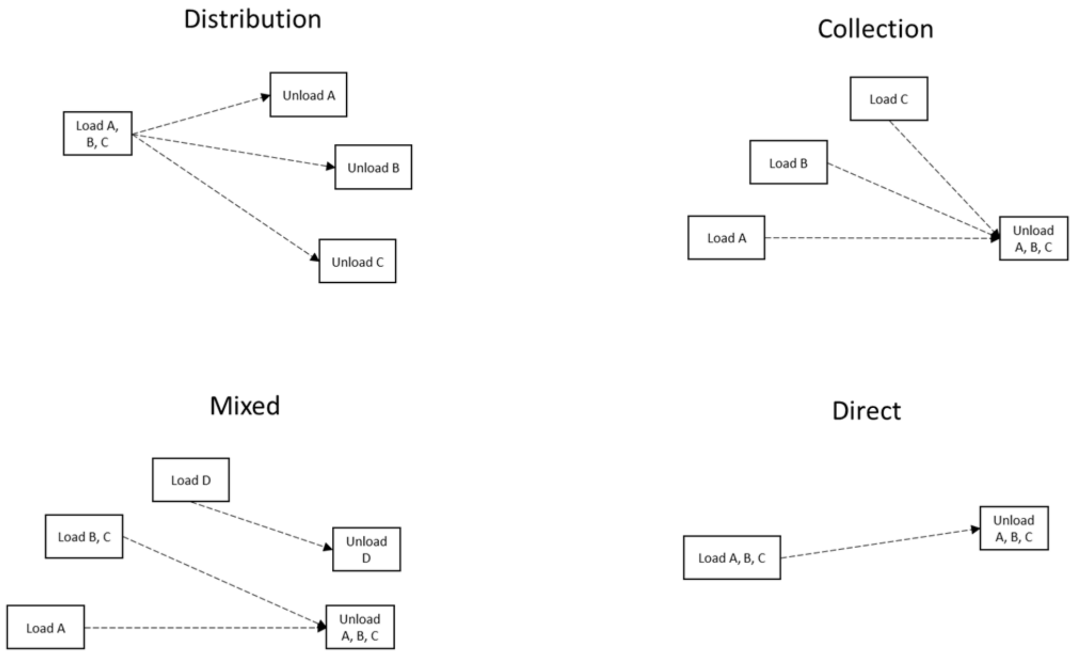

Figure 2.2. Different types of freight tours based on the activities performed at stop locations.

A third way to characterize tours is to look at their total distance or duration (Figliozzi et al., 2007; Hunt & Stefan, 2007; Pluvinet et al., 2012; You, 2012; Mohammadian et al., 2013; Zhou et al., 2014; Beziat et al., n.d.). The tour distance is sometimes aggregated to obtain the Vehicle Kilometers Traveled (VKT) of a whole study region (Figliozzi, 2007). The VKT serves as a Key Performance Indicator for policy makers as it may be used to proxy the impact of freight transportation on emissions and congestion for example. Both the distance and duration can also be disaggregated to trips instead of tours. One can distinguish the average trip distance (Mohammadian et al., 2013; Zhou et al., 2014), but also differences in distance between the first trip and later trips in a tour (Figliozzi, 2007; Mohammadian et al., 2013; Beziat et al., n.d.). The tour duration can be disaggregated to capture the duration of each trip in a tour but also to capture the time that is spent serving customers at a stop, i.e. the stop duration or dwell time (Hunt & Stefan, 2007, Mohammadian et al., 2013; Zhou et al., 2014; Beziat et al., n.d.). The duration and distance can also be combined to obtain an average tour and trip speed (Figliozzi et al., 2007).

Finally, capacity utilization is a term often used to describe the efficiency of freight tours (Abate & Kveiborg, 2010). Many different ways to operationalize capacity utilization have been proposed. The simplest ones only look at the percentage of empty trips (Zhou et al., 2014). More advanced measures take into account distances and the volume or weight of the shipments in comparison to the vehicle capacity (Abate & Kveiborg, 2010).

Table 2.3. Different ways to characterize tours.

Tour characteristic	Operationalization	References
Number of nodes	Number of customers visited Number of stop zones Number of shipments	Figliozzi (2007); Hunt & Stefan (2007); Pluvinet et al. (2012); You (2012); Mohammadian et al. (2013); Kuppam et al. (2014); Zhou et al. (2014); Beziat et al. (n.d.)
Type of activities	Loading/unloading stop Distribution/collection/mixed/direct tour	CBS (2017b), Beziat et al. (n.d.)
Distance	Total tour distance Average trip distance First vs. later trip distance Total VKT	Figliozzi et al. (2007); Hunt & Stefan (2007); Pluvinet et al. (2012); You (2012); Mohammadian et al. (2013); Zhou et al. (2014); Beziat et al. (n.d.)
Duration	Total tour duration Average trip duration First vs. later trip duration Travel tour duration Average dwelling time Total dwelling time	Figliozzi et al. (2007); Hunt & Stefan (2007); Pluvinet et al. (2012); You (2012); Mohammadian et al. (2013); Zhou et al. (2014); Beziat et al. (n.d.)
Capacity utilization	Percentage of empty trips Percentage of empty km Utilized volume-km / capacity volume-km Utilized weight-km / capacity weight-km	Figliozzi et al. (2007); Abate & Kveiborg (2010); Zhou et al. (2014)

2.3.2 Explaining tour differences

Differences in the number of stops per tour can largely be explained by the number, geographical dispersion, and compatibility of shipments. Companies with a high turnover and third-party carriers tend to have more shipments to transport per day and, therefore, more potential to group shipments into a tour and save transportation costs. For this reason, they have been found to construct tours with more stops on average (McCabe et al., 2006; Roorda et al., 2010; Nuzzolo et al., 2012; Beziat et al., n.d.). Tours that transport goods with a higher dispersion of demand and more severe constraints relating to goods compatibility and vehicle type (e.g. refrigerated transport for fresh foods), tend to have fewer stops as there is less potential to group shipments into efficient tours (Figliozzi, 2007; Nuzzolo et al., 2012; Zhou et al., 2014; Beziat et al., n.d.).

Zhou et al. (2014) also found that tours in Texas, USA that include a stop at a retail store tend to have fewer stops. They explain this by tight time windows in the retail sector, which negatively affect the possibility to combine shipments in one tour, echoing analytical findings of Figliozzi (2007) and Quak (2008). Tours that visit a distribution center tend to have more stops, as “the vehicle load gets replenished (for outbound delivery) or emptied (outbound pickup) at the distribution center, which allows the vehicle to go on with more customer visits” (Khan & Machemehl, 2017:95). In contrast, tours that visit a manufacturing establishment tend to have fewer stops, with larger shipment sizes and little remaining vehicle capacity for other shipments (Khan & Machemehl, 2017).

Nuzzolo et al. (2012) found that for restocking tours performed for and by retailers in Rome, Italy, accessibility to wholesalers decreases the probability of making tours with many stops, which they speculated to be caused by the desire to reduce complexity of operations. You (2012) found that drayage trucks in Southern California, USA on average make fewer stops per tour than trucks in Denver, Amsterdam, and Melbourne, which they explained by the fact that these drayage trucks perform longer (un)loading activities, which in combination with a maximum tour duration allows for fewer stops per tour. Different factors influence the dwelling time at a stop, including: truck size (i.e. more time to park), shipment size, if customers help (un)loading, and if both deliveries and pick-ups are performed (Quak, 2008). As these factors may differ for different types of customers and goods, Figliozzi et al. (2007) found that these significantly influence stop duration.

The average and distribution of tour/trip distance and duration is heavily dependent on the spatial layout of a region. Figliozzi et al. (2007) found peaks in the trip distance distribution which they were able to explain by the locations of major freight generating zones and facilities relative to each other when analyzing truck tours made by a freight forwarding company in Sydney, Australia. Working shift patterns explain why You (2012) mainly found tours from 3-9 hours and Figliozzi et al. (2007) found the median tour duration to be 8 hours and most tours to have a distance below 300 km. Figliozzi et al. (2007) also mentioned daily variations in demand to have a large impact on the daily average tour length and duration, underlining yet again that the carrier is constrained to the available shipments in its construction of tours. Beziat et al. (n.d.) and Figliozzi (2007) showed that urban freight tours often include a long first trip from the tour starting point, and shorter trips to later stops in the tour due to concentration of customers in a zone. Together with the objective of travel time minimization, serving all these customers in one tour makes sense, if the vehicle capacity allows.

3. STATE OF THE ART IN FREIGHT FORECASTING MODELS

This chapter provides an overview of previous freight forecasting modeling efforts. Section 3.1 presents a general typology of such freight models, after which we go into further detail discussing those models that incorporate tour formation in Section 3.2. The scientific gap that we fill with this research is defined in Section 3.3.

3.1 A TYPOLOGY OF FREIGHT FORECASTING MODELS

In this section, different types of freight forecasting models for road transportation are discussed. A clear typology of such models is given by Abate & Kveiborg (2010) and Kim & Park (2017). They distinguish four types of freight forecasting models: (1) vehicle-based models, (2) commodity-based models, (3) tour-based models, and (4) hybrid models. The remainder of this section discusses the philosophy and pros and cons of these four model types.

Vehicle-based models, alternatively called truck-based models, follow the four-step model developed for personal transportation and apply that model to freight road transportation. The four steps in this model are trip generation, trip distribution, vehicle type choice, and network assignment (Federal Highway Administration, 2007; Kim & Park, 2017). Vehicle-based models are the simplest and require the smallest amount of data (Kim & Park, 2017), and may be useful to quickly and cheaply assess the needs for infrastructural expansions. However, they ignore two crucial aspects of freight transportation: trips are often chained into complex tours (Hunt & Stefan, 2007; Doustmohammadi et al., 2016b) and the demand for freight transportation is derived from the flow of goods between areas (Boerkamps & van Binsbergen, 1999; Wisetjindawat et al., 2006).

To deal with the latter, commodity-based models have been proposed. Such models do not directly estimate vehicle trips, but estimate commodity flows between areas first. The same four steps as in vehicle-based models are followed, with an additional 'vehicle conversion' step in which commodity flows between areas are converted into vehicle flows based on average payloads per commodity type (Federal Highway Administration, 2007). Commodity productions and attractions per zone may be based on employment, Input-Output (IO) tables, and labor productivity. Many regional freight transportation forecasting models in the US have applied this approach (Federal Highway Administration, 2007). The strength of these models, compared to vehicle-based models, is their explicit consideration that freight transportation is derived from commodity flows (Wisetjindawat et al., 2006; Doustmohammadi et al., 2016a). However, due the highly simplified vehicle conversion, these models cannot accurately consider empty trips. Empty trips constitute a significant part of freight vehicle trips, as the demand is rarely bidirectional (Abate & Kveiborg, 2010). Furthermore, as multiple shipments are often transported in one tour, the origin and destination of a commodity flow do not need to match the origin and destination of its vehicle flow (Nuzzolo et al., 2012; Sánchez-Díaz et al., 2015). Therefore, an explicit consideration of tour formation is desired.

In the Netherlands, a national strategic freight model has been developed (BasGoed), which follows a similar commodity-based approach as described above. Commodity flows between zones are estimated, after which it is calculated how many vehicle trips are needed to deliver these commodity flows (Groot & Miete, 2016). This model implicitly also takes tour formation into account. The module that transforms commodity flows into vehicle trips (deelrittenmodule) uses a database of observed tours (Basisbestanden Goederenwegvervoer, shortly discussed in Chapter 4). The trips within these tours are multiplied based on the magnitude of the commodity flows between zones. Consequently, tours are included in a very implicit and non-behavioral manner in the BasGoed model. The trips within a tour are not connected and no new tours can be formed. As a result, no policy sensitivity is present in the model with respect to tour formation.

Tour-based and hybrid models consider tour formation in a very explicit manner. The difference between these two models is that the first directly estimates vehicle tours while the latter estimates commodity flows that are assigned to vehicle tours (Kim & Park, 2017). Hybrid models have the potential to estimate freight vehicle flows most accurately, as they consider that freight transportation is derived from the flow of goods between regions and that multiple stops may be visited in the same tour. However, hybrid models usually also require the most elaborate data collection efforts and complex calculations (Kim & Park, 2017). In the next section we will give an overview of previous modeling efforts that fall under these last two categories, i.e. models that incorporate tour formation explicitly.

3.2 Incorporating tour formation

3.2.1 Aggregate: Tour-based entropy maximization

Aggregate freight modeling approaches, in contrast to disaggregate approaches, do not explicitly model decisions at the individual level. Therefore, these models require less computation time and data for estimation which makes them more tractable and usable for large study areas (Wang, 2008; Wang & Holguín-Veras, 2009; You, 2012). To the best of the author's knowledge, the only aggregate approach that considers tour formation is the tour-based entropy maximization as developed by Wang & Holguín-Veras (2009).

Entropy maximization seeks the most likely meso state, i.e. a configuration of micro states that complies with constraints acting at the macro level. In standard estimation of Origin Destination (OD) matrices, a micro state is a trip between an origin zone and a destination zone, the meso state is a trip distribution matrix, and the macro state can consist of zonal trip productions and attractions, a travel impedance matrix, and a total travel impedance in the network. Given no further information, entropy maximization assumes that each micro state is equally probable to occur. In that case, the most likely meso state is the one that can be configured in the most possible ways, while complying with macro level constraints (Ortúzar & Willumsen, 2011; You, 2012).

In tour-based entropy maximization, the micro state is not a trip but a tour (here: a node sequence departing from and returning to the same node), and the meso state is a matrix with the number of vehicles that follow each tour (Wang & Holguín-Veras, 2009). Again, it is assumed that each micro state is equally probable to occur. The objective of the tour-based entropy maximization, is to find the most likely set of tour flows that complies with trip productions and attractions and a total travel impedance in the network. In trip-based entropy maximization, the enumeration of all possible micro states is feasible. However, if the micro state is a tour, the number of possible micro states is astronomically large for relatively small network sizes (Wang & Holguín-Veras, 2009). Therefore, Wang & Holguín-Veras (2009) use a heuristic algorithm developed by Wang (2008) to enumerate a large set of tours that are consistent with observed trip chaining behavior. This heuristic uses discrete choice models to represent the decision of the next stop location and tour termination. In Section 3.2.2, this heuristic is treated in further detail.

Instead of a heuristic to generate tours, You (2012) obtained a large set of observed tours, using GPS data of clean drayage trucks in the San Pedro Bay Ports (SPBPs) in Southern California, USA. You (2012) extended the tour-based entropy maximization to include tours with a different sequence of the same set of visited stops and tours that visit a certain node multiple times. Their extended model showed promising results in replicating observed tour flows and distributions of travel time, tour transaction time and tour time. Sánchez-Díaz et al. (2015) extended the tour-based entropy maximization to include a time of day component and used traffic counts as additional macro constraint.

Tour-based entropy maximization is presented as computationally more efficient and less data hungry than disaggregate agent-based modeling approaches (You, 2012). However, these models still require a large and representative set of tours between zones as input. A calibrated tour formation model or a large set of observed tours can be used, but both require large data collection and preparation efforts. While tour formation is considered in these models, there is no formation of tours from a set of shipments, the objective of this research.

3.2.2 Disaggregate approaches

Disaggregate approaches explicitly model individual components of freight operations, such as an individual vehicle or shipment (Wang, 2008). In general, two types of approaches to model freight tour formation at a disaggregate level can be distinguished: (1) approaches that use mathematical optimization in a behavioral way and (2) approaches that use the Random Utility Modeling (RUM) framework to model discrete choices. The literature on the Household Activity Pattern Problem (HAPP) in personal transportation also provides techniques that are useful for modeling freight tour formation but do not fall clearly into one of the two mentioned approaches. In the remainder of this section, we discuss previous disaggregate tour modeling efforts that can be categorized as optimization, RUM, and HAPP approaches respectively.

Optimization and heuristics to describe tour formation

To represent tour formation, optimization techniques from the field of operations research can be used. While these models were developed to prescribe the optimal solution to a decision maker with a set of objectives and constraints, they may also be used in a descriptive context. In fact, in freight transportation a lot of decisions are made with optimization software, which is why it intuitively makes sense to use such optimization techniques to describe behavior in freight forecasting models (You, 2012).

A specific instance of an optimization technique that is relevant to modeling tour formation is the Vehicle Routing Problem (VRP). The goal of the VRP is to allocate shipments optimally to a set of vehicles that depart from and return to a home depot (Solomon, 1987). Numerous variations have been proposed to the VRP, each with slightly different objectives or constraints. Because VRPs are NP-hard problems, they cannot be solved numerically in a feasible amount of time and thus require heuristic methods to approximate the optimal solution in a more efficient way. A large body of literature is dedicated to developing efficient heuristics for these problems (Taniguchi & van der Heijden, 2000; Poot et al., 2002). The goal of this section, however, is to show how these methods have been used in disaggregate freight forecasting models. Therefore, we do not go into further detail describing these different variations and heuristics for the VRP.

Taniguchi & van der Heijden (2000) developed a simulation framework for goods movement between nodes in a hypothetical road network. Ten freight carriers are placed randomly on the network and have a random set of customers to be served with a specified time window for pick-up or delivery. VRPs with Time Windows (VRPTW) are used to construct routes that serve all the customers of the carrier optimally. The routes are assigned to the network, together with passenger traffic, after which congested travel times are fed back into the VRP decision engine to construct new routes. While some of the model assumptions are based on observations, such as the time window width, this model should mostly be seen as a general city logistics evaluation model and a demonstration of how optimization methods can be used in a descriptive context.

Anand et al. (2014) provide another example of such a model for a hypothetical network to evaluate city logistics measures. Their focus is rather on the behavior of and interaction between different agents. Shop agents, for example, choose a shipper to buy from based on distance and optimize their order quantity and frequency. Carriers calculate the transportation costs required to deliver their orders, based on a VRP, and add a profit margin to place a bid on the market. Similarly, Irannezhad & Hickman (2016) propose a multi agent framework, with the extension that receivers set time windows, shippers bundle shipments into one contract based on transportation cost savings, and carriers have a perception of travel times between stops based on congestion in previous periods, which they use as input for a VRPTW. Polimeni et al. (2010) applied a VRP with soft time windows (VRPSTW) when modeling the restocking tours made by retailers on own-account (i.e. performing their own transportation), time windows are included in the VRP formulation as penalties for late and early arrivals outside the time window rather than as a hard constraint.

The models of Taniguchi & van der Heijden (2000), Anand et al. (2014), Irannezhad & Hickman (2016), and Polimeni et al. (2010) are generic frameworks. Consequently, these models have limited use to support decision making on policies for a specific city. Others have developed disaggregate freight models that incorporate tour formation for specific regions or cities. For example, Boerkamps & van Binsbergen (1999) developed a goods simulation model for Groningen, the Netherlands, and Alho et al. (2017) used the city of Singapore for their agent-based freight modeling framework. Boerkamps & van Binsbergen (1999) allocate synthesized goods flows to vehicle tours with a vehicle loading algorithm. Parameters of this algorithm include constraints such as the vehicle capacity, maximum loading factor, and maximum number of stops per tour, which are specified based on the activity type of the origin zone of the goods flow. Tours are only modeled for the food retail and book sector.

In the framework of Alho et al. (2017), shipments are sent and received by a synthetic firm population. They discuss two sub models of their simulation framework, a strategic model and a tactical model. The strategic model focuses on decisions relating to resource acquisition (e.g. commodities, vehicles), whereas the tactical modeling focuses on the allocation of these resources (e.g. tour formation, allocation of drivers and vehicles). Tours are constructed for a set of shipments with a heuristic algorithm, rather than a pure VRP optimization. Firstly, shippers maximize the amount of shipments allocated to their own vehicle fleet, if they own one. Secondly, remaining shipments are pooled together with those of other shippers and are outsourced to carriers. These carriers construct tours as follows: (1) a randomly chosen shipment is allocated to a vehicle, (2) other shipments are added based on proximity of origin and destination and compatibility until vehicle capacity or a maximum amount of deliveries per tour is reached, (3) a 'closest-node-next' heuristic is used to sequence the stops. If this tour exceeds a maximum total tour time, shipments are dropped. Afterwards the stop sequence is optimized with a Traveling Salesman Problem (TSP) based on either time or distance, depending on the priorities of the carrier.

Wisetjindawat & Sano (2003) and Wisetjindawat et al. (2006) developed an agent-based microsimulation framework for the metropolitan region of Tokyo, Japan. To construct tours out of shipments between synthetic firms, a VRP that minimizes total travel time and is constrained by the carrying capacity of a truck and maximum working hours of a driver is used. Donnelly et al. (2010) developed a model for the state of Oregon, USA, in which allocation of shipments to a tour takes place by partitioning them into their general direction, and the stop sequence is generated with a TSP that minimizes travel time.

All mentioned models in this section so far only make assumptions about parameters related to tour formation; these models do not include estimated coefficients. Consequently, the empirical and behavioral foundation is limited. Wisetjindawat et al. (2006) assumed travel time minimization as the objective and a maximum working hour limit as a constraint. Boerkamps & van Binsbergen (1999) and Alho et al. (2017) used observed statistics to substantiate assumptions, relating to the maximum number of stops per tour, as one example.

Instead of merely making assumptions about objectives and constraints, You et al. (2016) used inverse optimization to calibrate weights for different objectives of a VRP in a case study of clean drayage trucks in the San Pedro Bay Ports (SPBPs) in Southern California, USA. The following six objectives were assumed to represent trip chaining behavior of these truck operators: (1) maximize the number of visits to the same node in a tour, (2) minimize total travel time, (3) total emissions, (4) total truck operating hours, (5) early arrival times and (6) late arrival times. Because neither time windows nor goal arrival times could be directly obtained from GPS truck diary data, goal arrival times were made endogenous to the objective function, such that they could be estimated simultaneously with early and late arrival penalties, as proposed by Chow & Recker (2012) for the inverse HAPP.

In inverse optimization, the objective is to make a given solution optimal by minimally perturbing prior estimates of parameters. If the problem is a Linear Programming (LP) or 0-1 Integer Programming problem, its inverse problem can be formulated with a similar mathematical structure by making use of the complementary slackness conditions to prove optimality (Ahuja & Orlin, 2001). For Mixed Integer Linear Programming (MILP) problems, such as VRPs and HAPPs, another formulation and a cutting plane algorithm are proposed by Wang (2009). To infer findings from multiple observations, Chow & Recker (2012) propose a Method of Successive Averages (MSA) algorithm, which, although computationally burdensome, is proven to converge. Furthermore, to measure the goodness-of-fit, Chow & Recker (2012) obtain the ratio between the squared error of estimated ODs and arrival times against observed ODs and arrival times with the calibrated model, against the squared error with the model with prior information. The calibrated objective weights can be used in a VRP in an agent-based microsimulation framework to construct tours out of a set of synthetic shipments (You et al., 2016).

Random Utility Modeling

Instead of using optimization techniques from the field of operations research, some disaggregate freight forecasting models use the 'incremental tour growth' approach, first presented by Hunt & Stefan (2007). Conditional probabilities of selecting a location as the next stop and terminating the tour (i.e. returning to depot) are calculated, after which Monte Carlo simulation is used to pick one choice, which leads to an iteratively grown tour. Probabilities are calculated by estimating logit functions based on observed truck tour data.

Hunt & Stefan (2007) pioneered this approach for their tour-based microsimulation framework of commercial vehicle movements in the city of Calgary, Canada. Their model directly estimates vehicle tours and can, therefore, be classified as a tour-based model rather than a hybrid model; it does not explicitly consider the flow of goods between regions. To develop the model, truck trip diary data of only own-account commercial vehicle movements were used. Hunt & Stefan (2007) distinguish three types of movements: external-internal movements (with one trip end outside the study area), fleet-allocator movements (vehicles that need to cover an area or set of links, e.g. newspaper deliveries), and tour-based movements (tours comprised of a small set of individual shipments or services). The incremental tour growth is implemented for the third type of movements. The overall framework consists of the following steps that are iterated until the tour termination decision is made: (1) tour generation, (2) vehicle and tour purpose, (3) tour start, (4) next stop purpose, (5) next stop location, (6) stop duration.

To estimate the number of tours generated per zone, Hunt & Stefan (2007) used an exponential regression equation to estimate the number of tours per employee and establishment category, which was multiplied with the number of employees per establishment category in a zone. A logit model was estimated to determine the time period in which the tour starts, based on accessibility to employment and land use. In the next step, the tour purpose (goods, service, other) and vehicle type (light, medium, heavy) are chosen simultaneously with a logit model, with land use of the zone and company type generating the tour as independent variables. The exact tour start time is chosen with a Monte Carlo simulation on the observed distribution of tour start times in each time period. The next stop purpose has four choice options: goods, service, other, return-to-establishment, and is estimated separately for tours generated by different segments (company type and tour purpose). The choice depends on tour memory variables (e.g. number of goods stops already made in the tour), accessibility of the current location, and the generalized travel (dis)utility of a trip from the current location to the depot. For the next stop location, the choice set for estimation is constructed by the observed next zone and 80 unselected alternative zones randomly identified with stratified sampling based on zone characteristics. Influencing variables include accessibility, land use, travel (dis)utility, and the enclosed angle between the previous stop, the current stop and the (potential) next stop. The stop duration is obtained with a Monte Carlo draw from observed distributions. Finally, parameters were slightly changed to reproduce aggregate statistics relating to zonal tour productions and to the number of stops per tour in different segments.

Wang (2008) and You (2012) mentioned that in the incremental tour growth approach, next stop location and tour termination decisions are being made while performing the tour. This is, of course, a weak assumption, given the tendency of freight operators to make decisions at a more tactical planning level in an optimized way

(You, 2012). However, this approach may rather be seen as a calibrated greedy algorithm to construct tours, rather than a representation of decisions made during tours. Hunt & Stefan (2007) realized this and mentioned that parameters reflect the general structure of tours, rather than decisions being made in the moment. An example Hunt & Stefan (2007) mentioned is that a driver does not decide to go back to the depot because the distance to it from the current location is short, but that tours in general are often constructed to include a stop on the way back to the depot.

Wang (2008) developed a similar yet simpler tour growth approach than Hunt & Stefan (2007). Wang (2008) used this tour growth model as input for a tour-based entropy maximization (as discussed in Section 3.2.1) and for a hybrid microsimulation framework. The two choices in this framework are tour termination and next stop destination, which are represented with an MNL model. The two choice functions were calibrated with a data set of commercial vehicle tours in Denver, CO, USA for the entropy maximization and with a synthetic data set of tours for the hybrid microsimulation framework. For the next stop destination choice, stratified sampling based on distance to the current stop location is used to construct choice sets. The maximum number of stops in a tour is set at 20. In the hybrid microsimulation framework, the tour growth model is used to construct tours that satisfy an OD matrix of flows of a generic commodity in a test network of 84 nodes. For this purpose, the vehicle picks up an average payload of cargo on each OD it visits that has a commodity flow still to be transported, and the vehicle drives empty if no commodity flow is left to be transported for the OD the vehicle travels. The amount of cargo still left to be picked up and delivered at each stop, positively influence the probability of choosing this stop as the next destination. While this model makes heavy assumptions (e.g. one generic commodity type, shipments are always delivered at the next stop of a tour), it does show the possibilities of using the tour growth approach to construct tours that satisfy a given commodity OD matrix.

Some other tour-based models used the incremental tour growth approach in a more similar fashion to Hunt & Stefan (2007). Ferguson et al. (2012) investigated the transferability of the framework of Hunt & Stefan (2007) to the region of Toronto, Canada, by following the same six steps but with some newly estimated parameters based on local data. Comparison of estimated and observed traffic counts yielded promising results, but with an overestimation of medium size vehicle flows in the late morning and early afternoon. Kim et al. (2014) and Kim & Park (2017) developed a similar model for the Seoul metropolitan area in South Korea, and compared it with a traditional trip-based four stage modeling approach. The tour-based model performed marginally better in replicating the observed OD flows, trip length distribution, and average trip length. However, Kim & Park (2017) did not specify whether they separated estimation and validation data, nor does the quantitative comparison provide information about the models' abilities to predict future impacts.

Kuppam et al. (2014) extended the tour growth framework for their tour-based model of Phoenix, AZ, USA to include tours that do not end at their starting locations (i.e. paths). They also use GPS truck data instead of truck trip diary data. Furthermore, their model consists of a different set of choices than the model of Hunt & Stefan (2007): (1) tour generation, (2) stop frequency, (3) tour completion, (4) stop purpose, (5) stop location, and (6) time of day. Step 2 is an MNL model that determines the number of stops in the tour, based on accessibility and land use of the origin zone, constrained to eleven stops as very few tours had been observed containing more stops. Step 3 is Binomial Logit (BL) model that determines whether the tour ends at its starting location or not. Only eight percent of observed tours started and ended at the same location. The stop purpose (industry type) of each stop is chosen sequentially with an MNL, after which the location of each stop is chosen based on the travel time to its previous stop and depot and accessibility to employment. The time of day choice is estimated for each stop in the tour with an MNL model, the alternatives are the 24 hours of the day. A separate choice model is estimated for the first stop and subsequent stops, as for the latter the influence or earlier stops needs to be incorporated. The model replicates observed statistics quite well, yet it is not specified whether the same data is used for calibration and for estimation. Doustmohammadi et al. (2016b) developed a similar model for Birmingham, AL, USA and showed that their tour-based model performs well in replicating observed traffic count data.

Ruan et al. (2012), Zhou et al. (2014), and Khan & Machemehl (2017) used the Random Utility Modeling framework to model the choice for the number of trips in a tour based on commercial vehicle survey data from Texas, USA. Ruan et al. (2012) and Khan & Machemehl (2017) also included the daily tour chaining strategy of a vehicle in the choice, which can be a single or multiple tours on a day. Ruan et al. (2012) estimated a mixed logit, nested logit, and multinomial logit model, while Khan & Machemehl (2017) adopted a discrete-continuous extreme value modeling approach. Ruan et al. (2012) found that direct tours (one stop) are more likely to be made for shipments sent over a longer distance and for construction materials, while tours with more stops are made more often for tours to/from distribution centers and transporting food, health, and beauty products. In the model of Khan & Machemehl (2017) tours that include intermediate stops made at retail centers are also more likely to have more stops. While these models provide a deeper understanding of factors influencing the number of stops and tour chaining strategy, they cannot directly handle transformation of shipments into tours.

For this purpose, Outwater et al. (2013) developed a tour formation framework for Chicago, IL, USA, around a similar Multinomial Logit (MNL) choice model that assigns to each shipment the tour pattern (one tour, two tours, or three tours per day) and number of stops of the tour it will be part of. The same data as Ruan et al. (2012) is used for estimation. A hierarchical clustering is used to group nearby shipments with the same tour pattern and number of stops, and a greedy 'closest-node-next' algorithm constructs the tour sequence. Only tours with one loading point, an identified warehouse, are modeled. These tours can visit multiple stops for unloading.

Nuzzolo et al. (2012) developed a tour formation model for Rome, Italy, in order to convert a shipment OD matrix into a vehicle OD matrix. For this purpose, they estimate an MNL model to represent the choice for trip chain order (1, 2, 3, 3+ stops in the tour), and for next stop location. The trip chain order model shows that tours constructed by third-party carriers and for foodstuffs are more likely to include multiple stops. A higher accessibility to retailers in the depot zone leads to fewer stops in the tour, which Nuzzolo et al. (2012) explain by the desire to reduce the complexity of operations. The choice set of the next stop location model includes all zones less than 25 km away from the current stop, as very few observed trips are longer. Separate models were estimated for direct tours and multiple-stop tours, and for the first stop and later stops in a tour. Significant variables include tour memory variables, and the retailer and wholesaler accessibility in the destination zone. Choice models are estimated based on only 500 interviews with truck drivers and the scope is limited to restocking tours performed for and by retailers. The model was not validated on observed statistics.

Other approaches from the Household Activity Pattern Problem literature

The literature on activity-based modeling of personal transportation also provides techniques that may be used in freight tour formation. More specifically, the HAPP, is a MILP that is very similar to the VRP. The HAPP is used to represent the household interaction that leads to the allocation of time and vehicles to travel and activities, taking into account spatio-temporal constraints (Chow & Recker, 2012). The calibration of parameters of its objective function remains a challenge. Recker et al. (2008) mention five reasons for this: (1) the set of alternatives is infinitely large, (2) the solution vector includes both integer and continuous variables, (3) alternative household patterns are mutually exclusive, but parts of the solution vector may not be, (4) the components of the utility function cannot be directly interpreted as relative utility weights, and (5) closed-form probabilities can often not be achieved due to the complex constraint space. Previously in this section, we discussed inverse optimization, which has been used by Chow & Recker (2012) to calibrate the HAPP, and later by You et al. (2016) for the VRPTW. In this section, we discuss two other approaches that have been used to estimate parameters of the HAPP, and may similarly be used to estimate parameters of a VRP.

Firstly, Recker et al. (2008) used a genetic algorithm to work iteratively towards a set of parameters of the HAPP objective function with a minimal difference between its predicted household pattern and the observed household pattern. The general philosophy of this approach is as follows: (1) a population of strings with different objective function parameters is created; (2) the fitness is evaluated by comparing the predicted pattern with the observed pattern; (3) strings with a higher fitness receive a higher probability of reproduction; (4) reproduction takes place by crossover of existing strings and mutation; (5) this process is iterated until a set convergence criterion is reached. Recker et al. (2012) demonstrate their approach both for estimation of parameters of an individual household and of a sample of 65 households. Two weaknesses in this approach are the lack of statistically sound measures of goodness-of-fit and the computationally cumbersome estimation, similar to the inverse optimization technique of Chow & Recker (2012) and You et al. (2016).

Xu et al. (2017) provided an alternative approach for HAPP estimation. Instead of genetic algorithms or inverse optimization, they make use of the RUM framework. Statistically sound measures are available to judge the model fit and inclusion of parameters (Xu et al., 2017). Similarly to what Adler & Ben-Akiva (1979) did in the earliest stages of activity-based modeling, household patterns of other households are sampled to construct a discrete choice set, from which utility weights can be obtained using maximum likelihood estimation. Xu et al. (2017) further formalized this procedure in the following three step: (1) choice set generation, (2) choice set individualization, and (3) parameter estimation. The choice set generation is improved by Xu et al. (2017) by sampling those household patterns that maximize the information gain of the choice set. In choice set individualization, goal-programming is used to adjust alternative household patterns that do not comply with constraints of an individual household. Another improvement is that a Path-size Logit (PL) model is estimated, instead of an MNL. This means that a correction term is added to each alternative that accounts for overlap with other alternatives.

3.3 The scientific gap

Clearly, in the last ten to fifteen years, freight modelers have increasingly realized the importance of incorporating tour formation. While this is a positive development for the forecasting potential of these models, challenges and gaps in the proposed approaches still remain. In this section, we define the gap that our research fills.

Tour-based entropy maximization is an interesting concept that is able to connect microscopic tour formation models with aggregate regional freight forecasting models. However, these models answer the question regarding how to connect microscopic tours with traffic counts and a macroscopic model but do not provide answers regarding the behavioral formation of tours out of shipments, the objective of this research.

The disaggregate models that we have discussed in this chapter come closer to answering this question, yet many models do not consider commodity flows explicitly. The driver behind freight transportation, trade of goods between firms/regions, is not modeled explicitly (Wisetjindawat et al., 2006). As a consequence, the different economic characteristics, constraints (Holguín-Veras et al., 2014), and geographical distribution of these commodity types cannot be considered. For example, Hunt & Stefan (2007) directly estimate the number of tours originating in each zone and You et al. (2016) work with a set of nodes instead of goods that need to be visited on a day, which obstructs the consideration of vehicle capacity constraints.

Preferably we do not merely model commodity flows as kilograms between regions but represent distinct shipments explicitly. Many decisions in freight transportation are made at the shipment level (de Bok et al., 2018). Consequently, modeling shipments allows a more accurate representation of these decisions. In addition, shipment-based models allow for analysis of far more specific policies and scenarios, such as: new distribution centers, changes in delivery frequencies and shipment sizes, and increased cooperation of shippers and carriers (Boerkamps & van Binsbergen, 1999). While Wang (2008) assigns commodity flows to vehicle tours, no set of shipments is modeled.

In behavioral freight tour formation modeling we aim to reproduce observed tour formation behavior. Ideally, parameters of a statistical model are calibrated on empirical data. Here we define calibration as tweaking model parameters to reproduce empirical data. A statistical model provides inferential statistics about the population and accounts for correlations between predictors. Calibration of a statistical model provides a strong behavioral foundation and validity.

As the required data for disaggregate agent-based models is often hard to obtain (de Bok & Tavasszy, 2018) and tour formation cannot be narrowed down to a single choice, few studies calibrate their tour formation model on empirical data in a statistical way. VRPs (e.g. Wisetjindawat et al., 2006; Anand et al., 2014) or heuristic algorithms (e.g. Boerkamps & van Binsbergen, 1999; Alho et al., 2017) without calibrated parameters are often used instead. You et al. (2016) calibrate a VRP with a mathematical optimization approach, which does not only lack inferential statistics and controlling between parameters but is also computationally too heavy for application in a regional forecasting model.

Nuzzolo et al. (2012) and Outwater et al. (2013) developed the only tour formation models we found that are shipment-based and statistically calibrated. Nuzzolo et al. (2012) limit their scope to restocking tours made for and by retailers in Rome, Italy, while Outwater et al. (2013) only model tours that distribute food and manufactured goods from a warehouse or distribution center in Chicago, IL, USA.

A methodological weakness can be identified in these two similar tour formation models too. For each shipment of a carrier, Nuzzolo et al. (2012) and Outwater et al. (2013) choose the number of stops of the tour that this shipment will be part of. The number of stops per tour and the daily number of tours are determined before the choice is made which shipments are transported in the same tour, while in reality the number of stops and tours are outcomes of the process of allocating shipments to tours. As a consequence, important operational factors that actually impact the number of stops in a tour are not or very implicitly included. For example, in the model of Nuzzolo et al. (2012) the number of stops is not constrained explicitly by the vehicle capacity and not chosen in coordination with other shipments. In this model, the number of stops does not depend on whether a carrier has other shipments that can be transported in the same tour with little additional time. A model where the number of stops is an outcome instead of a decision made before the tour formation process is desired.

When we look at Table 3.1, we see that none of the identified tour formation models satisfy all requirements we described so far in this section: (1) shipment-based, (2) statistically calibrated, and (3) the number of stops is an outcome of the tour formation process instead of a decision made in advance. Our study fills this gap.

The data that we use covers all road freight transportation in the Netherlands and further distinguishes our study from others. Firstly, we do not limit our tour formation model to restocking tours of retailers (Nuzzolo et al., 2012) or tours that distribute goods from a warehouse (Outwater et al., 2013). Secondly, most tour formation models only model and/or use data about urban trips, while we also model trips between cities. Thirdly, our study is highly inclusive with respect to goods types. Boerkamps & van Binsbergen (1999) only model shipments related to books and food retail, while Outwater et al. (2013) developed their model for food and manufactured products. Finally, few tour formation models are developed in Europe and only two are identified in the Netherlands, which both do not have calibrated parameters. Because strong regional differences exist in commercial vehicle patterns, a model developed for one region (or country) is not applicable to the other (Zhou et al., 2014). For example, in the Netherlands cities tend to be much more dense but also closer to each other than in the USA, which is why tours might be able to visit multiple cities more often in the Netherlands.

To summarize, no tour formation model was found to satisfy the following three criteria:

- Shipment-based
- Statistical calibration of parameters
- The number of stops per tour and the number of tours per day is an outcome of the process of allocating shipments to tours

In addition, many studies:

- Make selections on the types of locations or establishments where goods are loaded or unloaded
- Make selections on goods types
- Do not model inter-urban tours
- Do not develop their model for the Dutch context

Table 3.1. Summary of literature review on behavioral freight models that include tour formation.

Reference	Tour model type	Study region	Geographical scope	Commodity-based	Shipment-based	Commodity types	Carrier types *(receiver/sender types)*	Calibrated tour model	Statistical tour model	Number of stops is outcome of tour formation
Adler & Ben-Akiva (1979)	RUM	Washington DC, **USA**	-	-	-	-	-	Y	Y	Y
Boerkamps & van Binsbergen (1999)	Algorithm	Groningen, **the Netherlands**	Urban	Y	Y	Food retail, bookstores	All	N	N	Y
Taniguchi & van der Heijden (2000)	VRP	*hypothetical network*	-	Y	Y	All	All	N	N	Y
Wisetjindawat et al. (2006)	VRP	Tokyo, **Japan**	Urban	Y	Y	All	All	N	N	Y
Hunt & Stefan (2007)	RUM	Calgary, **Canada**	Urban	N	N	All	Own-account	Y	Y	Y
Recker et al. (2008)	HAPP (VRP)	Washington & Oregon, **USA**	-	-	-	-	-	Y	N	Y
Wang (2008)	RUM	Denver, CO, **USA** & *hypothetical network*	-	Y	N	All *(generic commodity type)*	All	Y	Y	Y
Wang & Holguín-Veras (2009)	Entropy maximization	Denver, CO, **USA**	Urban	N	N	All	All	Y	N	-
Donnelly et al. (2010)	Algorithm & optimization	Oregon, **USA**	Urban and inter-urban	Y	Y	All	All	N	N	Y
Polimeni et al. (2010)	VRP	Palermo, **Italy**	Urban	Y	Y	All	All *(to/from retailers)*	N	N	Y
Chow & Recker (2012)	VRP	Orange County, CA, **USA**	Urban	-	-	-	-	Y	N	Y
Ferguson et al. (2012)	RUM	Toronto & Hamilton, **Canada**	Urban	N	N	All	All	Y	Y	Y
Nuzzolo et al. (2012)	RUM	Rome, **Italy**	Urban	Y	Y	All	All *(to/from retailers)*	Y	Y	N
Ruan et al. (2012)	RUM	Texas, **USA**	Urban	N	N	All	All	Y	Y	N
You (2012)	Entropy maximization	Southern California, **USA**	Urban	N	N	All	Third-party *(to/from port)*	Y	N	-
Outwater et al. (2013)	RUM	Chicago, IL, **USA**	Urban	Y	Y	Food products, manufactured products	All *(from warehouse/DC)*	Y	Y	N
Kuppam et al.	RUM	Phoenix, AZ,	Urban	N	N	All	All	Y	Y	N

(2014)		**USA**								
Kim et al. (2014) Kim & Park (2017)	RUM	Seoul, **South Korea**	Urban	N	N	All	Third-party couriers	Y	Y	Y
Anand et al. (2014)	VRP	*hypothetical network*	Urban	Y	Y	All	Third-party *(to/from retailers)*	N	N	Y
Zhou et al. (2014)	RUM	Texas, **USA**	Urban	N	N	All	All	Y	Y	N
Sánchez-Díaz et al. (2015)	Entropy maximization	Denver, CO, **USA**	Urban	N	N	All	All	Y	N	-
Irannezhad & Hickman (2016)	VRP	Brisbane, **Australia**	Urban	Y	Y	All	Third-party	N	N	Y
You et al. (2016)	VRP	Southern California, **USA**	Urban	N	N	All	Third-party *(to/from port)*	Y	N	Y
Doustmohammadi et al. (2016b)	RUM incremental	Birmingham, AL, **USA**	Urban	N	N	All	All	Y	Y	N
Alho et al. (2017)	Algorithm & optimization	**Singapore**	Urban	Y	Y	All	All	N	N	Y
Khan & Machemehl (2017)	RUM	Texas, **USA**	Urban	N	N	All	All	Y	Y	N
Xu et al. (2017)	RUM + HAPP (VRP)	California, **USA**	-	-	-	-	-	Y	Y	Y
BasGoed (see Groot & Miete, 2016; Significance, 2018)	Multiplication of observed trips	**The Netherlands**	Urban and inter-urban	Y	N	All	All	N	N	-
This research	RUM	**The Netherlands**	Urban and inter-urban	Y	Y	All	Third-party	Y	Y	Y

4. DATA

For this research, we have access to the XML microdata that is collected by the Central Bureau of Statistics of the Netherlands (CBS). The objective of this chapter is twofold: (1) understand this data and its usefulness for estimation of a tour formation model, (2) empirically explore other factors that can explain differences in tour formation. For this purpose, in Section 4.1 we discuss the general structure and variables in this data set. In Section 4.2 we discuss another source of data, skim matrices that provide distances and travel times between zones in the Netherlands. In Section 4.3 we go into further detail explaining relevant variables for tour formation and report descriptive statistics, after which we conclude with the implications of the findings of this chapter for model estimation in Section 4.4. We begin by providing background information about the data set.

CBS collects this XML microdata as one of the sources for their 'Basisbestanden Goederenwegvervoer' (BGW). This is a set of three files (at the level of tours, trips, and shipments) that CBS collects for Rijkswaterstaat, the executive body of the Dutch Ministry of Infrastructure and the Environment, who uses this data to develop models to evaluate new policies on the subject of freight road transportation (CBS, 2017b). CBS collects similar data sets for freight transportation by rail and over inland waterways.

The population for the BGW is the complete commercial vehicle fleet of the Netherlands and foreign trucks driving on the Dutch roads. CBS draws a sample from these vehicles based on 74 strata, including operator type, carrying capacity, vehicle age, vehicle type, and the vehicle fleet of its owner (Robroeks, 2016; CBS, 2017b). Companies are obligated to fill in the survey for all shipments transported by the sampled vehicle for a specified week and CBS calls companies when they have forgotten to comply (Robroeks, 2016). The chance that a vehicle is drawn for the sample in a year is approximately 1:3 (Robroeks, 2016).

CBS allows companies to fill out the survey in three different ways: internet, XML, and paper (Robroeks, 2016; CBS, 2017b). In this research, the XML data is made available, and we will only go into further detail explaining this data. The XML data is directly tapped from the planning software of a carrier. Companies can implement such XML collection extensions themselves, but CBS always has to perform a quality control before data can be provided in XML format (CBS, 2008). Approximately 80% of the XML data originates from real-time systems, and 20% from Transport Management Systems (TMS) (Robroeks, 2016). In most cases, these systems require the driver to interact with it in order to fill in the survey in a partially automated way (Robroeks, 2016). CBS specifies to those companies that have decided to install XML data collection tools into their software and trucks what the definitions are of terms such as 'tour' and 'shipment'. CBS realizes, however, that some companies have their own definitions, which can sometimes lead to inconsistent data. Furthermore, some companies provide data of all trucks for requested weeks, while some only provide those vehicles that CBS has requested, but it is not known which companies do so.

4.1 STRUCTURE OF THE DATA

The XML data that is available for this study consists of approximately 2.6 million records, stored in csv-format. The data has a relational structure, with three layers of information: (1) truck-weeks, (2) tours, and (3) shipments, each with its own IDs and variables (Robroeks, 2016). Two one-to-many relationships can be distinguished: one truck-week can consist of multiple tours, and one tour can consist of multiple shipments (Robroeks, 2016). The data is organized in such a way that each record is a shipment. For this reason, if a tour contains multiple shipments, information at the tour level is repeated for each shipment. In Table 4.1, this structure is shown with a pseudo data example.

At each level, multiple variables are present that relate to the object that this level represents (i.e. information about a shipment, a tour, or a vehicle-week). Figure 4.1 summarizes which information is reported in the data at each level. In the next section, we go into further detail explaining how the variables that are relevant for tour formation are measured, what their underlying assumptions are, and how their values are distributed in the sample. Most tour variables have been obtained with transformations using the variables in this data.

Table 4.1. Pseudo data. The structure of the data with three different layers (differently shaded text) and several accompanying variables is shown.

Truck-week ID	Week	Tour ID	Departure town	Departure time	Shipment ID	Loading location	Shipment distance
4510	1	30100	Amsterdam	1-1-15 8:15:00	80430	Amsterdam	50
4510	1	30105	Rotterdam	6-1-15 10:30:00	80431	Zwolle	13
4510	1	30105	Rotterdam	6-1-15 10:30:00	80432	Rotterdam	54
4510	1	30105	Rotterdam	6-1-15 10:30:00	80433	Gouda	28
4511	2	30107	Utrecht	1-1-15 6:00:00	80434	Utrecht	260

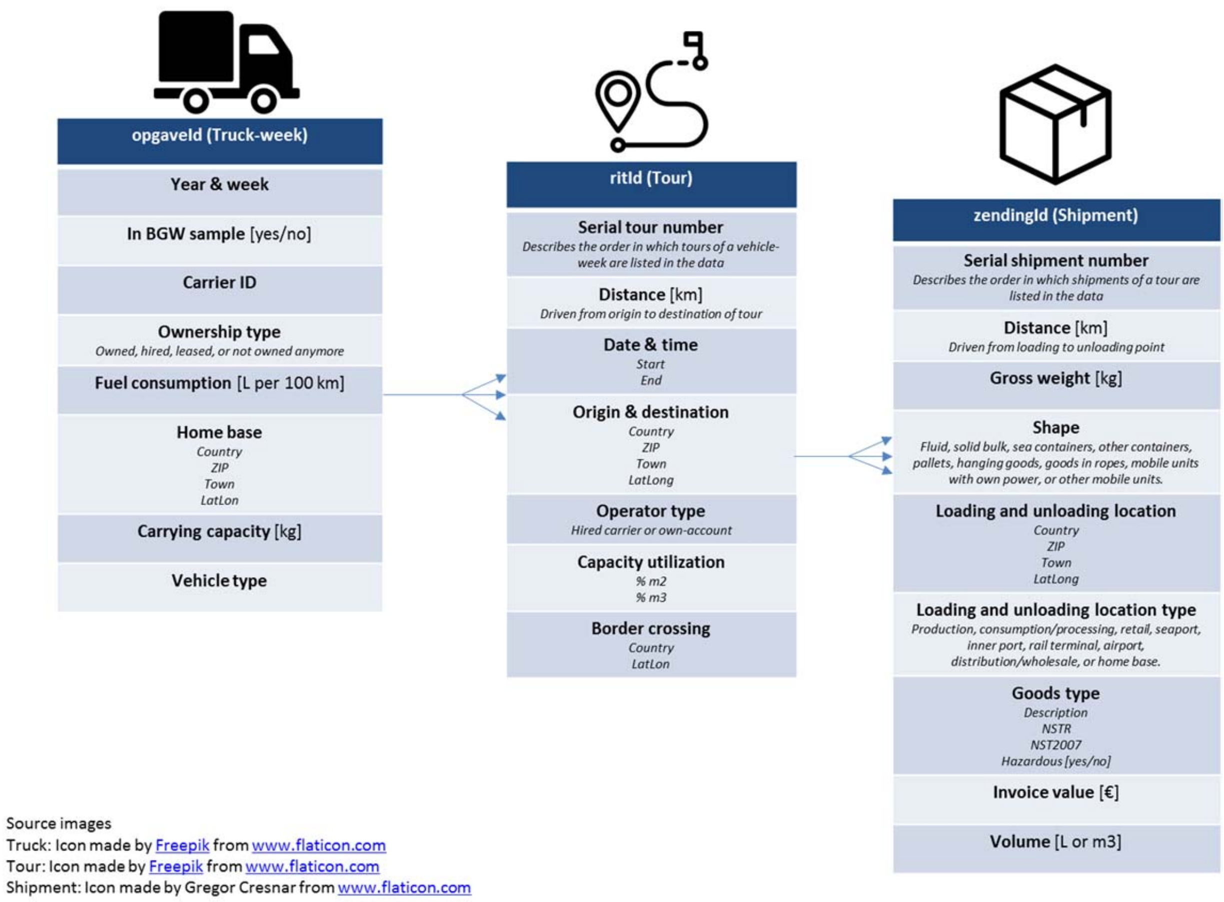

Figure 4.1. The available data at each layer of information.

4.2 Skim matrices

Besides the XML microdata, we also have access to skim matrices in this research. These skim matrices provide a travel impedance (travel time and distance) between all combinations of zones in the Netherlands. Zones are at the level of 'buurten', an administrative unit with a level of aggregation in between 4-character and 6-character ZIP codes in the Netherlands. The Netherlands consists of approximately 12,000 buurten. Separate skim matrices are available for the AM peak, PM peak, and the remainder of the day.

Road link impedances are used to find the shortest path between each combination of zones. These road link impedances originate from a traffic assignment in the calibrated NRM-West (Nederlands Regionaal Model), the Dutch transportation model, with most detail in the western region.

Information about these buurten are made publicly available by the CBS (Kerncijfers wijken en buurten 2015). These zonal characteristics relating to land use and degree of urbanization are used to enrich the XML microdata.

4.3 Analysis and discussion of variables

In this section, we discuss some of the relevant variables in the XML data set for this research and other variables that have been added to this data. The definitions and distributions of some of these variables are discussed. Variables at the level of shipments and tours are discussed respectively.

When analyzing this data, we have selected only the records that are part of tours which include the following:

- All loading and unloading locations can be obtained in the form of a buurt zone
- The goods type of all shipments is filled in (NSTR 1-digit classification)
- The gross weight of all shipments is filled in and non-zero
- The carrying capacity of the vehicle is filled in and non-zero

This leaves approximately 600,000 records of the roughly 2,600,000 records in total. These filters are applied because these variables need to be known in order to obtain the explanatory variables of the tour formation model and to be able to arrive at a consistent definition of a shipment in the data set (see Section 4.3.1).

The buurt zones of the loading and unloading location of each shipment can only be obtained for correctly filled out Dutch 4-character or 6-character ZIP codes. For this reason, the filtered data set only includes tours that do not cross the border.

Of all tours in the filtered data set, 54.1% were made in 2014, while the rest were made in 2013 or 2015. All records in the filtered data set were filled out in XML format. Since this requires advanced transportation planning systems, a self-selection of more advanced carriers has taken place. For this reason, almost all tours in the data are reported by third-party carriers.

4.3.1 Shipment attributes

The CBS guideline for distinguishing shipments requires the combination of loading location, unloading location, and goods type of the shipment to be unique in the tour. Companies are not consistent in this definition in their planning systems, however. Some provide far more detailed descriptions of the goods than others. For this reason, we sometimes see tours with dozens of shipments between the same loading and unloading location. Table 4.2 provides a pseudo data example to show this.

Additionally, only direct shipments are reported in the data (i.e. one loading and unloading location per shipment). We do not know whether the shipment was part of a complex logistics chain with multiple intermediate storage/transshipment locations.

To arrive at a more consistent data set regarding the definition of a shipment, shipments in the same tour with the same loading location, unloading location and goods type (which we operationalize with the NSTR 1-digit classification) are aggregated into one shipment. The gross weight of these aggregated shipments is the sum of the gross weights of each individual shipment that is aggregated. Consistency in this shipment definition is important when analyzing data statistics, especially when parameters of a tour formation model are estimated on it.

Table 4.2. Pseudo data. Some companies distinguish many similar shipments in a tour. In this example, five shipments are distinguished between the same locations, all with fruits/vegetables.

Tour ID	Shipment ID	Loading location	Unloading location	Goods type
4610	40101	Amsterdam	Rotterdam	Apples
4610	40102	Amsterdam	Rotterdam	Pears
4610	40103	Amsterdam	Rotterdam	Peaches
4610	40104	Amsterdam	Rotterdam	Pineapples
4610	40105	Amsterdam	Rotterdam	Tomatoes
4610	40106	Amsterdam	Utrecht	Tomatoes
4610	40107	Amsterdam	Gouda	Tomatoes

Serial shipment number

Each shipment has a serial shipment number, which denotes the order in which the carrier has listed the shipments in a tour. While CBS does not provide a guideline about what carriers should fill in for this variable, it appears that carriers usually fill this in the order that shipments were loaded, since the first listed shipment is usually loaded at the tour starting location.

Loading and unloading location

The loading location and unloading location of each shipment can be filled in with the following levels of detail:

- Country
- Town
- ZIP code (4-digit or 6-digit)
- Coordinates

For this research, we have decided to use the ZIP code variables. These are filled out much more often than coordinates. These can also rather easily be converted to buurten, which provide a workable skim matrix to obtain tour characteristics to estimate parameters of the tour formation model.

Goods type

The goods type of a shipment is filled out with the NSTR classification, NST2007 classification, and with a textual description. The NSTR code is used in this research, for it is filled out much more often than the NST2007, and the textual description provides too much detail to make general statements about goods types of shipments. At the 1-digit level, the following NSTR categories can be distinguished:

- 0: Agricultural products and livestock
- 1: Other foodstuffs and fodder
- 2: Solid mineral fuels
- 3: Petroleum and petroleum products
- 4: Ores, metal waste, and iron pyrites
- 5: Iron, steel, and non-ferrous metals (incl. semi-finished products)
- 6: Raw minerals and construction material
- 7: Manure/fertilizers
- 8: Chemical products
- 9: Vehicles, machines, and other goods

The distribution of the goods type of shipments is shown in Figure 4.2.

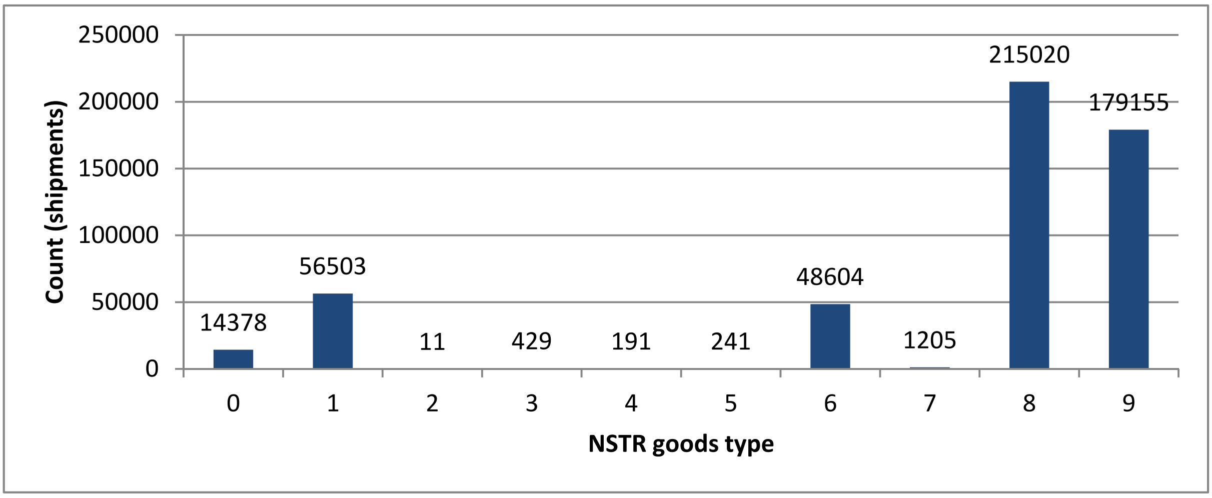

Figure 4.2. Goods type distribution of shipments (NSTR 1-digit).

A large part of the data consists of shipments of NSTR category 8. Further analyses showed that a striking portion of the shipments in this category (approximately 180,000 shipments) are concrete. Many shipments of goods type NSTR 9 are found in the data as well, although this may be mostly due to the fact that it includes other goods as a remainder category.

4.3.2 Tour attributes

CBS defines a tour as any journey made by a vehicle, which starts at the location where the first shipment was loaded into the vehicle, and which ends at the location where the last shipment was unloaded from the vehicle or when the vehicle returns to its home base.

For each tour, we know which shipments it contains. Geographically, we know the tour starting location, the tour end location, and the loading and unloading location of each shipment in the tour. We do not know, however, what the order of visiting the loading and unloading locations of a tour is. In Figure 4.3 it is graphically shown how a tour is represented in the data.

The definition of a tour in the data set has the following two consequences: (1) empty trips are not included (Figure 4.4), and (2) we cannot state with certainty whether an empty trip starts at the home base or at the end location of the previous tour (Figure 4.5). Empty trips are not in the data, since the tour does not start until the first shipment is loaded into the vehicle, and it ends when the last shipment is unloaded. Before 2010, CBS asked for these empty trips in their surveys but has stopped doing so to reduce the time and effort it takes for respondents to fill in the survey (CBS, 2017b). Since a tour ends when the last shipment is unloaded, its driver may very well have directly driven to the starting location of the next reported tour after it, instead of driving back to the home base, which makes obtaining objective information on these empty trips unfeasible (CBS, 2017b).

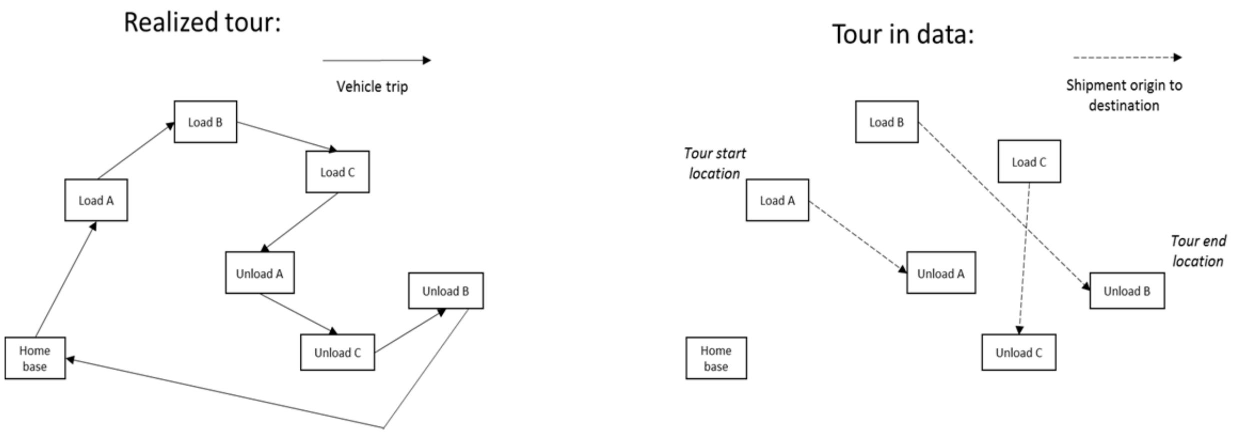

Figure 4.3. Left: a hypothetical example of a performed tour. Right: how this tour is represented in the XML-data.

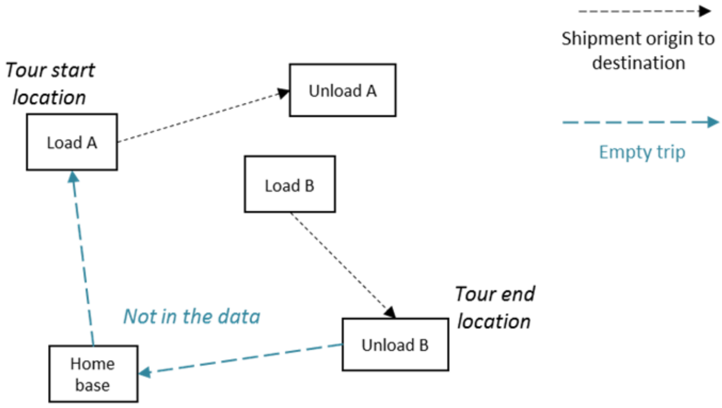

Figure 4.4. Empty trips in a tour are not included in the data.

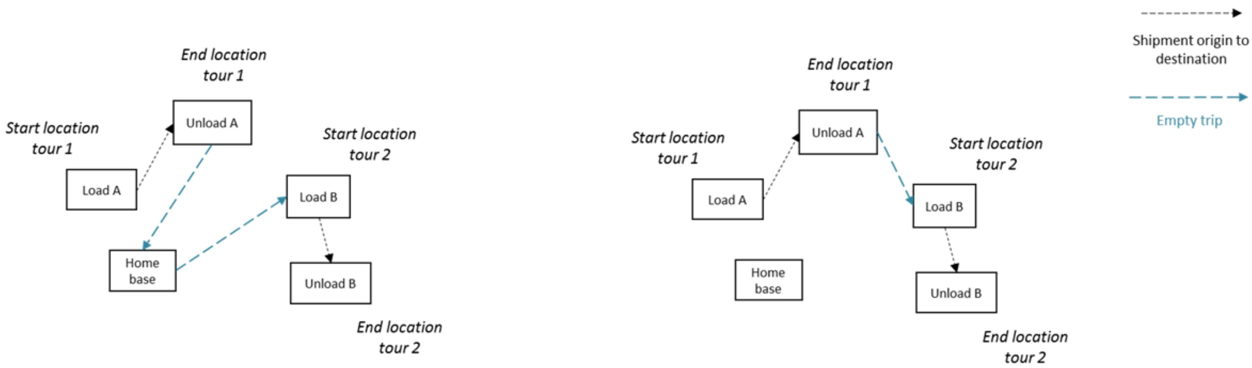

Figure 4.5. We do not know for certain whether an empty trip started at the home base of a vehicle (left) or whether the vehicle was driven from the end location of the previous tour to the starting location of the current tour (right).

Based on the distance of the home base to the tour starting location, it is possible to make assumptions about whether an empty start trip originates from the home base or from the end location of the previous tour. If the distance to the home base is much further away, then it more likely that the empty start trip originates from the end location of the previous tour. This is what CBS does to generate empty trips when XML data is transformed for the BGW data set (CBS, 2017b). However, in the filtered data set, a home base location ZIP is filled out for only 15% of the tours.

NUMBER OF SHIPMENTS

Since the objective of this research is to develop a tour formation model that allocates shipments to tours, it is highly valuable to know how many shipments are usually allocated to tours in reality. In the data, we see that the vast majority (91.5%) of the tours include only one shipment (Figure 4.6). Apparently, there is a strong tendency to complete the tour after the first shipment. Several constraints and preferences can explain this tendency. It may be due to the desire to reduce the complexity of operations (Nuzzolo et al., 2012) but the vehicle capacity may also already have been reached with one shipment. The definition of a tour in the data set impacts this distribution. If a vehicle transports multiple shipments on a day in such a sequence that it turns empty in between each shipment, all these shipments are listed as separate tours. Remarkably, tours that transport concrete virtually never include more than one shipment. Concrete is a highly time-sensitive product, it is usually transported in large volumes, and cannot be combined with other goods types; therefore, a strategy of combining shipments in a tour is usually not feasible for concrete (Khan & Machemehl, 2017).

Another important piece of information illustrated by Figure 4.6, is the fact that very rarely are tours with more than fifteen shipments observed.

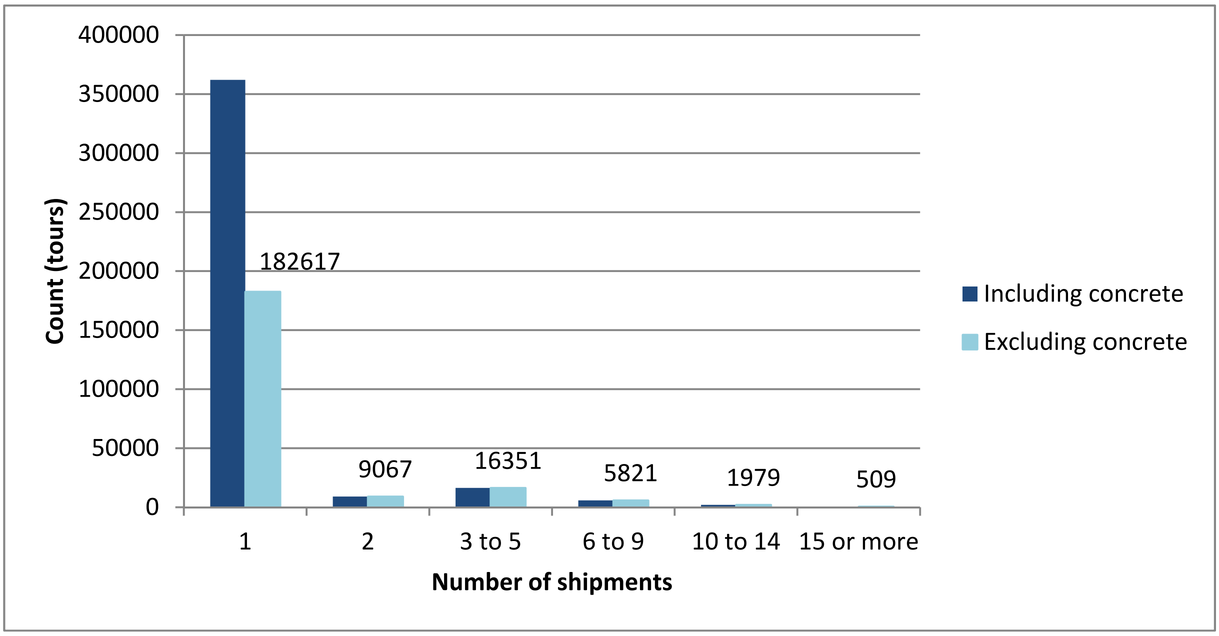

Figure 4.6. Distribution of the number of shipments per tour.

NUMBER OF STOP LOCATIONS

The number of stop locations has been obtained by counting the number of unique loading and unloading locations (at buurtcode level) in each tour. In contrast to a shipment, a stop location is a single unique location, while a shipment contains both a loading and an unloading location (Figure 4.7). Interestingly, the number of stop locations shows a distribution highly similar to that of the number of shipments (Figure 4.8). A shipment can add zero, one, or two new stop locations to a tour. On average a shipment adds approximately one stop to a tour.

Again, we see that only very rarely tours with more than fifteen stop locations occur. Such tours become too complex, last too long, and leave no more remaining capacity for more shipments. Concrete shipments, however, are only transported in tours with one or two stops.

20,1% of the tours visit only one unique location. In that case, either the loading and unloading location of a shipments are located so close to each other that they fall under the same buurtcode, or an error has been made in filling in loading and unloading locations.

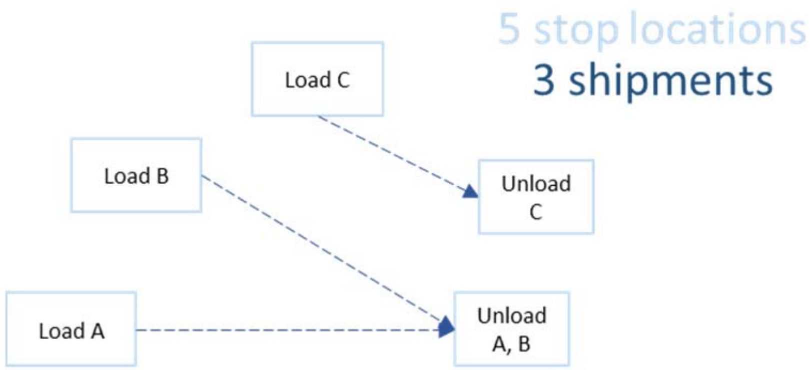

Figure 4.7. The difference between a stop location and a shipment, exemplified with a tour structure.

Most previous studies observed more stops per tour. For example, Figliozzi et al. (2007) found an average of 6.8 stops per tour in their analysis of tours made by a freight forwarding company in Sydney, Australia, and mentioned that previous studies in Amsterdam, Calgary, and Denver had similar findings. The different tour definition in the XML microdata and the large portion of concrete shipments may explain this difference.

In Section 2.3.2, we discussed the literature that suggests that third-party carriers are more likely to construct tours with many stops, as they have more shipments to combine. However, the transportation planner of Rensa BV, an own-account carrier, more often constructs tours with a number of stops (30-35) that we only rarely observe in the XML data set, which consists of virtually only third-party carriers. Perhaps we need to readjust this hypothesis from the literature, at least to address the situation in the Netherlands. For an own-account carrier like Rensa BV, the vehicle fleet size is a determining constraint in tour formation. They aim to fill their vehicles with as many shipments as possible. For large third-party carriers the vehicle fleet size may not be as constraining, which can explain why they do not appear to construct tours with this many shipments/stops and, therefore, why relatively few stops are made in the XML data set.

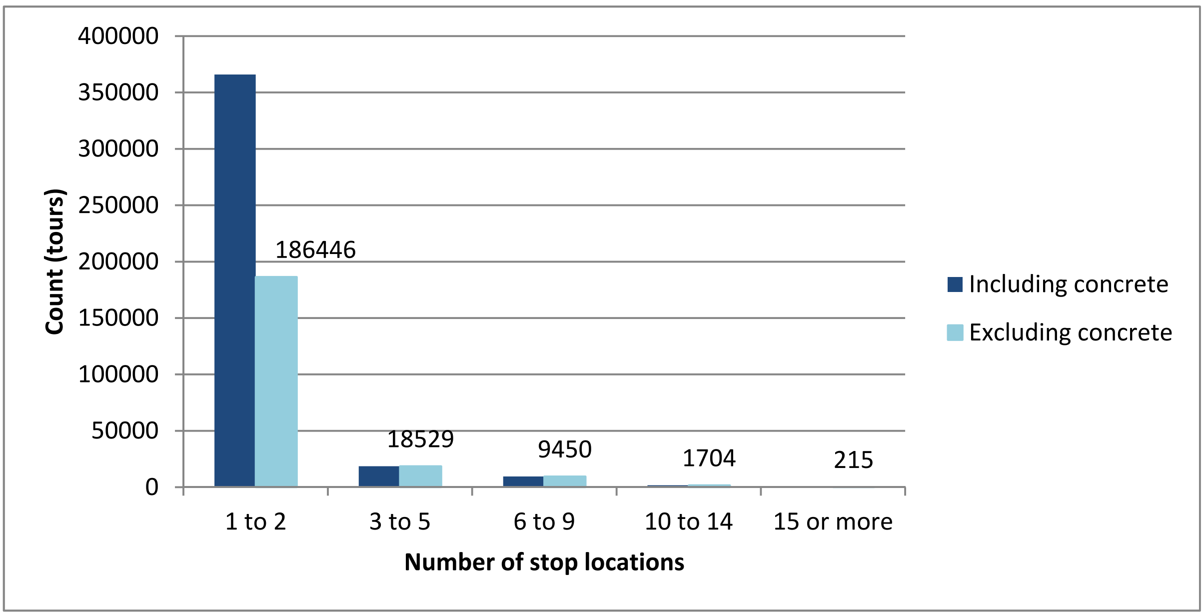

Figure 4.8. Distribution of the number of stop locations per tour.

In Figure 4.9, we see that the number of shipments and number of stops in a tour are very closely related, with both on average having nearly the same value in a tour. From this we can conclude that a shipment does not usually add two stops to a tour but more often only one. Shipments that have a location that is already part of the tour are more likely to be added to it.

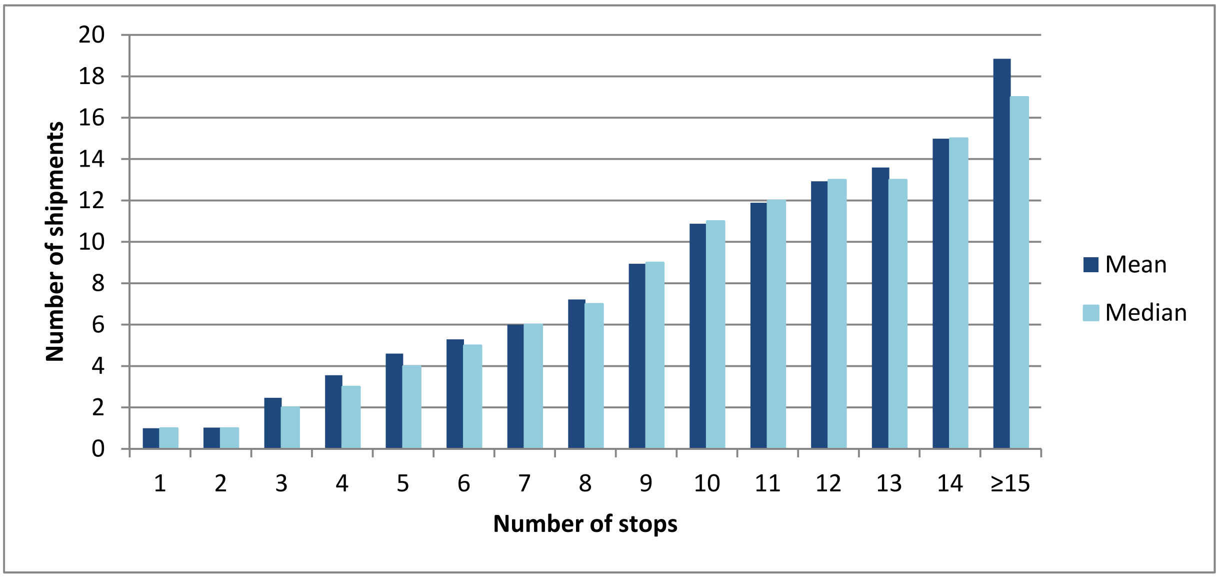

Figure 4.9. The mean and median number of shipments set out against the number of stops.

TOUR TYPE

The tour type, as explained in Section 2.3.1, provides information about the structure of tours. Since the vast majority of tours includes only one shipment (and therefore two stops), most tours can be identified as a direct tour. Of those tours that include more than two stops, most are mixed tours. Still, a notable portion of multiple stop tours are in the form of a collection or a distribution tour. Collection and distribution tours are less complex than mixed tours and are, therefore, easier to understand and plan. If there is only one loading location, a tendency to keep this the only loading location in the tour can be identified.

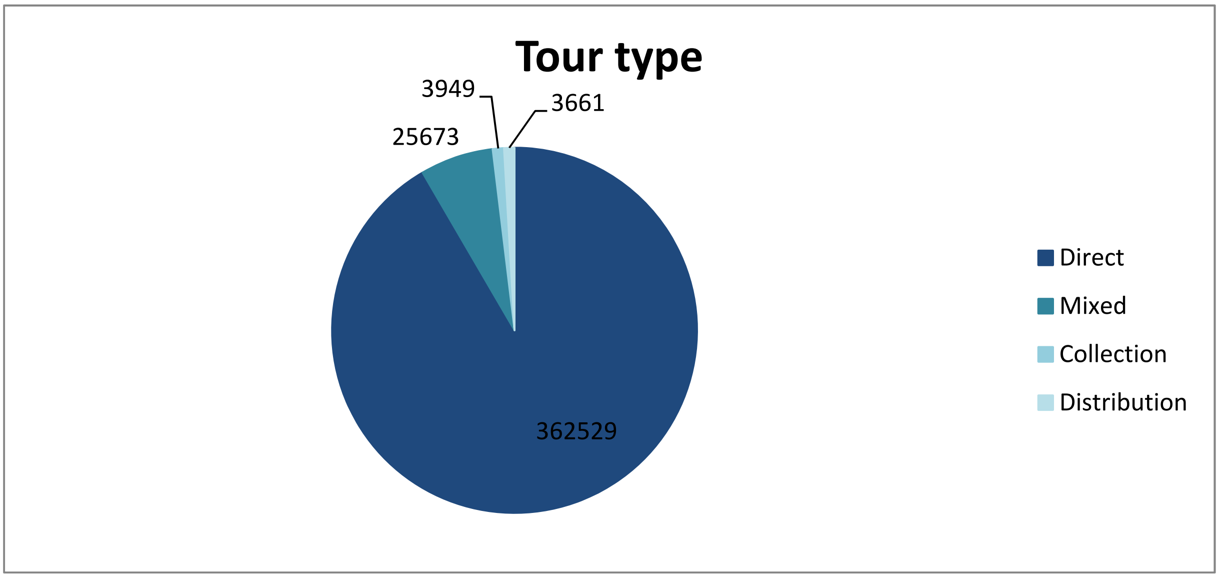

Figure 4.10. The distribution of tour types.

DURATION

For this data analysis, the tour duration is obtained using the start and end time (and date) of the tour, including both the time spent traveling on the road and time spent servicing customers. It should be noted that the tour start and end time are not very reliable variables. For this reason, we see quite a few tours that last exactly 24 hours and a filter has been applied removing tours that last less than five minutes. More short tours are to be found than in the studies of Figliozzi et al. (2007) and You (2012), which may in part be explained by the fact that a tour ends when it turns empty. These short tours also reflect the tendency to prefer less complex tours. Interestingly, quite a few tours last longer than the maximum allowed daily driving time of 9-10 hours (Figure 4.11). Some tours may include a night of sleep or a switch of driver.

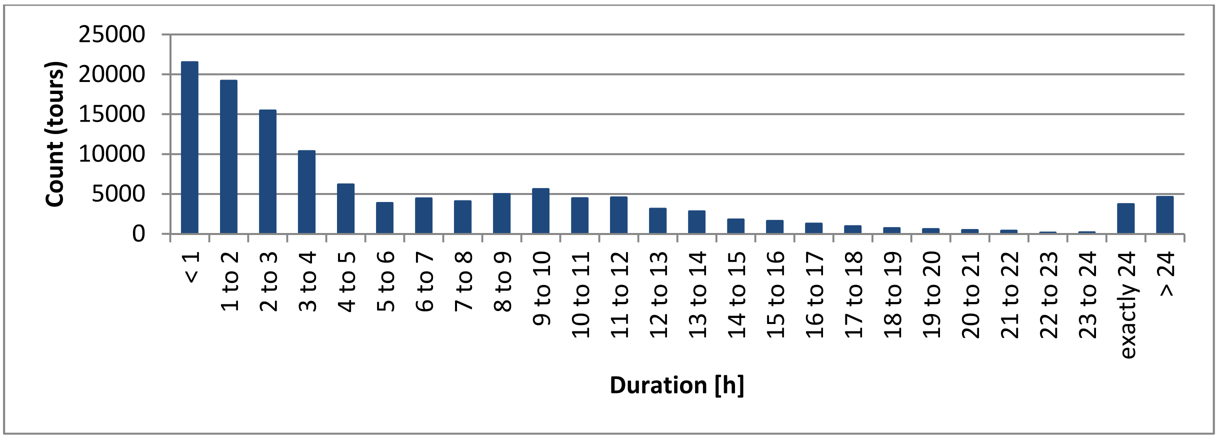

Figure 4.11. Tour duration distribution.

DISTANCE

The tour distance, a variable already present in the raw XML microdata, follows a somewhat similar exponential distribution to the tour duration. Tours much longer than 600 km are very rare (<1%). Besides the influences of constraints and preferences, the fact that only tours that stay in the Netherlands are selected can explain this. Many more short tours (< 100 km) are found than in Sydney, Australia by Figliozzi et al. (2007). Perhaps the difference in density of the built environment explains this, but most likely it is the definition of the term tour in the data, tours do not include empty trips and end when the vehicle turns empty.

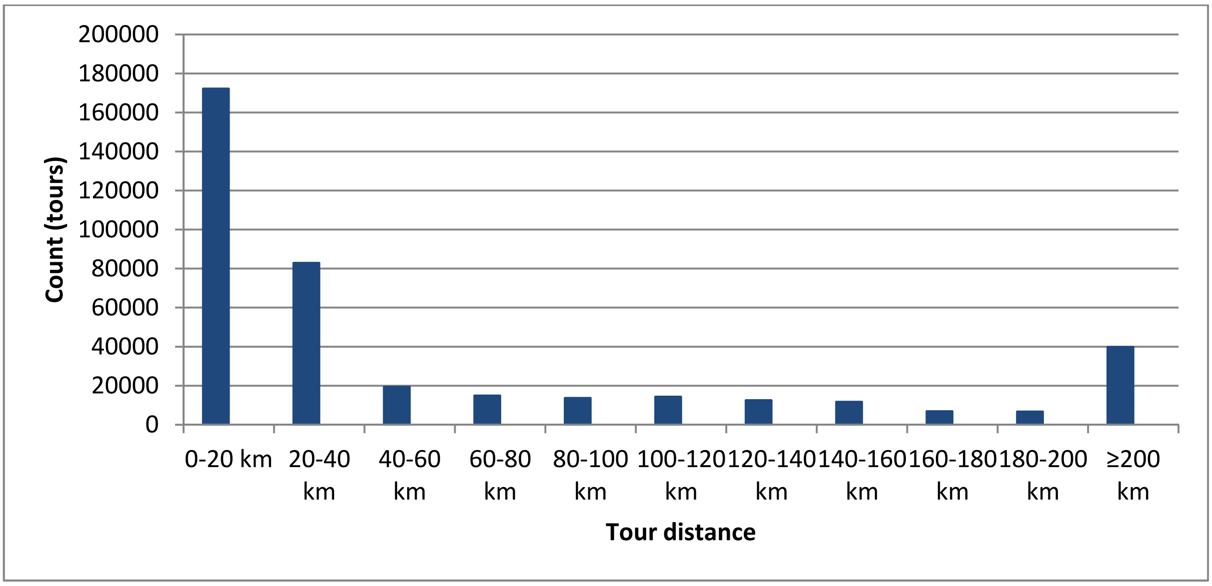

Figure 4.12. Tour distance distribution.

Maximum distance within tour

Tours can contain many stop locations, but often we see that these locations are not very far apart. A clear preference to group shipments with geographical proximity can be identified, which is logical as carriers prefer to construct efficient tours. Due to the selection of tours that stay within the borders of the Netherlands, only rarely distances over 250 km are found between locations in a tour.

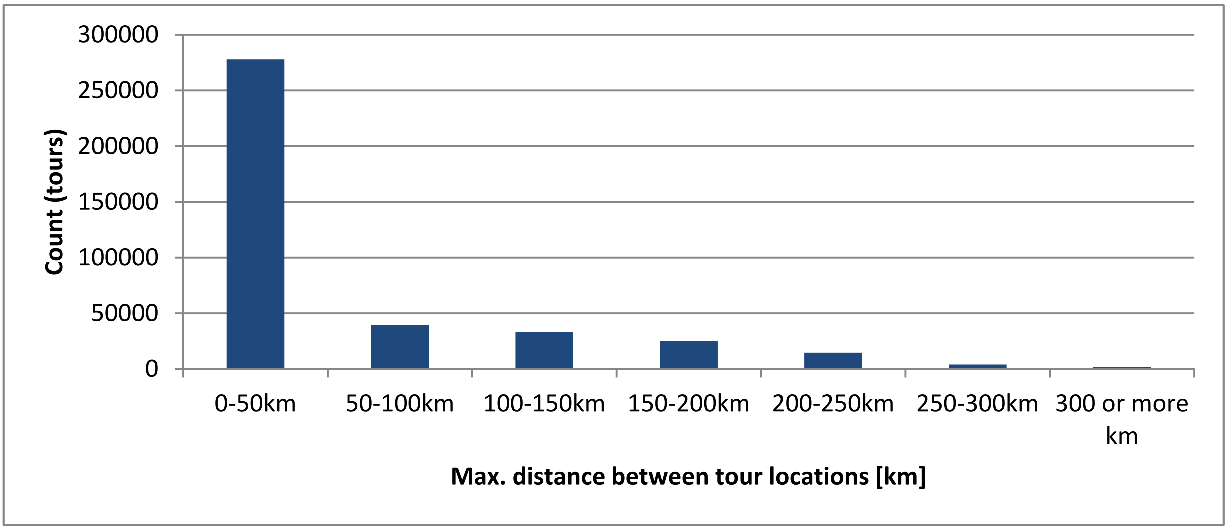

Figure 4.13. Distribution of the maximum distance between locations in a tour.

Capacity utilization

From the data we have obtained a crude and straightforward measure of capacity utilization. The total gross weight of all shipments in a tour is divided by the carrying capacity of the truck. Quite often this capacity utilization is very low, which can be due to other constraints that are present, such as the volume capacity and the availability of shipments to transport on day. Many tours do transport a full truckload (close to 100%). Unfortunately the data for the volume of the shipments is not usable for this research, as it is filled out poorly and inconsistently. The fact that the capacity utilization often exceeds 100% has the following possible explanations:

- Not all shipments are present in the truck at the same time. After some shipments are unloaded, another location is visited where a shipment is loaded.
- The carrying capacity of the vehicle is not filled out correctly or does not include the carrying capacity of a second carriage.
- The carrier is breaking the law.

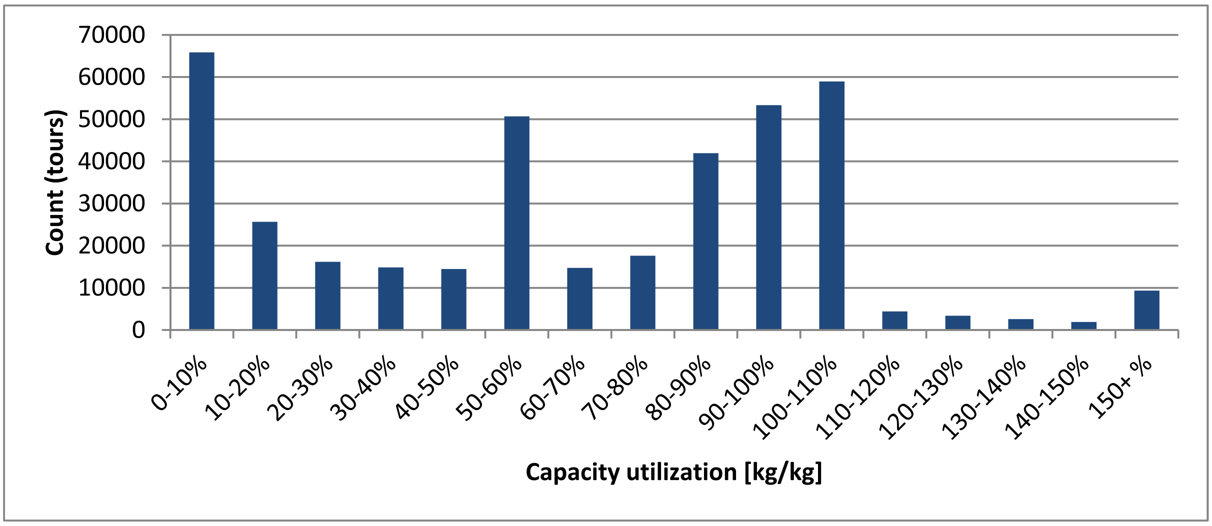

Figure 4.14. Capacity utilization distribution.

Combining different goods types

Carriers showcase a tendency to group shipments of the same goods type in the same tour. As can be seen in Figure 4.15, a large majority of multiple-stop tours only transports shipments of the same NSTR goods type. This can be due to restricted goods combinations, discussed in Section 2.3.2. Some carriers may also mainly transport shipments of the same goods type in general. Still, 6.8% of multiple-stop tours does combine different goods types.

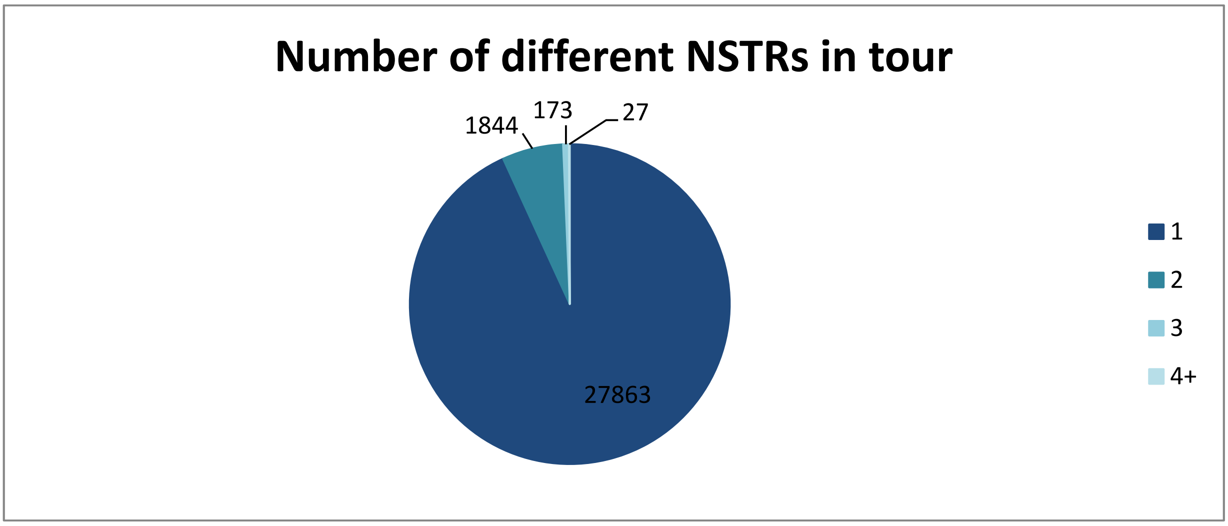

Figure 4.15. The share of tours transporting a certain number of different NSTR goods types (1-digit level). Only tours with more than two stops are included.

Vehicle type

Different types of vehicles are used by the carriers. Based on the variables in the data, the four vehicles types in Figure 4.16 can be distinguished. Other/special vehicles are usually vans, but these are only rarely observed.

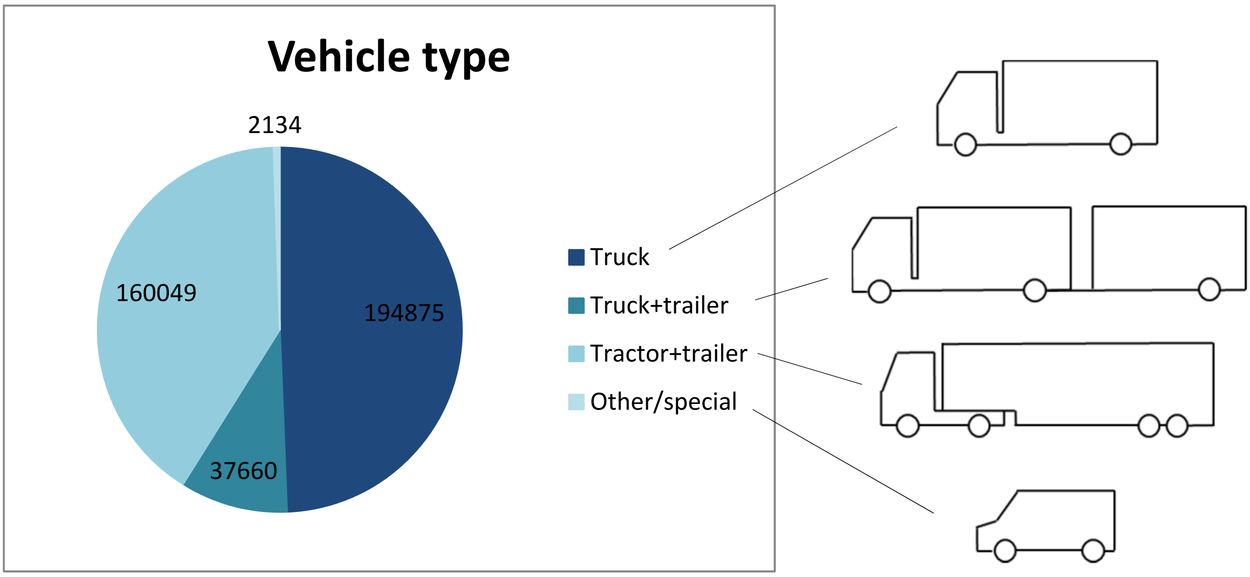

Figure 4.16. The share of vehicle types.

Location type

Tours can visit different location types. Different zones have been assigned a certain land use. Port transshipment nodes and distribution center zones are distinguished based on the number of employees of firms performing port or distribution activities (using SBI business information data), and the number of stops made in these zones in the XML microdata. Urban zones and retail zones are identified using the CBS zonal data (Kerncijfers wijken en buurten 2015). Urban zones are zones that CBS identifies as very strongly (>2500 addresses per km^2) or strongly urbanized (1500-2500 km^2) (CBS, 2015b). Retail zones have a number of business and food service establishments higher than 100 and a share higher than 40% (see Appendix D for further substantiation).

A substantial share of the tours visits a location where logistical activities (port transshipment or distribution) are performed. This may be related to the fact that we mainly have large third-party logistical service providers in the data set. Most tours do not visit an urban zone, even though 26.3% of the Dutch zones has been identified as one. Apparently city distribution is not a very prevalent type of freight activity in the data set. Still, quite a few tours visit a zone with a lot of retail activities. A carrier that is hired by a supermarket establishment located in a town center to transport restocking orders from its central warehouse would be one such example.

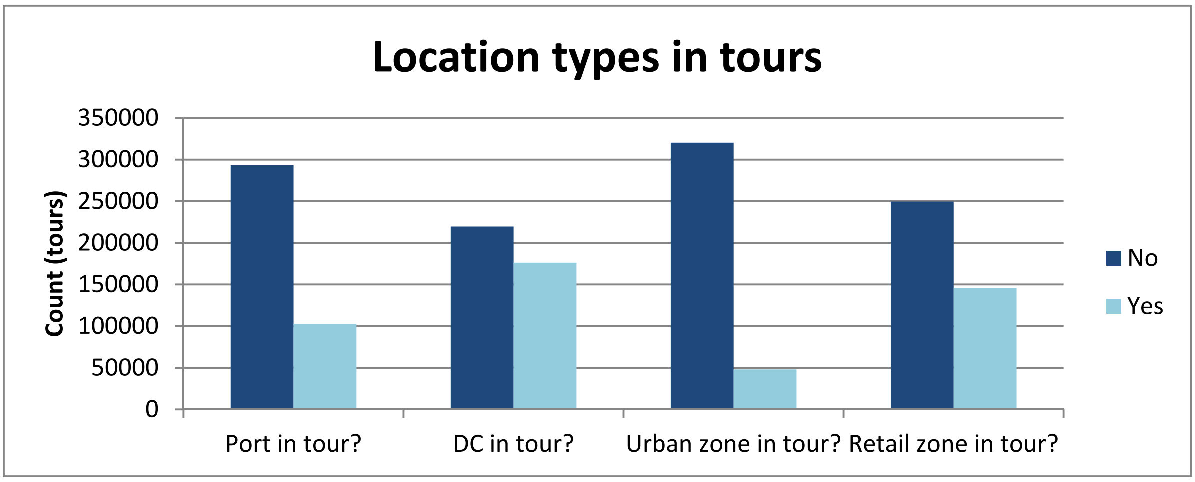

Figure 4.17. Distribution of tours visiting location types.

DEPARTURE TIME

The distribution of the departure time of tours shows peaks from 4AM-5AM and from 9AM-2PM (Figure 4.18). Both are periods outside the morning and afternoon peak of rush hour traffic. Hunt & Stefan (2007) found similar statistics. Clearly, carriers prefer to evade congestion. Quite a few tours still start during rush hour because carriers prefer to use their trucks all throughout the day and some carriers do not have enough trucks to perform all tours outside the peaks (de Jong et al., 2016).

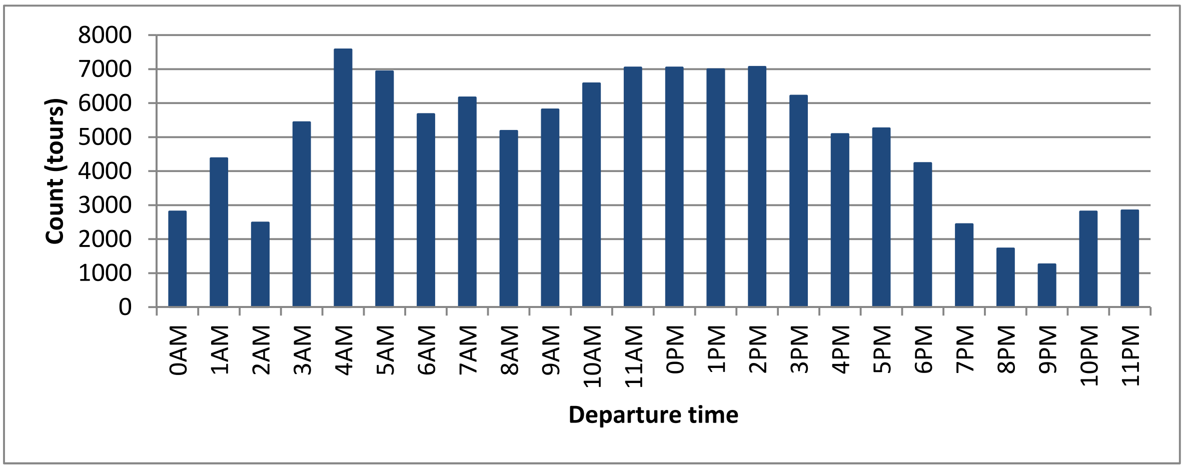

Figure 4.18. Tour departure time distribution.

4.3.3 Further analyses on number of stops

Different numbers of stops are observed for different types of tours. In this section, we discuss how this differs for different goods types, vehicle types, location types, and distances.

For different goods types we observe varying percentages of direct tours, i.e. tours with only 1-2 stops. In Figure 4.19, the NSTR goods type of the tour is calculated as the NSTR goods type of which the most weight is transported in the tour. This allows us to deal with tours that transport multiple goods types. NSTR goods types 2 to 5 are merged into one category, for they have a very low number of observations and they are all related to oils and metals. Tours transporting oils and metals *(NSTR 2 to 5)*, construction materials *(NSTR 6)*, or chemical products *(NSTR 8)* are much more often a direct tour than others. Tours transporting agricultural products and livestock *(NSTR 0)* or other foodstuffs and fodder *(NSTR 1)* often visit multiple stops.

These differences in number of stops between goods types can be explained in many ways: weight or volume differences, stricter restrictions on goods type combinations, ease of loading/unloading, and dispersion of demand. For example, a shipment of construction materials *(NSTR 6)* that is transported to a building site is intuitively more likely to have a high volume than a shipment of soda cans *(NSTR 1)* that are transported to a restaurant. Therefore, a vehicle transporting the latter is more likely to have remaining space available for other shipments. Chemical products *(NSTR 8)* might be more restricted in their combination with other goods if they are highly flammable. Construction materials such as long piles *(NSTR 6)* may take more time and effort to load than potatoes from a farmer *(NSTR 0)*. Finally, restaurants ordering foodstuffs *(NSTR 1)* can be highly concentrated (Beziat et al., n.d.), whereas factories producing steel *(NSTR 2 to 5)* may be highly dispersed. In the first case, it is easier to combine shipments of different customers in an efficient way.

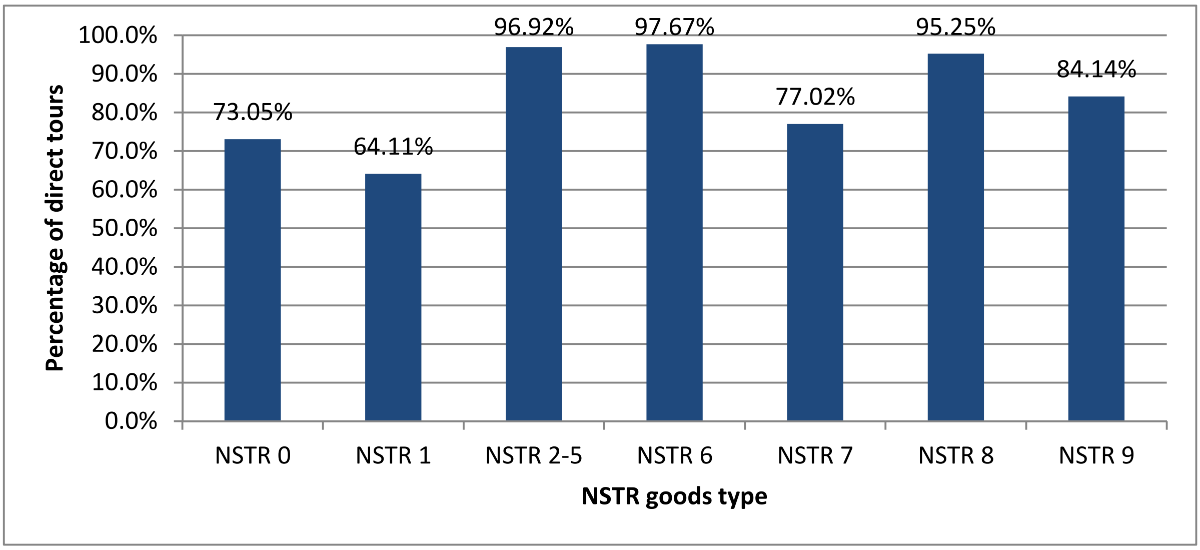

Figure 4.19. Percentage of direct tours (i.e. 1-2 stops) for different goods categories. Concrete shipments are excluded as we already know these are only transported in direct tours.

Multiple-stop tours are observed most often when a truck is used, less often with tractor + trailers, and only rarely with truck + trailers and other/special vehicles. Differences between these vehicle types can be related to capacity and the ease of loading/unloading. Other/special vehicles are mostly vans, which have a very low capacity, and therefore little room for the combination of shipments of many customers. A truck + trailer is less practical for unloading of goods for different customers than a simple truck, as the trailer needs to be uncoupled to unload the shipments in the first compartment. A tractor + trailer is more practical than a truck in this regard. Finally, different types of carriers may own different vehicle types. Carriers that mainly use tractor + trailers may have different planning strategies and leading constraints than carriers that mainly use trucks.

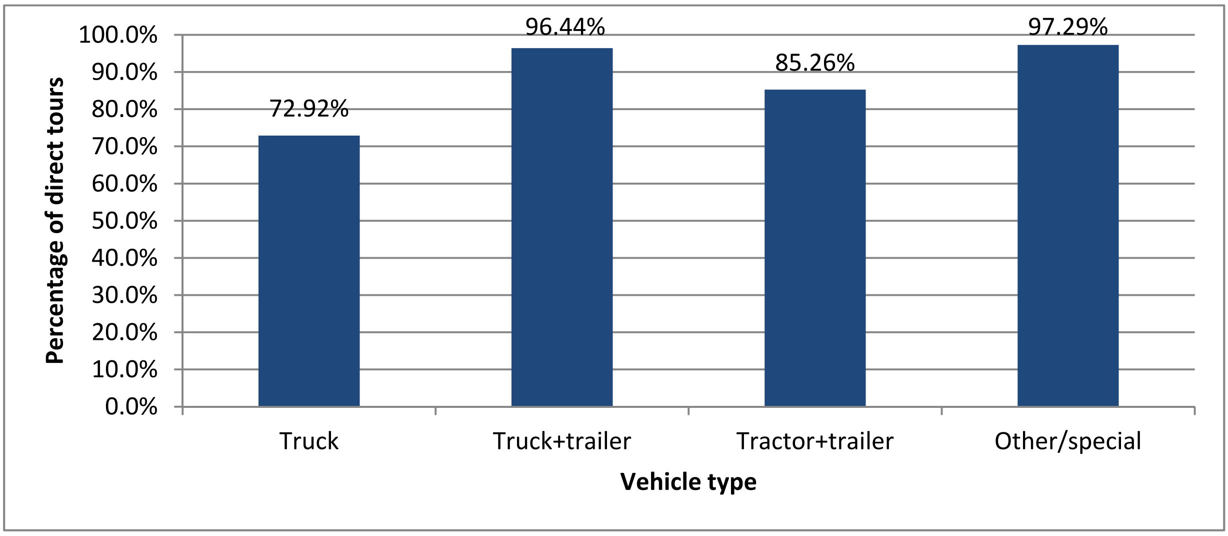

Figure 4.20. Percentage of direct tours (i.e. 1-2 stops) for tours using different vehicle types. Concrete shipments are excluded.

The large majority of tours that visit port transshipment zones are direct tours; combination of shipments of different customers is rarely observed here. In Figure 4.21, we can see that tours that do not visit port transshipment zones much more often have multiple stop locations. No clear difference is apparent between tours that load or unload goods at a port.

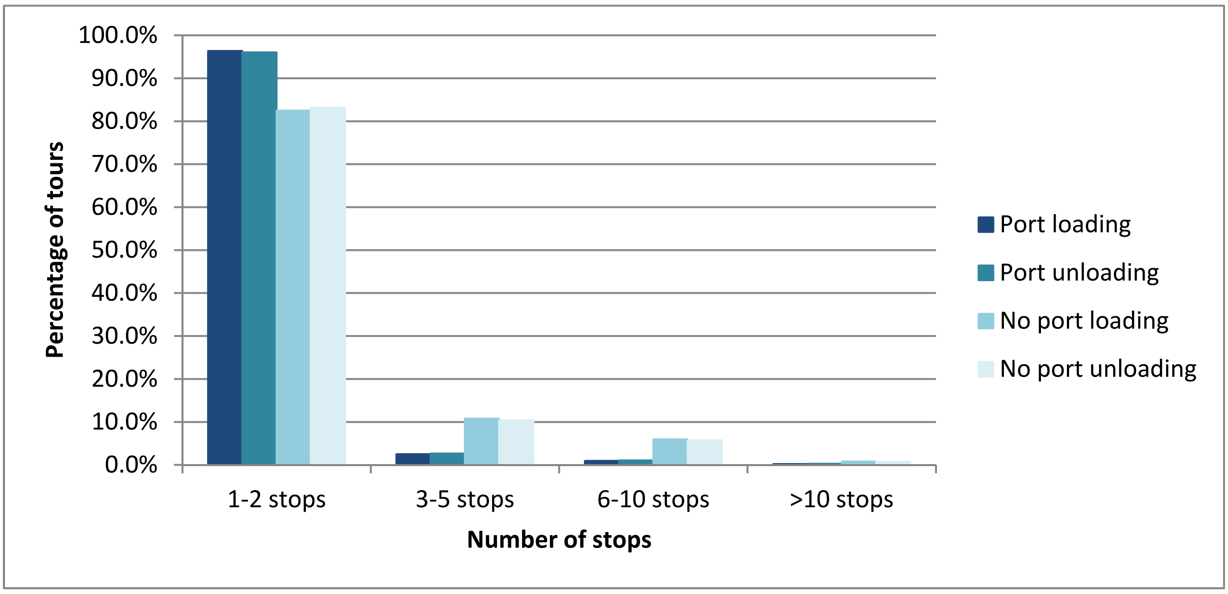

Figure 4.21. Percentage of tours with a certain number of stops that visit a port transshipment zone for loading/unloading or not. Concrete shipments are excluded.

For tours that visit distribution center zones the effect is quite the opposite. When goods are loaded at a distribution center in the tour, the tour is more likely to have multiple stops. When goods are unloaded at a distribution center in the tour, the tour is less likely to have multiple stops than when goods are not (Figure 4.22).

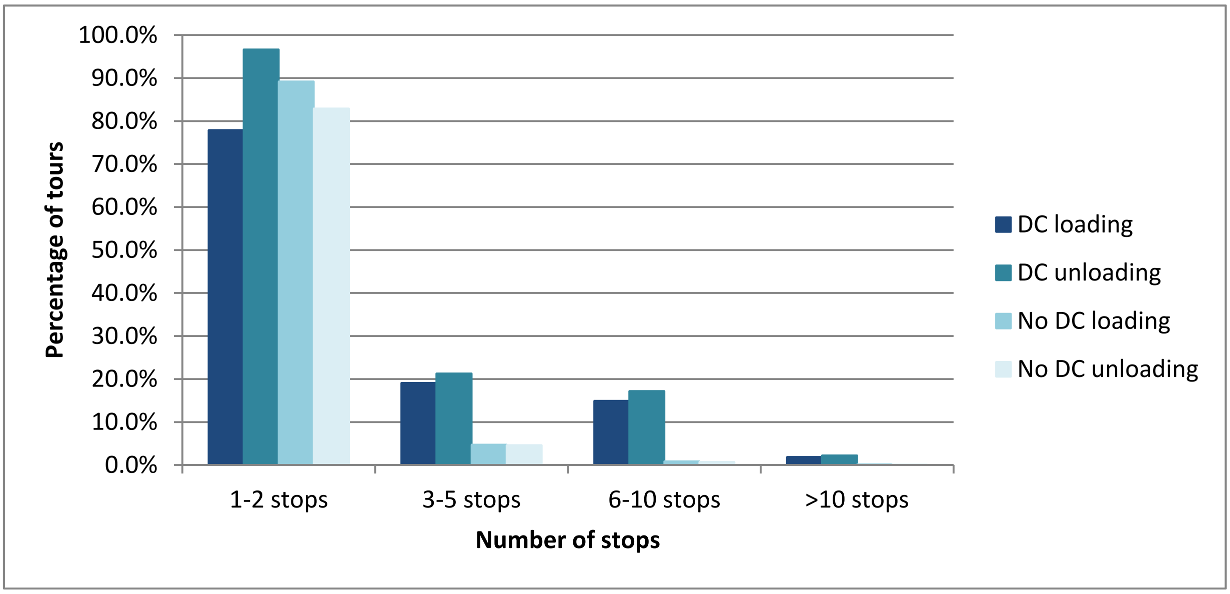

Figure 4.22. Percentage of tours with a certain number of stops that visit a distribution center zone for loading/unloading or not. Concrete shipments are excluded.

Tours that visit a port transshipment zone are more likely to transport shipments as part of a long-distance international logistics chain. The tour transports the shipment while it is on its way from a producer to perhaps a wholesaler or distribution center, in which case large shipment sizes are observed, often not smaller than a container (Friedrich et al., 2014). With such high volumes and weights, it is often not possible to combine shipments of different customers in one vehicle.

Tours that load goods at a distribution center most likely transport shipments to places of consumption, such as retail outlets, in which case smaller shipment sizes are observed (Friedrich et al., 2014). There are more possibilities to fit multiple shipments in a vehicle. Khan & Machemehl (2017) similarly found that tours visiting a distribution center tend to have more stops and reasoned that this may be due to the operations inside these centers: “the vehicle load gets replenished (for outbound delivery) or emptied (for outbound pickup) at the distribution center, which allows the vehicle to go on with more customer visits” (Khan & Machemehl, 2017:95). The distribution centers are also likely to have a large consolidation potential: they may have more shipments to transport on a day, which also have loading points (or unloading points) in common more often. Larger vehicles are sometimes observed at distribution centers too (van Duin et al., 2012). Tours that unload goods at a distribution center actually tend to have fewer stops. These tours might transport shipments more often that are part of the marketing channel related to the point of production instead of consumption.

In Figure 4.23 these different marketing channels and shipment sizes are depicted in a highly simplified manner. The real world shows highly diverse marketing channels. A shipment may, for example, skip the port transshipment step, and producing facilities can also receive shipments.

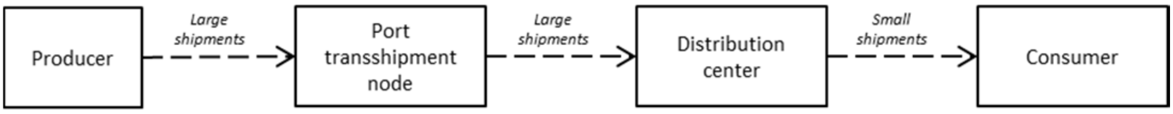

Figure 4.23. A simplified representation of marketing channels and according shipment sizes. Based on the concept of a large and small shipments network of Friedrich et al. (2014).

Tours that visit urban zones tend to include more stops (Figure 4.24). The demand is much more concentrated in urban areas; therefore, many customers may be visited in the same tour within a relatively short time. Especially if the vehicle has to be driven from some distant location into the city, it is efficient to visit multiple customers instead of making multiple direct tours each entering and leaving the city. These highly urbanized regions are also more likely to be places of consumption and part of the small shipments network (Friedrich et al., 2014).

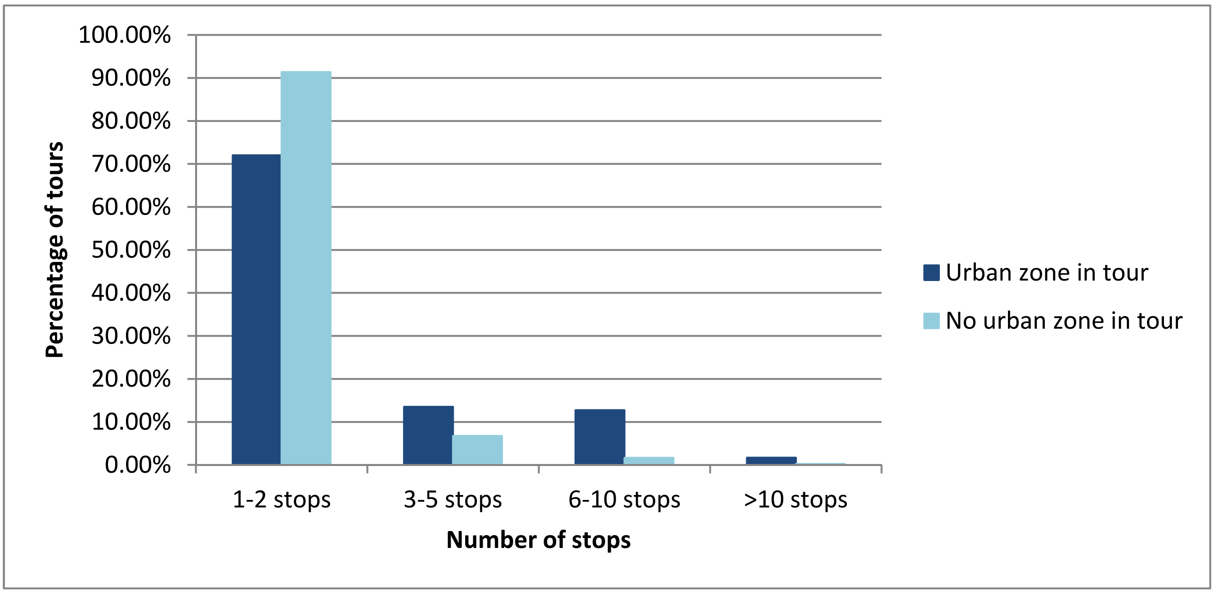

Figure 4.24. Percentage of tours with a certain number of stops and that visit an urban zone or not. Concrete shipments are excluded.

Retail shops are places of consumption and, therefore, more likely to receive small shipments (Friedrich et al., 2014). Consequently, tours visiting retail zones on average include more stops in the data. This in contrast with the findings in Texas, USA of Zhou et al. (2014), where more direct tours were observed when retail zones are visited due to time windows. Perhaps, because the Netherlands is a denser country, carriers can combine visits to retailers more often despite these time windows.

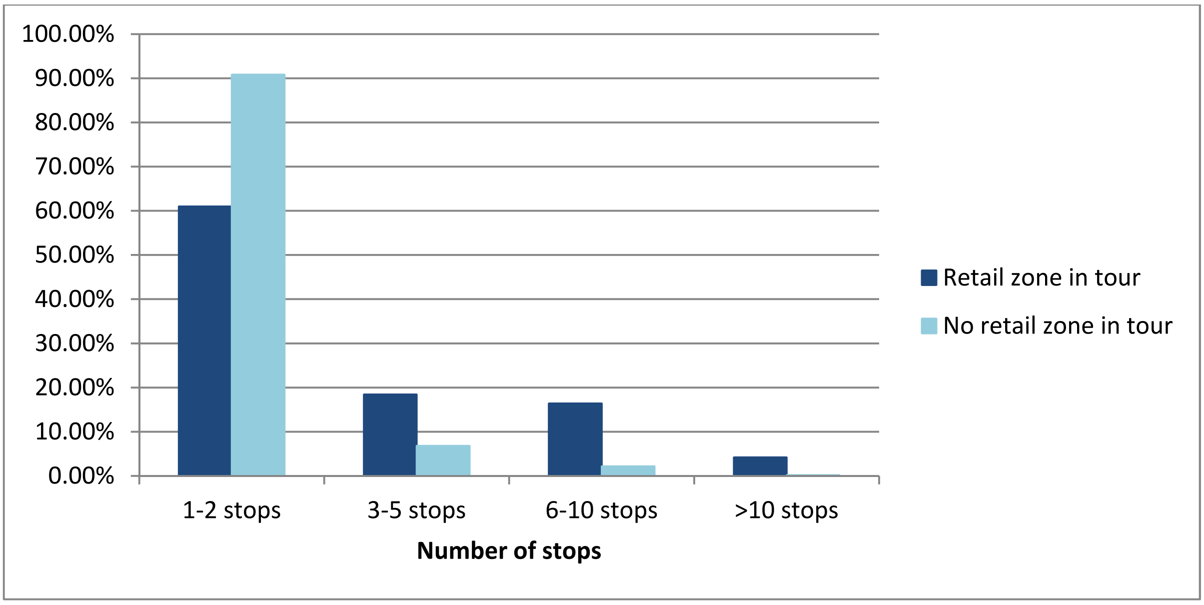

Figure 4.25. Percentage of tours with a certain number of stops and that visit a retail zone or not. Concrete shipments are excluded.

When we look at the mean and median tour distance set out against the number of stops per tour (Figure 4.25), we see that tours with more stops travel a longer distance. It is intuitively logical that an additional stop in a tour requires additional distance. However, the median distance is strikingly lower for tours with 1-2 stops. This further substantiates the idea of constructing simple (i.e. direct) tours for shipments that are located nearby (Nuzzolo et al., 2012). For these short-distance shipments, the efficiency gains of adding another shipment are also lower. If a shipment by itself already requires a rather long tour, carriers prefer to add more shipments to this tour, since the tour already lasts long anyway. Especially if other shipments can be added to this long tour with little additional distance, a cost minimizing VRP will combine these shipments into one tour, as making multiple long tours for all these shipments would be very inefficient.

Tours with more than ten stops do not clearly tend to have higher distance than those with six to ten stops. Probably once the tour is about 300-400 km long, usually no more shipments are added to the tour unless they require very little additional distance. Possibly the working hour constraint of 9-10 hours is reached with these long tours.

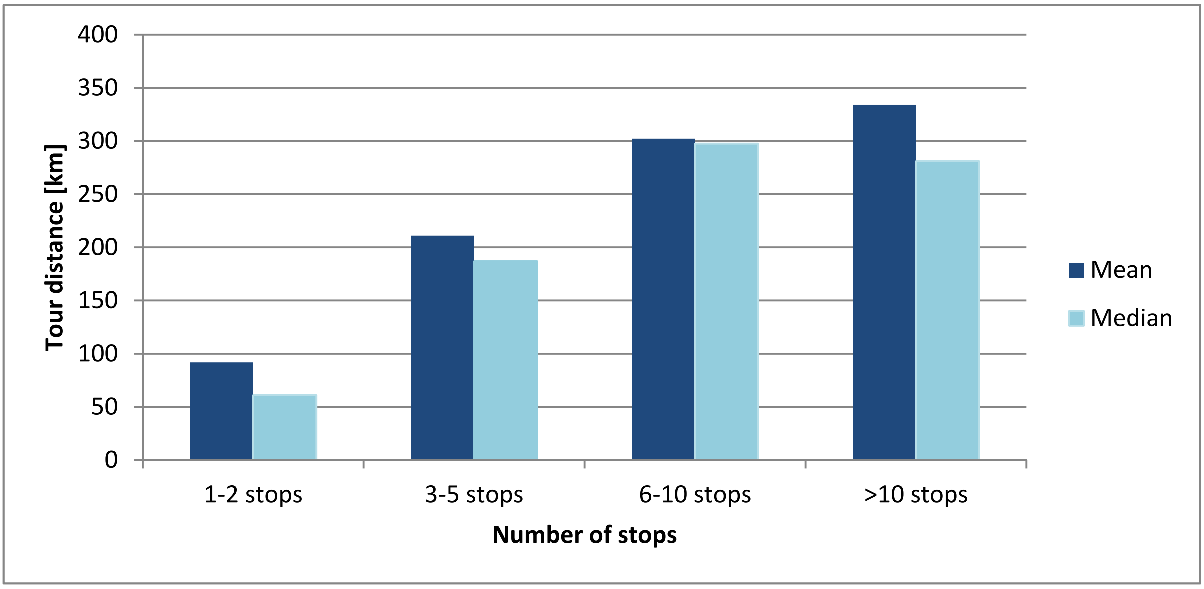

Figure 4.26. The mean and median distance of tours with a certain number of stops.

4.4 Concluding remarks

The XML microdata has its peculiarities that impact the distributions of variables and constrain what can be done with it for model development and estimation.

The conditions that start and end a tour in the data, lead to the exclusion of empty trips, and cuts up vehicle journeys that turn empty into multiple tours. For this reason, very few tours are observed with more than one shipment or more than two stop locations in comparison to previous studies (e.g. Figliozzi et al., 2007; Beziat et al., n.d.). In addition, the tour distance distribution shows a high peak for short distances for this reason.

The XML microdata imposes limitations on the model structure and parameters that can be estimated, most notably the following:

- Constructed tours should start at the loading location of the first loaded shipment, as empty trips from the home base are not included and the location of the home base of the vehicle is very poorly filled in.
- No time windows, estimation of goal arrival times like You et al. (2016), or dwelling times can be included, as only the start and end time of the full tour are known.
- Only direct shipments can be modeled, since the data does not provide information about the complete logistics chain a shipment goes through.

- The sequence of visiting loading and unloading locations is not known, we only know which shipments are part of the same tour. This complicates a VRP calibration approach, as we would need to make assumptions to obtain an observed tour sequence to calibrate a model that chooses a tour sequence. We do not observe the output of a VRP in sufficient detail. Not knowing these realized trips within tours also makes an incremental trip chaining approach like Hunt & Stefan (2007) problematic. Instead of a model that adds trips to tours, we can construct a model that adds shipments to tours.
- The definition of a tour might lead to a model that constructs many tours with one shipment allocated to it, and causes the specification of empty trips to be problematic and require crude assumptions.
- The data set is most likely not representative for all freight carriers in the Netherlands, since the XML tool causes a self-selection of third-party carriers with more advanced planning software. Conservatism is required when formulating statements about freight transportation in the Netherlands in general.

However, we do know all the shipments that were transported in a tour and the loading and unloading locations of all shipments are known for quite a substantial portion of the tours. This leaves room for calibration of a tour formation model through specification of choice situations in an algorithm that allocates shipments to tours. Combined with the enormous number of observed tours compared to other data sets, this is a unique data set that has definite value for calibration of a tour formation model.

Enrichment of the data with other variables can also provide a lot of useful insights that help us further understand tour formation, most notably:

- Tours visiting port transshipment zones have fewer stops, while tours visiting distribution centers have more stops. Different shipments size, number of available (similar) shipments, and the logistic operations inside distribution centers can explain this.
- In urban areas and retail zones, more stops are made, as demand is more concentrated and entering and leaving a city can be very time consuming.
- Fewer stops are observed for tours transporting oils and metals, construction materials, and chemical products. This is due to restricted goods combinations, dispersion of demand, volume, and of ease of loading/unloading.
- Tours with more than fifteen shipments or stop locations are only very rarely observed.
- Tour locations are usually quite close to each other, and rarely more than 250 km apart.
- If a tour already covers a long distance, there is a stronger tendency to allocate more shipments to it, as this is far more efficient than constructing multiple tours. For short-distance shipments, the efficiency gains of tours are smaller and a desire to construct simple tours can be identified.
- A substantial portion of tours transports much less weight than its vehicle capacity, other preferences and constraints (e.g. volume) can thus be leading too.

5. METHODOLOGY

In this chapter, we explain and substantiate the structure of our developed tour formation model.

5.1 GENERAL STRUCTURE OF THE TOUR FORMATION MODEL

The tour formation model allocates shipments to tours. A shipment is defined in this research as a physical object with a unique combination of loading location, unloading location, goods type (NSTR 1-digit classification), and tour that it is allocated to. A tour is defined as a sequence of visiting locations to load and unload shipments. To remain consistent with the dataset, we define that a tour starts when a shipment is loaded into an empty vehicle, and ends when the last shipment is unloaded causing the vehicle to turn empty.

The tour formation model consists of two sub models, an 'End Tour' (ET) sub model, and a 'Select Shipment' (SS) sub model. A tour is grown iteratively by allocating one additional shipment to it in each iteration, until the tour is ended with the ET sub model, in which case we start constructing the next tour. If the ET model does not end the tour, a shipment is chosen with the SS model and added to the tour.

This model structure is similar to that of Hunt & Stefan (2007), but the key difference is that shipments instead of stop locations are added iteratively to tours (see Figure 4.7 for the difference between a shipment and a stop location). In addition, in our model the sequence of visiting locations is chosen with a separate algorithm instead of being dependent on the iteration in which a location was added to the tour.

In the ET model, the dependent variable is binary, with the categories: '0 = continue adding shipments to tour' and '1 = end tour'. A Binary Logistic Regression (BLR) is estimated to explain this binary variable. The SS model has multiple shipment alternatives, each with alternative specific attributes. A Multinomial Logit (MNL) is estimated to explain the selection of a shipment alternative. Additionally, constraints are specified that end a tour regardless of the ET model and limit which shipments can be chosen in the SS model.

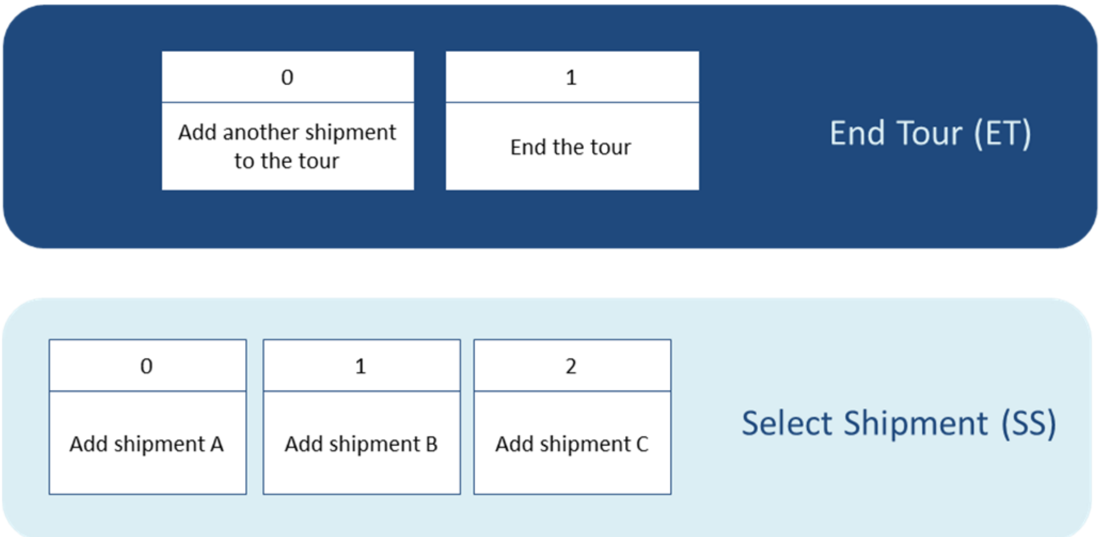

Figure 5.1. The alternatives of the two sub models.

The 'choices' modeled in the ET and SS models are not assumed to be choices that transportation planners necessarily make. Instead, we intend to represent tour formation behavior with the full model structure. Tour formation behavior is the result of a set of complex unobserved choices for which transportation planners might use different planning methods and/or software. Consequently, development of a model structure that represents explicitly each choice made in the tour formation process is an unrealistic goal.

The ET and SS models should be seen as a statistical model, instead of a choice model that represents an actual choice context. This allows us to represent behavior in a tour formation algorithm. In order to represent the behavior accurately, though, it is important that we can find meaningful variables that can explain the dependent variable in these statistical models. These effects are meaningful if they are explicable based on the objectives and constraints that lead to behavior. For example, in the ET model preferences for tours with few stops can be considered through a higher or lower probability of ending the tour, while in the SS model the desire to save transportation costs can be considered through a higher probability of selecting a shipment with little additional time to the tour.

The required input for application of the developed model in a freight microsimulation framework is a synthesized set of shipments, a skim matrix, and a vehicle type choice model. The output is a set of tours containing these synthesized shipments.

5.2 Detailed model steps

This section describes the steps in the tour formation model in further detail. In Figure 5.2, a flow diagram shows this detailed model formulation.

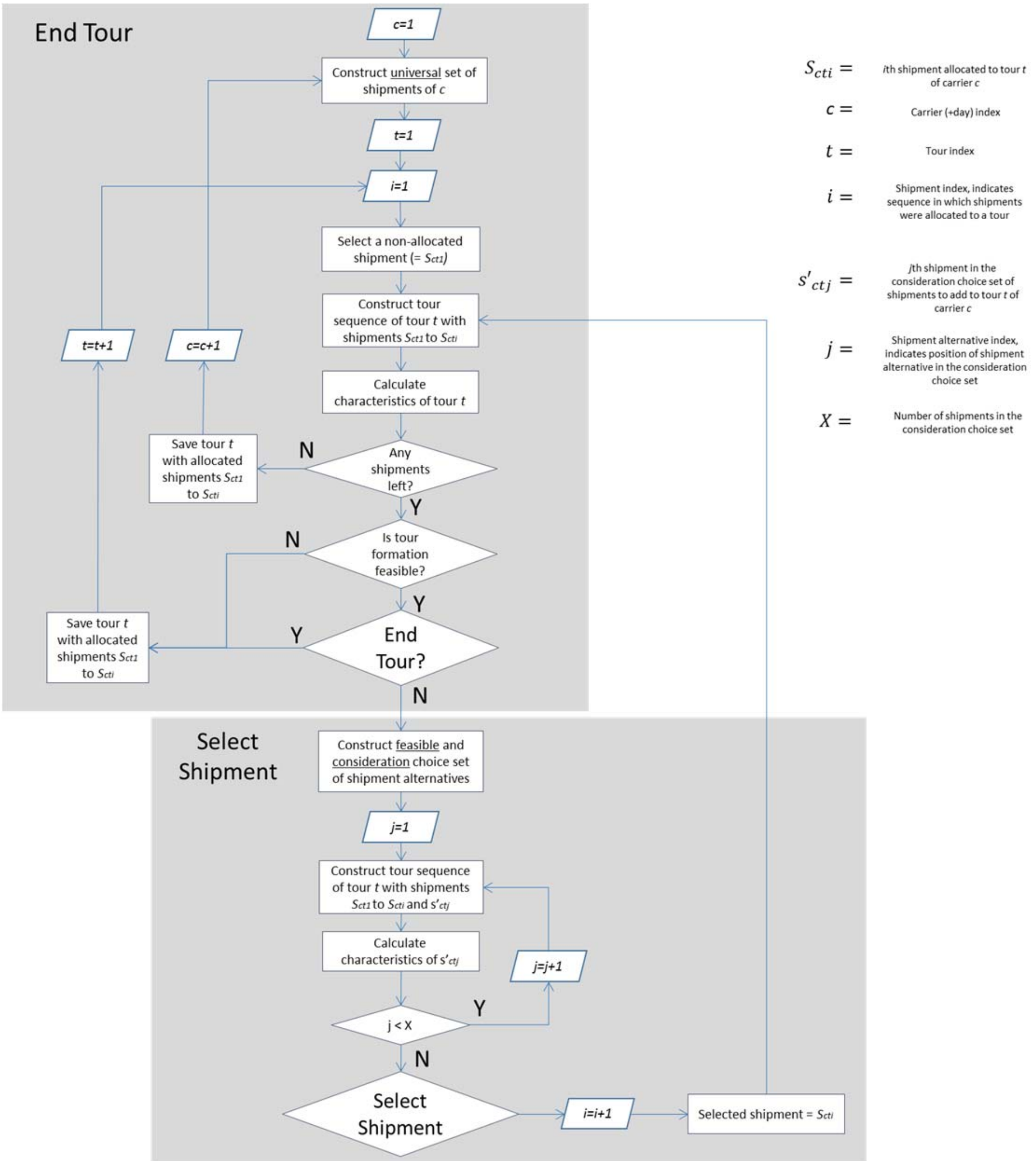

Figure 5.2. Conceptual diagram of the tour formation model.

The model thus loops through all shipments that a carrier has to transport on a day (the universal set), and exhaustively forms tours with these shipments. Next, it loops through all days and carriers, such that all shipments are allocated to a tour. We have decided to define this universal set because some limitation is necessary with respect to which shipments can end up in the same tour. It is not reasonable to assume that carriers can also select shipments from other carriers, and a carrier is not be able to combine two shipments that need to be received several months apart. A day is chosen as the time unit because the interview showed that (for at least some carriers) the day that a shipment needs to be transported is set in stone. Practically, this is also advantageous, since considering many hundreds of shipments as the next shipment might lead to unpractical running times.

Two parts to the model can be distinguished: the End Tour part (from the top to and including the End Tour block) and the Select Shipment part (below the End Tour block). In the next two sections we elaborate on the steps in these two model parts as shown in Figure 5.2.

5.2.1 End Tour sub model

Before the decision to end the tour is made, several steps need to be taken.

New tour

Firstly, we need to pick a first shipment from which we start growing the tour. This is done at random, there is no clear reason to assume that a certain shipment is more likely to be chosen for a new tour, since all shipments need to be allocated to a tour anyway, and the (still empty) tour has no characteristics yet to base this choice on.

Tour sequence algorithm

After each shipment that is allocated to a tour, the tour sequence is constructed with all shipments in the tour so far. The tour sequence is the order of visiting the loading and unloading locations of all shipments in the tour. It is required in order to obtain features of the tour so far (i.e. with all shipments allocated up to this iteration), such as the tour duration and number of stop locations, which are explanatory variables for the ET choice model. We emphasize that the tour sequence algorithm is not the SS model. The SS model chooses which shipment is added to a tour. After the shipment is added, the tour sequence algorithm is used to pick a sequence of visiting the loading and unloading locations of all shipments in the tour.

Two different tour sequence algorithms have been developed. Both tour sequence algorithms use a nearest neighbor search. The first algorithm first visits all loading locations and then all unloading locations, while the second algorithm visits alternately loading and unloading locations. The second algorithm was developed in addition to the first algorithm, for it was observed that a substantial share of observed tours in the data have a set of loading and unloading locations for which such a sequence of alternately loading and unloading would be much more logical (Table 5.1). Each time the tour sequence is constructed in Figure 5.2 (both in the ET and SS part), both algorithms are used and the sequence with the shortest duration is selected.

Table 5.1. An example of a tour structure in the data for which a loading-unloading-loading-unloading sequence would be more logical. The first algorithm may generate a sequence of A-B-C-D-D-C-B-E, while the second algorithm would simply visit A-B-C-D-E.

Shipment	Loading location	Unloading location
1	A	B
2	B	C
3	C	D
4	D	E

The nearest neighbor search is a simplistic greedy tour construction algorithm that iteratively chooses the unvisited location that is nearest to the most recently visited location (Kizilates & Nuriyeva, 2013; Hougardy &

Wilke, 2015). The advantage of the nearest neighbor search is that it requires little computation time and is easy to implement, while showing good results in practice (Kizilates & Nuriyeva, 2013). However, it does sometimes miss visually obvious shorter tours, and in general performs worse than more complex and heavy algorithms (AlSalibi et al., 2013). Since the tour sequence needs to be constructed after each shipment allocation step (and with each considered shipment alternative in the SS sub model), and tours can have more than ten stops, the low computation time of the nearest neighbor search is of large importance, making it more attractive than more complex algorithms and enumeration of all possible sequences. Furthermore, our variation of the nearest neighbor algorithm is able to take into account precedence constraints, since all loading locations are visited before unloading locations.

In the first tour sequence algorithm, we first need to know the loading and unloading locations of all shipments in the tour. The tour starts at the loading location of the shipment that we selected as the first shipment of the tour (see section 'new tour' above). The algorithm then removes this loading location from an imaginary list of 'loading locations to be visited'. The next inserted loading location is the one that is still to be visited and can be reached within the smallest amount of time (based on the obtained skim matrix discussed in Chapter 4). This process is iterated until all loading locations are visited. Next, the unloading locations are visited, for which the same nearest neighbor search based on travel time is used. This first tour sequence algorithm is summarized with a flow diagram in Figure 5.3. A resulting tour sequence with this algorithm could look like Figure 5.4.

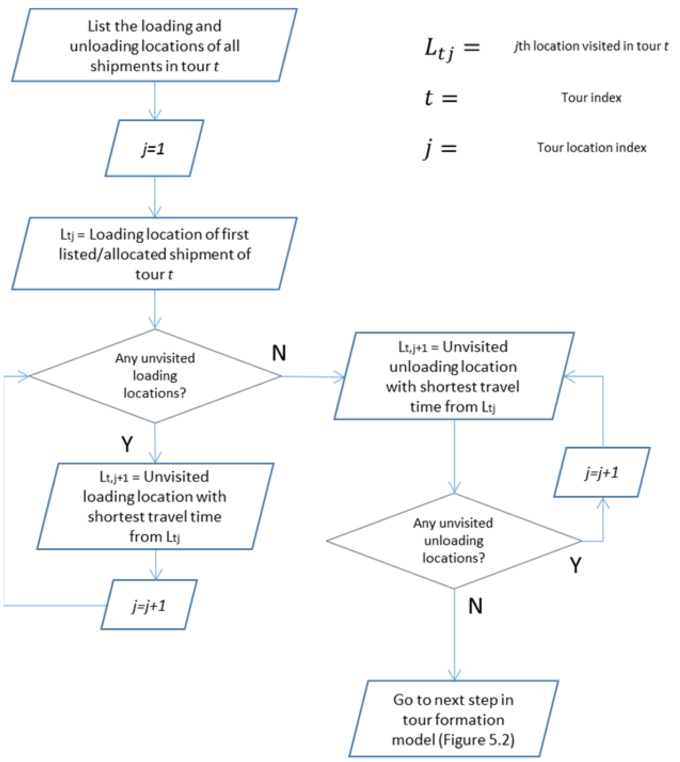

Figure 5.3. A flow diagram representing the first tour sequence algorithm.

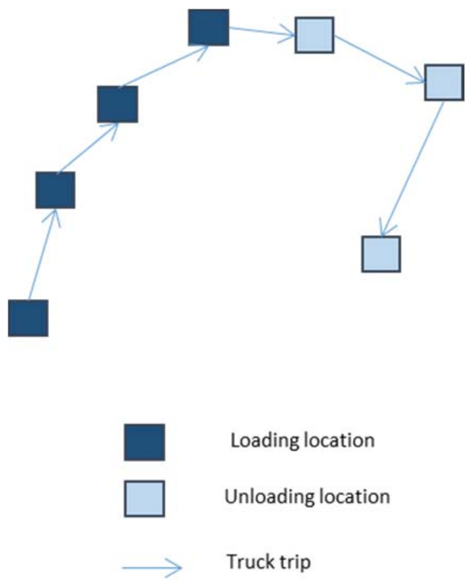

Figure 5.4. A hypothetical example of a tour sequence constructed with the first tour sequence algorithm.

CBS uses a similar algorithm to prepare the raw XML microdata into trip data for their BGW dataset. Similarly, first all loading locations are visited and then all unloading locations are visited (CBS, 2017b). However, they sort shipments based on shipment distance, the reported distance driven by the vehicle from loading until unloading. This may lead to logical tour sequences when observed XML data is transformed, but such an algorithm cannot be applied in a simulation model, as the shipment distance is one of the resulting outputs of a tour and is not known in advance when a tour needs to be constructed from scratch. Therefore, we have chosen to use a nearest neighbor search instead to order loading and unloading locations.

In the second tour sequence algorithm the tour also starts at the loading location of the first allocated shipment. However, the next visited location in the tour is the unloading location of this first shipment. Next, we keep track of all shipments whose loading location has not been visited yet and pick the shipment whose loading location can be reached within the smallest amount of time from the current tour location. The loading and unloading location of this shipment are the two locations that are visited next. If there are multiple shipments with the same travel time to its loading location, we simply pick the one that was allocated first. This process is repeated until all loading locations are visited.

In this second algorithm, we assume that at each visited loading location all shipments at this loading location are loaded. However, only the unloading location of first allocated shipment is visited after loading here. Therefore, it can occur that the unloading location of the other shipments picked up at this loading location have not yet been visited. When this occurs, these unvisited unloading locations are visited after all shipments have been loaded. This is done with the same nearest neighbor search as in the first algorithm. This way we ensure that the loading and unloading locations of all shipments in the tour are visited. This second tour sequence algorithm is summarized with a flow diagram in Figure 5.5. Figure 5.6 shows a hypothetical example of a tour constructed with this algorithm.

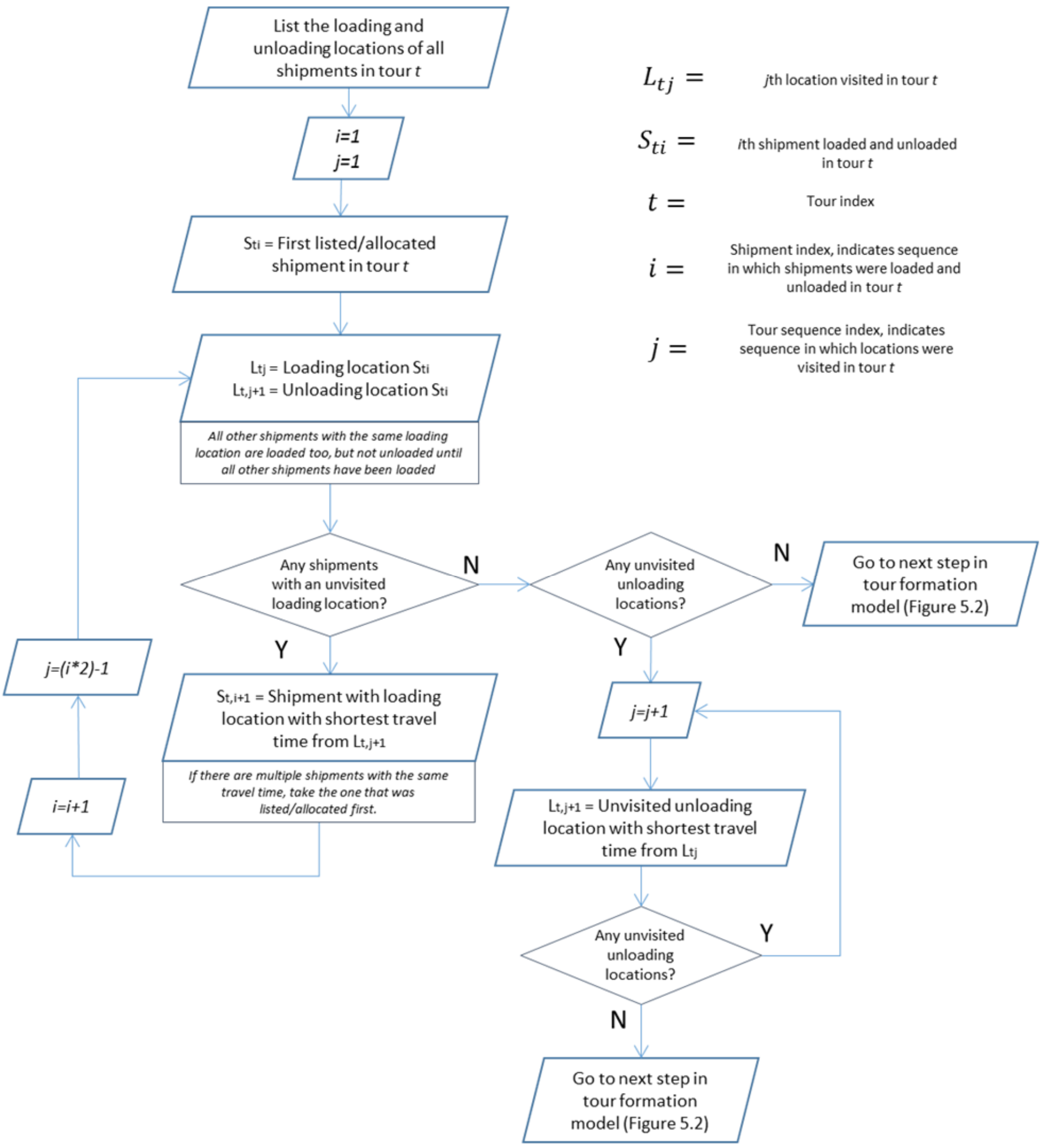

Figure 5.5. A flow diagram representing the second tour sequence algorithm.

Figure 5.6. A hypothetical example of a tour sequence constructed with the second tour sequence algorithm. We visit alternately loading and unloading locations, and visit remaining unloading locations afterwards.

TOUR CHARACTERISTICS

After the tour sequence is constructed, characteristics of the tour so far are obtained. These tour characteristics provide the required input for the ET choice model and determine whether it is feasible to add more shipments to the tour. In Chapter 6, we will discuss which variables are used in the ET choice model and interpret their meaning.

TOUR CONSTRAINTS

Tour formation is not considered in the following cases: (1) no non-allocated shipments are to be found with a 'proximity' lower than 100 km to the tour; (2) the tour transports a concrete shipment; (3) the capacity utilization is above 100%; (4) the tour lasts longer than nine hours. These constraints improve the behavioral foundation of the model and allow us to skip unnecessary calculations to model the ET and SS choices.

To connect the ET model with the SS model, we calculate a measure which we call 'proximity'. For each non-allocated shipment of this carrier (+ day), we calculate proximity as the sum of the distance of its loading and its unloading location to the nearest point in the tour so far (Figure 5.7). Besides as a constraint, proximity is also included as an explanatory variable in the ET choice model. Inclusion of this constraint allows us to incorporate that it might not be reasonable to continue adding more shipments if all remaining shipments are located relatively far away from the tour locations. Of course, if no shipments are left to be transported to begin with, no shipments are within a distance radius of 100 km either. Therefore, this constraint also ends the tour when no shipments are available anymore regardless of their proximity. In 88.3% of the ET choice observations there is a shipment within this distance radius of 100 km (based on Table E.1 in Appendix E). In Chapter 6 we also test the model with a more lenient constraint of 150 km to see if this improves the predictive performance.

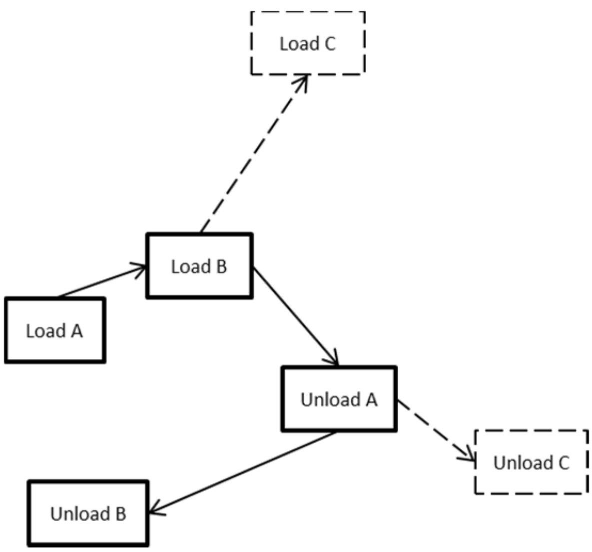

Figure 5.7. Proximity measure. If the current tour consists of shipments A and B, then the 'proximity' of shipment C is the sum of the distance of the two dashed arrows.

In Section 4.3.2 (number of shipments) we observed that and explained why multiple-shipment/multiple-stop tours are virtually never made with concrete shipments. For this reason, a tour is automatically ended when a concrete shipment is selected.

Capacity utilization is operationalized in the same way as in the data analysis in Figure 4.14. It is the total transported weight divided by the carrying capacity of the vehicle. It is also included as an explanatory variable in the ET choice model. Inclusion as a constraint allows us to consider that a vehicle simply cannot (or is not legally allowed to) transport more weight than its carrying capacity. Otherwise a high capacity utilization would only increase the probability of ending the tour. Note that we can still construct tours with more than 100% of the capacity used when a heavy shipment is added to a tour with 95% of capacity used. To keep this effect limited, capacity utilization is also included as a constraint in the SS sub model (Section 5.2.2).

Finally, the tour is ended when it lasts more than nine hours (excluding dwelling time at stops). In the Netherlands, working and break hours of truck drivers are regulated. Most notably, a truck driver is only allowed to drive for ten hours per day twice a week and for nine hours per day during the rest of the week (Inspectie Leefomgeving en Transport, 2018). Our constraint is already on the mild side, as the calculated tour duration does not include empty trips to and from the home base. Tour duration is also included as an explanatory variable in the ET choice model but inclusion as a constraint improves the behavioral foundation of the model. If we do not end the tour after nine hours, it is merely the probability of ending the tour that increases with a high tour duration, so we would still potentially construct tours that last unreasonably long due to other parameters explaining the choice.

MODELING THE END TOUR CHOICE

If tour constraints have not been violated, the characteristics of the tour as constructed so far are used to calculate the systematic utility and the probability of ending the tour. The systematic utility of ending the tour (U_{ET}) is calculated with equation 5.1, where β_i is the estimated parameter for explanatory variable x_i, and n is the number of explanatory variables in the estimated ET models in Chapter 6.

$$U_{ET} = Constant + \sum_{i=1}^{n} \beta_i * x_i \qquad \textit{Eq. 5.1}$$

To obtain the total utility, a random error component is added to the systematic utility. This error component is added because the researcher cannot observe the total utility. Some explanatory variables and taste differences may not be observable to the researcher and measurement errors may be made (Ben-Akiva & Lerman, 1985). If we assume that this error component is logistically distributed, we can calculate the probability that the tour is ended with equation 5.2 (Ben-Akiva & Lerman, 1985).

$$\pi_{ET} = \frac{e^{U_{ET}}}{1 + e^{U_{ET}}} \qquad \textit{Eq. 5.2}$$

A choice is selected based on these probabilities with a Monte Carlo simulation. If the tour is ended, its shipments are labeled as 'allocated'; they cannot be allocated to other tours anymore. If the tour is not ended, a shipment is chosen to add to the tour, the SS procedure starts.

5.2.2 SELECT SHIPMENT SUB MODEL

The key aspect of the SS sub model is choice set formation. We need to form a choice set that consists of multiple shipments that may or may not be added to the tour. We define a universal, feasible, and consideration choice set (UC, FC, and CC), which we explain respectively in this section.

UNIVERSAL CHOICE SET

It is useful to distinguish first a universal choice set (UC) that consists of all alternatives present in the model (de Bok, 2007). As defined in the previous section, the universal set consists of all shipments that a carrier has reported for a day. These are all shipment alternatives that may be selected as the next shipment. Note that carriers do not have to report all transported shipments, but only those transported with requested vehicles. It is likely that the UC usually does not consist of every shipment that a carrier transported on a day.

FEASIBLE CHOICE SET

Since this UC may consist of many shipments that are not feasible or reasonable to add to the tour, it is appropriate to construct a feasible choice set (FC) from the UC. The following constraints lead to the formation of a feasible choice set:

- The shipment has not been allocated to another tour yet
- The shipment is not a concrete shipment
- The shipment has a proximity below 100 km
- The shipment does not bring the capacity utilization of the tour that it is added to above 110%

It is simply not possible to allocate the same shipment to two different vehicles. Furthermore, we already identified that concrete shipments are virtually never combined with other shipments in a tour. This 'concrete' constraint in the SS part leads to consistency between the ET and SS constraints.

The same proximity measure as discussed in the previous section (Figure 5.7) is calculated for each shipment in the UC, and those above 100 km are excluded. This proximity measure is calculated to proxy the 'additional cost' of each shipment. We did not calculate the actual additional cost, as this requires the construction of the tour sequence with and without the alternative shipment, which would be a very heavy calculation considering that the UC can consist of dozens to hundreds of shipments (Table 5.2).

Table 5.2. Frequency table of day + carrier observations by number of shipments.

Number of shipments per day + carrier	Frequency
1	1907
2	1640
3	1080
4	832
5-10	1864
10-15	733
15-20	442
20-25	390
25-30	218
30-35	189
35-40	179
40-45	167
45-50	134
50-100	713
100-150	218
150-200	56
200-250	36
250-300	54
300-350	48
350-400	28
400-450	9
>=500	36

The last constraint ensures that tours do not transport far more weight than allowed or possible. This constraint is set higher than 100%, as in Figure 4.14 we observed that for varying reasons tours with a capacity utilization that is a little over 100% are observed quite frequently. In combination with the capacity utilization constraint in the ET sub model of 100% (Section 5.2.1), we can still construct tours with a capacity utilization above 100%. For example, if a tour with a capacity utilization of 99% is not ended, then it is still possible to add a shipment that brings the capacity utilization to 105%.

If all shipments in UC violate one or more constraints, then the FC is an empty set and the tour is ended.

CONSIDERATION CHOICE SET

After the FC has been obtained, a consideration choice set (CC) is constructed. The CC consists of a randomly sampled subset of the FC. Each shipment in the FC has an equal probability of being sampled for the CC. It has been shown that consistent parameter estimates of the MNL are obtained with small well-sampled choice sets (Bovy, 2009; Prato, 2009). Doing so allows us to calculate more advanced explanatory variables within a reasonable amount of time.

MODELING THE SELECT SHIPMENT CHOICE

Once the CC has been constructed and explanatory variables of the alternative shipments have been obtained, the choice for one of these shipments can be modeled. For this purpose, the systematic utility (U_j) and the probability (π_j) of each alternative *j* is calculated. The systematic utility of an alternative is calculated in an MNL in the same way as in a BLR, see equation 5.1 in Section 5.2.1. When we assume that the error component that is added to the systematic utility of each alternative has a Gumbel distribution with mean zero and is independently and identically distributed over observations, then the probability of choosing an alternative shipment *j* can be calculated with equation 5.3, where *m* is the number of alternative shipments in the consideration choice set (Ben-Akiva & Lerman, 1985).

Eq. 5.3

$$\pi_j = \frac{e^{U_j}}{\sum_{j=1}^{m} e^{U_j}}$$

Again, with a Monte Carlo simulation, one of the alternative shipments is chosen based on these probabilities.

5.3 MODEL ESTIMATION STEPS

To estimate the parameters of the ET and SS sub models, the XML microdata is used. For this purpose, we need to obtain choice data from the raw XML microdata. In other words, the data needs to be transformed from 'a list of tours containing shipments' to 'a list of observed ET choices and SS choices with explanatory variables'. In the remainder of this section, we explain the process of obtaining choice data.

5.2.1 ESTIMATION OF THE END TOUR SUB MODEL

The ET sub model is used to predict the probability of ending a tour, which is the probability that a shipment is the last to be allocated to a tour. For this reason, we need to know for each shipment in the data whether it was the last shipment allocated to this tour or not. This is the dependent variable of the ET choice model.

The order of allocating shipments to a tour is a theoretical construct that is not directly observed from the data. However, we do know all the shipments that are part of a tour. Consequently, we know that no more shipments have been allocated to a tour, other than its listed shipments. The 'ET decision' is therefore 1 for each complete tour (i.e. a tour including all its listed shipments). For any sub tour (i.e. a tour with a subset of its listed shipments), we therefore also know that its 'ET decision' is 0. Each tour thus provides one observed 'ET=1' observation, and $\sum_{r=1}^{n-1} \frac{n!}{r!(n-r)!}$ 'ET=0' observations[2].

Listing all these possible sub tours would lead to an explosion of 'ET=0' observations. For a tour with seven shipments, 126 possible sub tours can be found. Instead, we propose to use the order in which shipments are listed in the data to generate the 'ET=0' observations. In the XML microdata, a variable called 'zendingvolgnummer' is present, which gives each shipment in a tour a number based on its order of listing in the data. For each shipment, we construct the tour sequence with all shipments listed in this tour up to this current shipment. Doing so allows us to calculate a sub tour's characteristics, which serve as explanatory variables for the observed 'ET=0' decision. This process provides (*n*-1) 'ET=0' observations per tour, a far more practical solution than listing all possible sub tours.

In Table 5.3, it is shown what this generated choice data for the ET sub model could look like. For example, the tour with Tour ID=12 contains four shipments and has three 'ET=0' observations. Explanatory variables are calculated for the (sub) tour that consists of all shipments listed up to the current one. If variable A is the tour duration, then its third row will contain the duration of the sub tour with shipments with Tour ID=12 and Zendingvolgnummer=1:3.

To remain consistent between the model as we will apply it (Figure 5.2) and the way we estimate its parameters, observations where the (sub) tour violates the constraints mentioned in Section 5.2.1 (i.e. concrete, proximity, capacity, duration) are excluded from model estimation.

Table 5.3. Pseudo choice data for estimation of the ET sub model.

Tour ID	Zendingvolgnummer	Choice ET	Variable A	Variable B	Variable C
12	1	0	…	…	…
12	2	0	…	…	…
12	3	0	…	…	…
12	4	1	…	…	…
13	1	0	…	…	…
13	2	0	…	…	…
13	3	1	…	…	…

5.3.2 Estimation of the Select Shipment sub model

To estimate the SS choice model, we need a choice set that consists of an observed added shipment and one or multiple shipments that were not added to the tour.

Each tour provides *n-1* sub tours, as discussed in the previous section. For each sub tour, we take the next listed shipment as the observed added shipment, since we know with certainty that this shipment was part of the same tour. We could take any shipment listed below this shipment as an observed added shipment, as long as they are still part of the same tour. However, similar to what we described for the ET choice data generation, this would lead to an unnecessarily large number of observations. In addition, the process of choice data generation would be less consistent and structured.

An observed choice, as well as unchosen alternatives, is needed. To maximize the consistency between the tour formation model as we will apply it (in Figure 5.2) and the way we estimate it on the observed data, we sample several shipments from the UC (i.e. shipments transported by the same carrier on the same day). In doing so, we only sample shipments that are not part of the observed tour. Shipments that violate the constraints listed in Section 5.2.2 (concrete, proximity, capacity) do not end up in the generated SS choice sets. Full consistency between application and estimation cannot be reached. In model application we run out of shipments to

[2] *n* = number of shipments listed in this tour, *r* = number of elements of sub group (from 1 to *n-1*)

sample from as we allocate them to different tours, while in choice data generation for estimation the only allocated shipments we cannot sample from are those part of the current observed tour.

In Table 5.4, a pseudo example of SS choice data is shown. In this example, two unchosen alternatives are sampled. Note how the choice is always 0, as this is how the observed next shipment is coded here. For the last shipment in the tour, there is no observed choice, as no other shipments were allocated to the tour after this one.

Table 5.4. Pseudo choice data for estimation of the SS sub model.

Tour ID	Zendingvolgnummer	Choice SS	Variable A *(0: observed shipment)*	Variable A *(1: sampled shipment 1)*	Variable A *(2: sampled shipment 2)*	Variable B *(0: observed shipment)*	Variable B *(1: sampled shipment 1)*	Variable B *(2: sampled shipment 2)*
12	1	0	…	…	…	…	…	…
12	2	0	…	…	…	…	…	…
12	3	0	…	…	…	…	…	…
12	4	-						
13	1	0	…	…	…	…	…	…
13	2	0	…	…	…	…	…	…
13	3	-						

6 ESTIMATION RESULTS

In this chapter, the parameter estimates of the choice models are reported and interpreted. Section 6.1 to Section 6.3 discuss the ET first shipment, ET later shipments, and SS model respectively. In Section 6.4, we investigate the changes in results when a subset of carriers is used for estimation.

As noted in Figure 4.6, a very large portion of tours is ended already after the first shipment. Different effects may explain the probability of ending the tour after first shipment, which is why the ET model is estimated separately for the first shipment and for later shipments.

The data used to estimate the choice models is kept separate from the data for validation. For the choice models in Section 6.1 to Section 6.3, fifty percent of the day + carrier combinations are selected at random with an equal probability. Similar statistics are found regarding number of stops, vehicle type, and goods types in the validation and estimation data set (Appendix I). While we have an enormous number of observed shipments, the number of carriers in the data is quite limited and these carriers do not form a representative sample of the population of firms performing freight transportation in the Netherlands, as identified in Chapter 4. For this reason, we have chosen to divide the data based on day + carrier combinations and not on carriers. With such a small and unrepresentative group of carriers, it was deemed more important to improve the model that we estimate by maximizing the number and diversity of carriers whose tours are used for it, than to test how well a model estimated for a subset of carriers is able to reproduce tours of other carriers in the data.

However, after these models are estimated we do also test how parameters change when only a subset of carriers is selected for estimation. This allows us to test the robustness of the estimates and use these estimates in the validation study in order to formulate more solid statements regarding the external validity of the model.

6.1 ESTIMATES END TOUR FIRST SHIPMENT

The explanatory variables of the ET sub model can be divided into three categories: instrumental variables, location variables, and vehicle/goods type variables. Table 6.1 shows the process of adding instrumental variables while Table 6.2 shows the successive process of adding location/vehicle/goods type variables. Instrumental variables were added first, these reflect the actual decision-making process of a transportation planner and are most intuitive. Variables are added consecutively to the model and removed when the p-value is larger than 0.05, when multicollinearity issues arise, or when interpretation is difficult. For non-categorical variables we test non-linear effects with the square and the square root of the variable. A non-linear specification is chosen if it leads to a higher pseudo-R^2 and when the non-linearity is clearly interpretable.

The most right column in Table 6.2 shows the final model. All parameters have a Tolerance much larger than 0.10 and a VIF far below 10 (Appendix B). Consequently, we can conclude that multicollinearity is no issue in the ET first shipment model.

In the remainder of this section we discuss the variables included in the final model, the variables that were tested but excluded from the final model, and the estimates with different model specifications.

INSTRUMENTAL VARIABLES

Tour duration is calculated with the two tour sequence algorithms mentioned in Section 5.2.1. Due to data availability, tour duration does not include dwelling time at stops. In the ET first shipment model we take the square root of tour duration. It shows a negative parameter sign, which can be interpreted as follows: after the first shipment, the probability of ending the tour is higher when the tour has a lower duration. This is in line with Figure 4.25, where we saw that tours with two stops have a particularly low distance on average. We reasoned that for short-distance shipments the desire to plan simple tours plays a large role (Nuzzolo et al., 2012) and efficiency gains of grouping shipments are smaller. Because we take the square root of tour duration, the effect is stronger for lower durations. Shipments that require a very long tour, whether this is two hours or four hours, are more or less considered equally as distant shipments for which a direct tour is not preferred.

This effect is in contrast with the findings of the interview with the transportation planner, whose practice is to add more shipments to a tour that lasts very shortly after the first stop. Different carriers may experience different constraints in tour formation. If the vehicle fleet size and a tour duration of nine hours are the most determining constraints for a carrier, then it may be more logical to add more shipments to tours that last shortly. Perhaps for the large third-party carriers in the XML microdata the vehicle fleet size is a constraint that plays a role in tour formation less often.

The squared capacity utilization shows a positive parameter. As a larger share of the vehicle capacity (in kilograms) is used, the probability of ending the tour is higher. Intuitively this makes a lot of sense. It is also in line with the findings of the interview, where we identified that the vehicle capacity is not necessarily only a constraint. Maximizing the capacity utilization can also be considered a strategy for the transportation planner to minimize transportation costs. Since capacity utilization could only be obtained with respect to weight and the interview underlined the importance of volume as a constraining factor, many of the location and vehicle/goods type variables are likely to reflect differences in volume. The square root implies that not until we nearly reach the capacity, the probability of ending the tour becomes much higher, because sparsely filled vehicles are very inefficient.

LOCATION VARIABLES

Two types of location variables are included in the final model: logistical zones and urban zones. Two types of logistical zones are distinguished: port transshipment nodes and distribution centers. Implicitly we assume that when a zone that is identified as a 'distribution center zone' is visited, the sending or receiving firm is indeed a distribution center. In Section 4.3.2 we discussed how these zonal variables were obtained. Because a tour can consist of multiple locations, we distinguish tours that visit any zone of a certain type and tours that do not.

When a tour visits a port transshipment node *(any port)*, the probability of ending the tour is higher than when it does not visit a port transshipment node. In contrast, when a tour visits a distribution center *(any loading DC; any unloading DC)*, the probability of ending the tour is lower. This effect is stronger when goods are loaded than when they are unloaded at a distribution center. As discussed in Section 4.3.3, these differences can be attributed to shipment sizes and vehicle capacities in different marketing channels (volume, as we control for weight capacity utilization), the number of available (similar) shipments at these locations, and the operations inside distribution centers.

Tours that visit an urban zone are more likely to add more shipments *(any urban zone)*. The tour may require the driver to enter the city from some distant rural location. Entering a city by road can be a very time- and energy-consuming task, which is why carriers may prefer to deliver/pick-up other shipments in the city too. Other studies also mention that tours in urban areas tend to visit multiple stops (Hunt & Stefan, 2007; Khan & Machemehl, 2017). Due to a high density of commercial activity, it is likely that multiple shipments in this city can be picked up efficiently in one tour.

VEHICLE/GOODS VARIABLES

The same vehicle types as shown in Figure 4.16 are used as a categorical variable in the model.

- 0: Truck
- 1: Truck + trailer
- 2: Tractor + trailer
- 3: Other/special (e.g. van)

Because the integration with a vehicle type choice model is outside of the scope of this research, we simply assume that the vehicle type used for the tour is the vehicle type reported in the data for the first allocated shipment. Shipments that were actually transported with another vehicle type can still be added to the tour.

The estimated parameters are consistent with the percentage of direct tours for these vehicle types in Figure 4.16, and can be related to the ease of loading/unloading and volume capacity of these vehicles. Compared to the reference category (*2: tractor + trailer* and *3: special/other),* tours driven with a truck + trailer (*1*) have quite a substantially higher probability of being ended after the first shipment. Truck + trailers are less practical for multiple-stop tours as the trailer must be uncoupled to unload goods from the truck.

Substantial differences in parameters are found for different goods types, which can be explained by their varying volumes, dispersion of demand, ease of loading/unloading, and rigidity of goods combination restrictions. We calculate the goods type of the tour as the NSTR code (1-digit) with the highest transported weight in the tour. In the first shipment model, the tour only includes one shipment, so it is simply the NSTR code of that shipment. Because of the low number of observations and similar goods types (fuels/oils and metals), NSTR 2 to 5 are grouped into one category. The reference category is NSTR 9 (other). Consistent with the percentage of direct tours we observed for different goods types in Figure 4.19, tours transporting agricultural products & livestock (*0*) and foodstuffs and fodder (*1*) have a lower probability of being ended after the first shipment, while tours transporting other products, especially construction materials (*6*), are more likely to be ended after the first shipment. The parameter for manure (*7*) is positive, in contrast with Figure 4.19, which may be due to the low number of observations and controlling for other variables such as vehicle type.

Excluded variables

Tables 6.1 and 6.2 show the process of adding and removing variables to arrive at the final model. In this section, we briefly discuss some of the variables that were not included.

Both tour duration and tour distance were tested as linear and non-linear (square root and squared) variables. Tour duration was chosen, for we expect this to reflect better the tour length that drivers and planners actually experience. Two locations may feel equally far away when they can be reached within the same amount of time, even if one is located further away in terms of distance. The square root of tour duration also leads to a higher pseudo-R^2.

The proximity of the nearest shipment, as explained in Section 5.2.1, was also tested. While it improves the model fit, it was excluded because of multicollinearity issues (Appendix B). In Table 6.1, we can see that the parameter for tour duration changes with a factor three to four when proximity is included. It is still included as a constraint that can end the tour, though.

Several variables reflecting the logistical nodes were tested. First both (port and distribution center) were added with an operationalization that does not consider loading or unloading locations separately *(any port; any DC)*, and then with an operationalization that does consider both locations separately *(any loading port, any unloading port, any loading DC, any unloading DC)*. Next, we also tested variables *(cat: any port; cat: any DC)* that distinguish four categories: (*0*) no port, (*1*) only loading port, (*2*) only unloading port, and (*3*) loading and unloading port. This was done to test if category *3* leads to substantially different effects compared to adding up *any loading* and *any unloading* and a better model fit. The categorical variables do not improve the model fit a lot, have higher standard errors, and category *3* does not show substantially different effects. For ports, we chose *any port*, as only small differences in the parameters for *any loading port* and *any unloading port* are observed, which can also not be clearly explained. For distribution centers, we chose *any loading DC* and *any unloading DC*, for the effects differ more and can be explained well.

We also tested if tours that visit retail zones have a significantly lower or higher probability of being ended. In advance, a negative parameter sign was expected, as we observed more stops for tours visiting retail zones in Figure 4.25. As the parameter is not significant at the 0.05 level when vehicle type is added to the model, it was not included in the final model.

Finally, a variable that distinguishes whether the tours transports palletized goods was tested. From the interview we found out that goods of different customers can easily be stacked onto the same pallet, even in the order of delivery. For this reason, tours with more shipments/stops were expected. Because the 'unknown' category (-*1*) is so highly different from the 'other' (*0*) category, the variable is not considered to be sufficiently reliable and meaningful. Furthermore, the parameter shows an effect in contrast to the explanation above.

DIFFERENT MODEL SPECIFICATIONS

The ET first shipment model was estimated with different model specifications (Table 6.3) to check the robustness of the estimates of the final model and to be able to use different model specifications in the validation in Chapter 7. We vary in the tour sequence algorithm that is used and the constraints that are applied (concrete, vehicle capacity, tour duration, proximity). In Table 6.4, model specifications A to G are listed.

The estimated parameters are quite robust with varying model specifications. Other than for the NSTR 7 parameter, no sign changes are found and most betas do not change a lot. Most importantly, the capacity utilization parameter increases when it is constrained to 100% (C to D), as there are fewer outliers to bias the effects. Furthermore, the NSTR 8 parameter decreases when concrete shipments are removed (A to B), as concrete shipments fall under this category and are virtually only found in tours with one shipment. The model fit is the highest when concrete shipments are included. This is misleading, though, as inclusion of concrete shipments increases the explanatory power of the ET choice model but not of the whole tour formation model in which we incorporate constraints that can also end the tour.

Table 6.1. Process of adding instrumental variables to the ET first shipment model with specification F (see Table 6.4). Cells below the bold line show the beta and standard error. Estimates with a p-value higher than 0.05 are grey.

$R^2_{Nagelkerke}$	0.026	0.044	0.011	0.020	0.037	0.007	0.178	0.170	0.196	0.250
-2 LL	69775	68955	70482	70082	69294	70663	62331	62724	61409	58520
Percentage correct	82.0	82.0	82.4	82.0	82.0	82.0	82.2	82.5	82.0	82.4
N	75255	75255	75255	75255	75255	75255	75255	75255	75255	75255
Constant	1.906 (0.015)	2.253 (0.021)	1.676 (0.012)	1.828 (0.015)	2.134 (0.019)	1.642 (0.012)	1.536 (0.022)	0.950 (0.026)	1.793 (0.021)	2.545 (0.029)
Tour duration [h]	-0.990 (0.028)									
$\sqrt{\textit{Tour duration}}$		-1.347 (0.028)					-1.629 (0.031)	-1.578 (0.032)	-1.658 (0.031)	-6.146 (0.096)
Tour duration^2			-0.627 (0.027)							
Tour distance [km]				-0.009 (0.000)						
$\sqrt{\textit{Tour distance}}$					-0.124 (0.003)					
Tour distance^2						-0.000 (0.000)				
Weight/capacity							3.328 (0.048)			
$\sqrt{\textit{Weight/capacity}}$								3.153 (0.044)		
(Weight/capacity)^2									5.221 (0.081)	5.848 (0.085)
Proximity nearest shipment [km]										0.049 (0.001)

Table 6.2. Process of adding other variables to the ET first shipment model with specification F (see Table 6.4). Cells below the bold line show the beta and standard error. Estimates with a p-value higher than 0.05 are grey.

$R^2_{Nagelkerke}$	0.328	0.324	0.330	0.330	0.334	0.334	0.346	0.411	0.442	0.442
-2 LL	50379	54399	54035	54049	53841	53834	53123	49197	47313	47315
Percentage correct	81.8	82.2	81.9	81.9	82.0	82.0	82.9	83.7	84.8	84.8
N	75255	75255	75255	75255	75255	75255	75255	75255	75255	75255
Constant	1.535 (0.024)	1.525 (0.025)	1.589 (0.026)	1.602 (0.025)	1.633 (0.026)	1.629 (0.026)	1.125 (0.037)	1.660 (0.027)	1.685 (0.029)	1.684 (0.029)
$\sqrt{\textit{Tour duration [h]}}$	-1.702 (0.033)	-1.623 (0.033)	-1.801 (0.036)	-1.793 (0.033)	-1.771 (0.034)	-1.777 (0.034)	-1.847 (0.035)	-1.946 (0.036)	-1.693 (0.037)	-1.698 (0.037)
(Weight/capacity)^2	4.900 (0.082)	4.862 (0.082)	4.905 (0.083)	4.910 (0.083)	4.900 (0.083)	4.912 (0.083)	5.332 (0.088)	6.097 (0.093)	5.465 (0.102)	5.471 (0.102)
any port	2.184 (0.035)			2.141 (0.035)	2.136 (0.035)	2.143 (0.035)	2.275 (0.036)	1.714 (0.036)	1.585 (0.037)	1.588 (0.037)

any DC	-0.359									
	(0.022)									
any loading port		1.568								
		(0.043)								
any unloading port		1.415								
		(0.045)								
any loading DC		-0.378		-0.400	-0.384	-0.384	-0.353	-0.513	-0.576	-0.578
		(0.023)		(0.023)	(0.023)	(0.023)	(0.023)	(0.025)	(0.026)	(0.026)
any unloading DC		-0.128		-0.158	-0.149	-0.155	-0.130	-0.317	-0.473	-0.475
		(0.024)		(0.023)	(0.023)	(0.023)	(0.024)	(0.025)	(0.026)	(0.026)
cat: any port [0]										
[1]			2.084							
			(0.058)							
[2]			2.037							
			(0.064)							
[3]			2.259							
			(0.056)							
cat: any DC [0]										
[1]			-0.345							
			(0.030)							
[2]			-0.098							
			(0.032)							
[3]			-0.570							
			(0.030)							
any urban zone					-0.512	-0.510	-0.447	-0.599	-0.462	-0.461
					(0.035)	(0.035)	(0.036)	(0.038)	(0.038)	(0.038)
any retail zone						0.212	0.137	0.128	-0.083	
						(0.045)	(0.046)	(0.048)	(0.052)	
any pallets [0]										
[-1]							0.416			
							(0.030)			
[1]							0.913			
							(0.035)			
vehicle type [0]										
[1]								-1.174	-1.295	-1.295
								(0.037)	(0.039)	(0.039)
[2]								2.159	1.847	1.850
								(0.047)	(0.049)	(0.049)
[3]										
NSTR tour [0]									-0.739	-0.736
									(0.047)	(0.047)
[1]									-0.659	-0.659
									(0.032)	(0.032)
[2-5]									1.497	1.495
									(0.337)	(0.337)
[6]									1.462	1.452
									(0.058)	(0.058)

	0.710	0.713
[7]	(0.253)	(0.253)
	0.582	0.583
[8]	(0.045)	(0.045)
[9]		

Table 6.3. Estimation results of ET (first shipment) model with different model specifications.

Specification	A	B	C	D	E	F	G
$R^2_{Nagelkerke}$	0.477	0.335	0.335	0.424	0.424	0.442	0.439
-2 LL	68704.864a	70949,509a	70949,509a	61362,447a	61362,447a	47315,066a	55185,542a
Percentage correct	92.3	85.6	85.6	85.2	85.8	84.8	85.3
N	194691	109207	109207	102654	102654	75255	90000
Constant	1.509 (0.024)	1.442 (0.024)	1.442 (0.024)	1.281 (0.025)	1.281 (0.025)	1.684 (0.029)	1.473 (0.027)
$\sqrt{\textit{Tour duration [h]}}$	-0.750 (0.024)	-0.561 (0.023)	-0.561 (0.023)	-0.772 (0.025)	-0.772 (0.025)	-1.698 (0.037)	-1.112 (0.030)
$(\textit{Weight/capacity})^2$	0.894 (0.032)	0.896 (0.030)	0.896 (0.030)	6.185 (0.095)	6.185 (0.095)	5.471 (0.102)	6.022 (0.098)
any port	1.765 (0.033)	1.766 (0.033)	1.766 (0.033)	1.509 (0.034)	1.509 (0.034)	1.588 (0.037)	1.484 (0.035)
any loading DC	-0.339 (0.021)	-0.397 (0.021)	-0.397 (0.021)	-0.452 (0.022)	-0.452 (0.022)	-0.578 (0.026)	-0.517 (0.024)
any unloading DC	-0.442 (0.021)	-0.314 (0.021)	-0.314 (0.021)	-0.362 (0.023)	-0.362 (0.023)	-0.475 (0.026)	-0.450 (0.024)
any urban zone	-0.264 (0.031)	-0.428 (0.033)	-0.428 (0.033)	-0.580 (0.035)	-0.580 (0.035)	-0.461 (0.038)	-0.605 (0.036)
vehicle type [0]	-	-	-	-	-	-	-
[1]	-0.452 (0.033)	-1.060 (0.031)	-1.060 (0.031)	-1.350 (0.036)	-1.350 (0.036)	-1.295 (0.039)	-1.370 (0.037)
[2]	1.447 (0.041)	1.432 (0.041)	1.432 (0.041)	1.943 (0.044)	1.943 (0.044)	1.850 (0.049)	1.980 (0.045)
[3]	-	-	-	-	-	-	-
NSTR tour [0]	-0.357 (0.037)	-0.346 (0.036)	-0.346 (0.036)	-0.556 (0.040)	-0.556 (0.040)	-0.736 (0.047)	-0.881 (0.044)
[1]	-0.685 (0.025)	-0.714 (0.025)	-0.714 (0.025)	-0.834 (0.027)	-0.834 (0.027)	-0.659 (0.032)	-0.808 (0.029)
[2-5]	2.143 (0.276)	2.225 (0.276)	2.225 (0276)	1.428 (0.297)	1.428 (0.297)	1.495 (0.337)	1.298 (0.324)
[6]	2.302 (0.046)	2.213 (0.046)	2.213 (0.046)	1.678 (0.048)	1.678 (0.048)	1.452 (0.058)	1.472 (0.051)
[7]	-0.120 (0.159)	0.090 (0.159)	0.090 (0.159)	-0.408 (0.193)	-0.408 (0.193)	0.713 (0.253)	-
[8]	3.251 (0.042)	1.250 (0.036)	1.250 (0.036)	0.508 (0.040)	0.508 (0.040)	0.583 (0.045)	0.530 (0.042)
[9]	-	-	-	-	-	-	-

Table 6.4. Description of tested model specifications for the ET model.

Specification	Tour sequence algorithm	Concrete excluded	'Weight/capacity > 1' excluded	'Tour duration > 9h' excluded	'Proximity > X' excluded
A	1	No	No	No	No
B	1	Yes	No	No	No
C	1 & 2	Yes	No	No	No
D	1 & 2	Yes	Yes	No	No
E	1 & 2	Yes	Yes	Yes	No
F	1 & 2	Yes	Yes	Yes	Yes (X=100)
G	1 & 2	Yes	Yes	Yes	Yes (X=150)

6.2 Estimates End Tour later shipments

The estimation process for ET later shipments choice model is reported in Table 6.5 (instrumental variables) and Table 6.6 (other variables) in a similar fashion as for the ET first shipments choice model. Most of the parameters of the ET first shipment choice model are present in this final model with similar betas. In this section, we discuss the parameters that differ most notably from the first model.

The final ET later shipments model is found in the most right column in Table 6.6. The Tolerance and VIF statistics indicate that there are no multicollinearity issues in the ET later shipments model (Appendix B).

EXPLANATORY VARIABLES

In contrast to the ET first shipment choice model, tour duration has a linear effect and a positive parameter. When a tour lasts longer with multiple shipments, the probability that it is ended is higher. When the tour consists of two or more shipments, the desire to construct a simple tour is not as present anymore and other effects play a role. Tours that last longer are more expensive due to labor and fuel costs (Anand et al., 2014). As the tour lasts longer, the chance of violating working hour constraints due to unexpected delays also increases. For this reason, carriers prefer to construct tours that do not last close to the maximum work shift duration.

Proximity is included in the ET later shipments model, it does not lead to multicollinearity issues. The positive parameter can be interpreted as follows: when the nearest shipment is located further away from the tour, the probability of ending the tour is higher, because there are no remaining possibilities to extend this tour efficiently.

A variable not present in the first model is the number of stops, for it is always one or two for a tour with one shipment. In the later shipments model this is not the case, however. Besides categorical variables, we tested the square, the square root, the natural logarithm, and a linear parameter of number of stops. The natural logarithm was chosen because it leads to a high pseudo-R^2. The negative parameter indicates that the probability of ending the tour is lower when more stops are visited. As the tour has more stops, adding more shipments is not as unattractive anymore because the tour is more complex already. This effect is stronger for tours with few stops. A tour with fifteen stops and a tour with sixteen stops may not be considered as dissimilar as a tour with four stops and a tour with five stops.

The effects for the location variables are similar, but not as strong as in the first shipment model. As the tour visits more stops, there is a higher chance that any of the stops is located in one of the identified location types and effects may be less pronounced. Visiting a port transshipment node still increases the probability of ending the tour and visiting a distribution center to load goods or visiting an urban zone still decreases the probability of ending the tour. However, visiting a distribution center to unload goods now increases the probability of ending the tour, which is more in line with what we discussed in Section 4.3.3, i.e. goods unloaded at a distribution center are more likely to be large-volume shipments coming from producers.

Compared to tractor + trailers (*2*), the probability to end the tour is lower with trucks (0) and truck + trailers (*1*). While truck + trailers (*1*) are more likely to be used for direct tours, these vehicles tend to be used for tours that visit many stops when no direct tour is made.

Due to the low number of observations, NSTR categories 2-5 are included in the reference category in the later shipments model with NSTR 9. Quite a few effects for goods types are different than in the first shipment model. This might relate mostly to the NSTR 9 category. This category has a relatively large percentage of direct tours, but when a multiple-stop tour is constructed, it is more likely to have many shipments/stops.

Table 6.5. Process of adding instrumental variables to the ET later shipments model with specification F.

$R^2_{Nagelkerke}$	0.028	0.026	0.020	0.041	0.038	0.032	0.099	0.108	0.082	0.132	0.119	0.128	0.178	0.170	0.161	0.168	0.149	0.173
-2 LL	46516	46572	46758	46131	46227	464067	44349	44076	44902	43316	43742	43437	41834	42080	42390	42170	42781	41983
Percentage correct	77.4	77.2	77.4	77.5	77.6	77.5	78.0	78.0	78.0	78.5	78.6	78.3	78.2	78.2	78.5	78.4	78.6	78.5
N	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622	44622
Constant	-1.730 (0.021)	-2.175 (0.036)	-1.430 (0.014)	-1.806 (0.020)	-2.293 (0.035)	-1.478 (0.014)	-2.466 (0.028)	-3.126 (0.039)	-2.093 (0.024)	-2.805 (0.031)	-2.855 (0.033)	-2.671 (0.030)	-1.890 (0.042)	-1.811 (0.042)	-2.185 (0.037)	-0.882 (0.066)	-2.727 (0.032)	-1.390 (0.049)
Tour duration [h]	0.267 (0.009)						0.228 (0.010)	0.231 (0.010)	0.230 (0.009)	0.188 (0.010)	0.201 (0.010)	0.193 (0.010)	0.426 (0.013)	0.342 (0.011)	0.426 (0.013)	0.458 (0.013)	0.348 (0.012)	0.472 (0.013)
$\sqrt{\textit{Tour duration}}$		0.729 (0.026)																
Tour duration^2			0.039 (0.002)															
Tour distance [km]				0.004 (0.000)														
$\sqrt{\textit{Tour distance}}$					0.091 (0.003)													
Tour distance^2						0.000 (0.000)												
Weight/capacity							1.924 (0.042)			2.057 (0.043)	1.971 (0.043)	2.075 (0.043)	2.237 (0.045)	2.124 (0.044)	2.097 (0.044)	2.144 (0.044)	2.039 (0.043)	2.192 (0.044)
$\sqrt{\textit{Weight/capacity}}$								2.421 (0.052)										
(Weight/capacity)^2									1.721 (0.043)									
Proximity nearest shipment [km]										0.015 (0.000)			0.012 (0.001)	0.014 (0.000)	0.011 (0.001)	0.011 (0.001)	0.013 (0.000)	0.011 (0.001)
$\sqrt{\textit{Proximity}}$											0.108 (0.004)							
Proximity^2												0.000 (0.000)						
Number of stops [1-2]													-					
[3]													-1.092 (0.043)					
[4]													-1.455 (0.047)					
[5]													-1.618 (0.052)					
[6-10]													-1.849 (0.052)					
[>10]													-2.564 (0.093)					
Number of stops [1-2]														-				
[3-6]														-1.305 (0.040)				
[>6]														-1.878 (0.055)				

Number of stops	-0.211 (0.007)			
$\sqrt{\textit{Number of stops}}$		-1.121 (0.035)		
Number of stops^2			-0.011 (0.001)	
ln(Number of stops)				-1.320 (0.037)

Table 6.6. Process of adding other variables to the ET later shipments model with specification F.

$R^2_{Nagelkerke}$	0.190	0.197	0.197	0.190	0.190	0.190	0.232	0.291	0.292
-2 LL	41435	41202	41201	41426	41426	41417	39996	37920	37894
Percentage correct	78.7	79.4	79.4	78.6	78.6	78.8	79.4	82.1	81.8
N	44622	44622	44622	44622	44622	44622	44618	44618	44618
Constant	-1.396 (0.050)	-1.457 (0.050)	-1.463 (0.050)	-1.403 (0.050)	-1.403 (0.050)	-1.416 (0.050)	-1.401 (0.052)	-2.568 (0.063)	-2.526 (0.062)
Tour duration [h]	0.474 (0.013)	0.479 (0.013)	0.481 (0.013)	0.468 (0.013)	0.468 (0.013)	0.471 (0.013)	0.386 (0.014)	0.388 (0.014)	0.386 (0.014)
Weight/capacity	2.300 (0.046)	2.366 (0.047)	2.369 (0.047)	2.300 (0.046)	2.300 (0.046)	2.301 (0.046)	2.604 (0.048)	3.255 (0.056)	3.286 (0.057)
Proximity nearest shipment [km]	0.011 (0.001)	0.012 (0.001)	0.012 (0.001)	0.011 (0.001)	0.011 (0.001)	0.011 (0.001)	0.011 (0.001)	0.009 (0.001)	0.009 (0.001)
ln(Number of stops)	-1.249 (0.038)	-1.238 (0.038)	-1.243 (0.038)	-1.223 (0.039)	-1.223 (0.039)	-1.213 (0.039)	-0.995 (0.040)	-0.910 (0.043)	-0.911 (0.042)
any port	0.818 (0.040)			0.814 (0.040)	0.814 (0.040)	0.808 (0.040)	0.706 (0.041)	0.526 (0.047)	0.526 (0.047)
any DC	-0.338 (0.050)			-0.331 (0.028)	-0.331 (0.028)	-0.317 (0.029)	-0.445 (0.030)	-0.023 (0.033)	
any loading port		1.180 (0.061)							
any unloading port		-0.103 (0.062)							
any loading DC		-0.294 (0.033)							-0.191 (0.036)
any unloading DC		-0.090 (0.033)							0.094 (0.036)
cat: any port [0]			-						
[1]			1.203 (0.077)						
[2]			-0.078 (0.084)						
[3]			1.066 (0.055)						
cat: any DC [0]			-						
[1]			-0.261 (0.046)						

[2]	-0.051 (0.050)						
[3]	-0.382 (0.030)						
any urban zone		-0.089 (0.031)	-0.089 (0.031)	-0.093 (0.031)	-0.128 (0.031)	-0.148 (0.032)	-0.145 (0.032)
any retail zone				-0.100 (0.034)	-0.068 (0.035)	-0.043 (0.036)	
vehicle type [0]					-1.789 (0.057)	-1.956 (0.060)	-1.968 (0.061)
[1]					-0.285 (0.079)	-0.942 (0.088)	-0.954 (0.088)
[2]						-	-
[3]						-	-
NSTR tour [0]						2.256 (0.058)	2.226 (0.059)
[1]						0.882 (0.035)	0.871 (0.035)
[2-5]						-	-
[6]						0.570 (0.080)	0.556 (0.081)
[7]						-0.937 (0.325)	-1.105 (0.327)
[8]						1.542 (0.063)	1.517 (0.063)
[9]						-	-

DIFFERENT MODEL SPECIFICATIONS

The later shipments ET choice model was also tested with the model specifications described in Table 6.4. Again, most parameters are rather robust and incorporating a vehicle capacity constraint (C to D) increases the value of the capacity utilization parameter.

In contrast to the first shipment model, exclusion of concrete shipments has little impact (from A to B), as tours transporting a concrete shipment very rarely have more than one shipment. Furthermore, the second tour sequence algorithm now improves model fit substantially (from B to C). In the first shipment model the second tour sequence algorithm does not improve the model fit, as those tours always have only one loading and one unloading location, and both algorithms construct the same tour that goes from the loading to the unloading location. Usage of the second tour sequence algorithm also increases the value of the tour duration parameter. We have shorter tour durations on average when both tour sequences are constructed since we always pick the shortest of the two tour sequences.

Table 6.7. Estimation results of the ET later shipments model with different model specifications.

Specification	A	B	C	D	E	F	G
$R^2_{Nagelkerke}$	0.186	0.186	0.200	0.291	0.298	0.292	0.293
-2 LL	62022	62008	61363	41203	40846	37894	39933
Percentage correct	75.1	75.1	75.4	81.5	81.5	81.8	81.6
N	59869	59863	59863	47244	47115	44618	46336
Constant	-1.536 (0.045)	-1.538 (0.045)	-1.548 (0.044)	-2.472 (0.059)	-2.492 (0.059)	-2.526 (0.062)	-2.547 (0.060)
Tour duration [h]	0.237 (0.008)	0.237 (0.008)	0.382 (0.010)	0.250 (0.012)	0.356 (0.014)	0.386 (0.014)	0.364 (0.014)
Weight/capacity	0.880 (0.017)	0.880 (0.017)	0.851 (0.017)	3.313 (0.055)	3.290 (0.055)	3.286 (0.057)	3.285 (0.055)
Proximity nearest shipment [km]	0.007 (0.000)	0.007 (0.000)	0.005 (0.000)	0.008 (0.000)	0.007 (0.000)	0.009 (0.001)	0.008 (0.000)
Ln(# stops)	-0.592 (0.031)	-0.592 (0.031)	-0.702 (0.030)	-0.752 (0.039)	-0.847 (0.040)	-0.911 (0.042)	-0.841 (0.041)
any port	0.420 (0.037)	0.423 (0.037)	0.453 (0.038)	0.487 (0.045)	0.522 (0.045)	0.526 (0.047)	0.545 (0.046)
any loading DC	-0.007 (0.028)	-0.008 (0.028)	-0.011 (0.028)	-0.190 (0.034)	-0.191 (0.034)	-0.191 (0.036)	-0.179 (0.035)
any unloading DC	0.305 (0.028)	0.307 (0.028)	0.275 (0.029)	0.107 (0.034)	0.093 (0.034)	0.094 (0.036)	0.078 (0.035)
any urban zone	0.021 (0.024)	0.021 (0.024)	0.044 (0.024)	-0.218 (0.031)	-0.184 (0.031)	-0.145 (0.032)	-0.175 (0.032)
vehicle type [0]	-1.279 (0.033)	-1.280 (0.033)	-1.188 (0.033)	-2.001 (0.057)	-1.981 (0.058)	-1.968 (0.061)	-1.968 (0.058)
[1]	-0.556 (0.058)	-0.562 (0.058)	-0.507 (0.058)	-1.016 (0.084)	-0.980 (0.084)	-0.954 (0.088)	-1.003 (0.086)
[2]	-	-	-	-	-	-	-
[3]	-	-	-	-	-	-	-
NSTR tour [0]	1.641 (0.049)	1.642 (0.049)	1.692 (0.049)	2.121 (0.055)	2.178 (0.055)	2.226 (0.059)	2.203 (0.056)
[1]	0.069 (0.025)	0.069 (0.025)	0.080 (0.025)	0.895 (0.033)	0.873 (0.033)	0.871 (0.035)	0.873 (0.033)
[2-5]	-	-	-	-	-	-	-
[6]	0.478 (0.060)	0.479 (0.060)	0.524 (0.061)	0.536 (0.077)	0.525 (0.077)	0.556 (0.081)	0.538 (0.078)
[7]	-2.240 (0.215)	-2.242 (0.215)	-2.157 (0.215)	-2.166 (0.226)	-2.115 (0.226)	-1.105 (0.327)	-1.702 (0.289)
[8]	0.994 (0.051)	0.986 (0.051)	0.971 (0.051)	1.418 (0.061)	1.397 (0.061)	1.517 (0.063)	1.468 (0.060)
[9]	-	-	-	-	-	-	-

6.3 Estimates Select Shipment

Explanatory variables

In the final SS model, in Table 6.8 on the right, three variables are present. All three variables can be considered instrumental, they reflect the actual decision-making process of the transportation planner.

First, let us consider the additional generalized cost. Its negative parameter implies that a shipment in the consideration choice set has a lower probability of being selected as the next shipment if it adds more generalized cost to the tour. This reflects the desire of carriers to minimize costs by constructing efficient tours. Generalized cost is a weighted sum of the additional travel time and distance, where the weights reflect the costs that carriers make for each driven hour and kilometer. These weights are obtained from an estimation report for the Dutch freight model BasGoed (Significance, 2018) and are based on, for example, fuel and labor costs. The costs per hour are €45.12, while the costs per kilometer are €0.45. To calculate the additional time and distance of a shipment alternative, the tour sequence is constructed with and without the shipment alternative, and the duration and distance of these two tours are obtained with the skim matrices discussed in Section 4.2.

The second explanatory variable is the additional number of stops. This variable can take on values zero, one, or two, depending on how many locations an alternative shipment has in common with the constructed tour so far. Controlled for the additional generalized cost, a shipment that requires the tour to visit two additional locations has a lower probability of being added to the tour than a shipment that requires only one or no additional stop location. A shipment that adds more stops to the tour adds more complexity to the tour and might require more additional time for loading/unloading and leaving/entering the premises of customers.

As we observed in Figure 4.15 that multiple-stop tours usually contain only goods of the same NSTR category, a third variable was tested that reflects whether the alternative shipment has the same NSTR goods type as that of the tour as constructed so far (i.e. NSTR with maximum total weight in the tour). This variable shows a strongly positive parameter and improves the pseudo-R^2 of the model with only additional generalized cost and additional number of stops (from 0.151 to 0.187). Restricted goods combinations are the key explanation for this effect.

Excluded variables

Additional time and distance are not included in the final model, as additional generalized cost reflects both variables in one variable.

The additional number of stops was also tested as a categorical variable for non-linear effects. It was expected that two additional stops have a disproportionally higher disutility than one additional stop, for we observe a substantial share of distribution and collection tours (where all shipments have one location in common) (Figure 4.10). This is indeed the case; however, one additional stop has a positive utility compared to no additional stop with such a categorical variable. As this is a very counterintuitive and inexplicable effect, we decided to follow the simple linear variable.

Table 6.8. Estimation process of SS sub model with SS model specification I (see Table 6.10).

$R^2_{McFadden}$	0.075	0.062	0.070	0.151	0.214	0.187
LL	-71910	-72936	-72353	-65060	-61008	-63256
N	43409	43409	43409	43409	43409	43409
additional time [h]	-1.254 (0.013)					
additional distance [km]		-0.012 (0.000)				
additional generalized cost [€]			-0.014 (0.000)	-0.005 (0.000)	-0.007 (0.000)	-0.005 (0.000)
additional number of stops				-1.089 (0.010)		-1.039 (0.010)
cat: additional # stops [1]					0.114 (0.016)	
cat: additional # stops [2]					-2.248 (0.023)	
Same NSTR						2.313 (0.038)

Table 6.9. Estimation results of the SS sub model with different model specifications.

Specification	A	B	C	D	E	F	G	H	I	J	K	L
$R^2_{McFadden}$	0.543	0.454	0.422	0.429	0.189	0.170	0.277	0.251	0.187	0.169	0.277	0.249
LL	-18440	-54204	-57312	-49486	-63109	-73839	-59403	-73614	-63256	-73929	-59413	-73834
N	58184	55376	55376	48396	43409	37112	45851	41001	43409	37112	45851	41001
additional generalized cost [€]	-0.014 (0.000)	-0.015 (0.000)	-0.016 (0.000)	-0.015 (0.000)	-0.005 (0.000)	-0.005 (0.000)	-0.010 (0.000)	-0.009 (0.000)	-0.005 (0.000)	-0.005 (0.010)	-0.009 (0.000)	-0.010 (0.000)
additional number of stops	-0.932 (0.017)	-0.912 (0.011)	-1.023 (0.011)	-1.180 (0.012)	-1.043 (0.010)	-1.187 (0.010)	-1.150 (0.011)	-1.189 (0.010)	-1.039 (0.010)	-1.088 (0.010)	-1.160 (0.011)	-1.176 (0.010)
Same NSTR	1.540 (0.046)	2.047 (0.037)	2.095 (0.038)	2.109 (0.040)	2.332 (0.038)	2.759 (0.042)	2.235 (0.038)	2.657 (0.041)	2.313 (0.038)	2.712 (0.042)	2.218 (0.038)	2.627 (0.041)

Table 6.10. Description of tested model specifications for the SS model.

Specification	Choice set size	Tour sequence algorithm	'Weight/capacity > 1.1' excluded	'Proximity > X' excluded	Concrete excluded
A	2	1	No	No	No
B	6	1	No	No	No
C	6	1 & 2	No	No	No
D	6	1 & 2	Yes	No	No
E	6	1 & 2	Yes	Yes (X=100)	No
F	11	1 & 2	Yes	Yes (X=100)	No
G	6	1 & 2	Yes	Yes (X=150)	No
H	11	1 & 2	Yes	Yes (X=150)	No
I	6	1 & 2	Yes	Yes (X=100)	Yes
J	11	1 & 2	Yes	Yes (X=100)	Yes
K	6	1 & 2	Yes	Yes (X=150)	Yes
L	11	1 & 2	Yes	Yes (X=150)	Yes

DIFFERENT MODEL SPECIFICATIONS

To test the robustness of the estimated model and use different specification in the validation, we estimated the model with different specifications (Table 6.9). In Table 6.10 these specifications of the SS model are listed.

We find a lower pseudo-R^2 when the choice set is larger (we tested choice sets with one, five, and ten unchosen sampled alternatives) and when proximity constraints are included. This is the case because we construct a better choice set with more reasonably considered alternatives. Therefore, the model has more difficulty predicting the observed choice in this sampled choice set. The formation of good choice sets and prediction of the right choice in this choice set are both important aspects for the general predictive power of the whole model. Consequently, a lower pseudo-R^2 does not necessarily imply a worse model, and we will test this in the validation chapter.

Inclusion of proximity constraints (D to E) and making these constraints more strict (G to E) also makes the additional generalized cost parameter less negative, as the choice set contains alternatives with a lower generalized cost on average.

The estimated parameters are quite robust with different model specifications. Due to randomness in the process of choice set formation for the SS choice model, we also estimated the parameters with different runs, which indicate a high robustness (Appendix C).

6.4 ESTIMATION WITH A SUBSET OF CARRIERS

The estimations results reported up to this point are based on data including fifty percent of the *day + carriers*. Although different days are used for estimation and validation, data of all carriers is present in both the estimation and validation data. When we select only fifty percent of the *carriers* for the estimation data instead, the results in Table 6.11 to Table 6.13 are obtained. None of the parameters in the three models change sign, except for *any unloading DC* in the ET later shipments models, although it is still less negative than the parameter for *any loading DC*. While some parameters do change substantially, overall we can conclude that the direction and magnitude of the effects are similar for different carriers. Consequently, our interpretation of these effects can be generalized towards the population of the data sample: carriers with advanced transport planning software. In Section 7.1.3, we will use these estimates to investigate the external validity of our model.

Table 6.11. Estimations results of the ET first shipment model for a subset of carriers. ET specification F is used (see Table 6.4).

Estimation sample	50% of day + carriers	50 % of carriers
$R^2_{Nagelkerke}$	0.442	0.570
-2 LL	47315	55866
Percentage correct	84.8	87.8
N	75255	99273
Constant	1.684 (0.029)	1.681 (0.024)
$\sqrt{\textit{Tour duration [h]}}$	-1.698 (0.037)	-2.403 (0.034)
(Weight/capacity)^2	5.471 (0.102)	5.258 (0.088)
any port	1.588 (0.037)	2.354 (0.037)
any loading DC	-0.578 (0.026)	-0.942 (0.025)
any unloading DC	-0.475 (0.026)	-0.765 (0.025)
any urban zone	-0.461 (0.038)	-0.499 (0.037)
vehicle type [0]	-1.295 (0.039)	-1.684 (0.039)
[1]	1.850 (0.049)	2.508 (0.047)
[2]	-	-
[3]	-	-
NSTR tour [0]	-0.736 (0.047)	-0.271 (0.037)
[1]	-0.659 (0.032)	-0.672 (0.037)
[2-5]	1.495 (0.337)	1.121 (0.311)
[6]	1.452 (0.058)	2.253 (0.048)
[7]	0.713 (0.253)	0.878 (0.237)
[8]	0.583 (0.045)	1.821 (0.053)
[9]	-	-

Table 6.12. Estimations results of the ET later shipments model for a subset of carriers. ET specification F is used (see Table 6.4).

Estimation sample	50% of day + carriers	50 % of carriers
$R^2_{Nagelkerke}$	0.292	0.186
-2 LL	37894	62022
Percentage correct	81.8	75.1
N	44618	59869
Constant	-2.526 (0.062)	-2.516 (0.054)
Tour duration [h]	0.386 (0.014)	0.449 (0.012)
Weight/capacity	3.286 (0.057)	3.122 (0.048)
Proximity [km]	0.009 (0.001)	0.008 (0.000)
Ln(# stops)	-0.911 (0.042)	-0.828 (0.036)
any port	0.526 (0.047)	0.450 (0.040)
any loading DC	-0.191 (0.036)	-0.281 (0.031)
any unloading DC	0.094 (0.036)	-0.150 (0.031)
any urban zone	-0.145 (0.032)	-0.036 (0.027)
vehicle type [0]	-1.968 (0.061)	-2.354 (0.059)
[1]	-0.954 (0.088)	-0.845 (0.085)
[2]	-	-
[3]	-	-
NSTR tour [0]	2.226 (0.059)	2.182 (0.045)
[1]	0.871 (0.035)	0.546 (0.031)
[2-5]	-	-
[6]	0.556 (0.081)	0.396 (0.068)
[7]	-1.105 (0.327)	-0.888 (0.244)
[8]	1.517 (0.063)	1.168 (0.062)
[9]	-	-

Table 6.13. Estimations results of the SS model for a subset of carriers. SS specification I is used (see Table 6.10).

Estimation sample	50% of day + carriers	50 % of carriers
$R^2_{McFadden}$	0.187	0.156
LL	-63256	-101620
N	43409	67181
Additional generalized cost [€]	-0.005 (0.000)	-0.006 (0.000)
Additional number of stops	-1.039 (0.010)	-0.918 (0.008)
Same NSTR	2.313 (0.038)	2.176 (0.031)

7 VALIDATION AND SENSITIVITY ANALYSIS

7.1 VALIDATION

The objective of this validation study is to inspect how well our developed tour formation model is able to reproduce how the carriers in the data construct their tours from a given set of shipments. It is not sufficient to merely estimate the ET and SS choice models and report a pseudo-R^2 and percentage of correctly predicted choices. These two choice models are part of a larger framework: the tour formation model depicted in the flow diagram in Figure 5.2. Parts of this tour formation model, other than the choice models, influence which tours are formed and impact the validity of the model. When we estimate the choice models, we have a list of observed tours that we can loop through, while when we apply the whole tour formation model, we incrementally add another shipment to the tour and previously modeled decisions influence the next choice situation. Another example of an aspect of the model that is not tested sufficiently in estimation is the choice set formation in the SS model, as discussed above.

In this section, we validate the developed tour formation model. We do so by applying the model to a separate part of the XML data, which we denote as the validation data set. In other words, we construct tours with the shipments in the validation data set. These predicted tours are then compared to the observed tours as they are reported in the validation data set. Note that predicted tours are constructed with the same set of reported shipments as the observed tours. In a freight simulation framework that includes our tour formation model, the accuracy of traffic flow predictions is greatly influenced by our ability to generate a realistic set of shipments, which is not tested in this validation.

To compare the observed and our predicted tours, we measure the level of similarity between the observed and predicted frequency distribution of tour distance and number of stops. Both a quantitative comparison with coincidence ratios and a qualitative/visual comparison with histograms is made. Tour distance and number of stops are two key statistics that characterize different types of tours, as we identified in Section 2.3.1. The number of stops is the key tour characteristic that we try to explain in this research, while tour distances influence relevant factors for policymakers, such as emissions and congestion. Tour distance is used instead of tour duration for validation, as we noted in Chapter 4 that the observed tour duration, which also includes dwelling time at stops, is a rather unreliable variable in the XML data.

To compare two unpaired distributions (i.e. predicted and observed tours), a Chi square test could be used. However, a Chi square test assumes a normal distribution and cell counts below five are not allowed (Levine, 2010). Both restrictions are violated in the observed and predicted number of stops and distance, due to their highly skewed distributions (see also Figure 4.8 and Figure 4.12). For this reason, the coincidence ratio is a good alternative to compare two distributions. In freight modelling the coincidence ratio is often used to compare observed and predicted zonal trip distance distributions (Federal Highway Administration, 2007).

The coincidence ratio is calculated by dividing the observed and predicted distribution into bins of constant step size. For each bin, the cumulative percentages of both distributions are calculated. The *coincidence* is the sum over all bins of the minimum of the two cumulative percentages per bin, while the *total* is the sum over all bins of the maximum of the two. The *coincidence ratio* is then obtained through division of the *coincidence* by the *total.* It ranges from 0% to 100%, where a higher percentage indicates a higher similarity between the two distributions (Federal Highway Administration, 2007; Levine, 2010). In validation of freight trip distance distributions, a value higher than 80% is usually considered good (National Cooperative Highway Research Program, 2008), although others consider a value above 90% good (Federal Highway Administration, 2007).

The validation is performed with three different models, denoted A to C. In these models we only vary the proximity constraint and choice set size, for these are more methodological than the other specifications and, therefore, difficult to specify intuitively in a satisfactory way (Table 7.1). The other constraints mentioned in

Table 6.4 and Table 6.10 are included in all three models (i.e. capacity utilization, concrete, tour duration), and both tour sequence algorithms are used. For each model, results are averaged over three runs because these models include random components. Results for each separate run are reported in Appendix G.

In Section 7.1.1 and Section 7.1.2, the number of stops and tour distance are validated respectively using Models A to C estimated on fifty percent of the day + carriers. In Section 7.1.3, this validation is performed using Model A estimated on fifty percent of the carriers, in order to formulate statements about the applicability of the model to other carriers than those that provide data for estimation.

Table 7.1 Specifications of the three models tested in the validation.

Model	Proximity constraint (ET and SS)	Choice set size SS
A	Yes (100 km)	6
B	Yes (100 km)	11
C	Yes (150 km)	11

7.1.1 Number of stops

In Table 7.2, we see that our tour formation model is able to replicate the distribution of number of stops very well, as the coincidence ratio is above 90% with each of the three models. The differences between the three models are negligible. Table 7.3 shows that the coincidence ratio is also satisfactory when we analyze specific types of tours, i.e. tours visiting a distribution center or not and tours transporting certain goods types, although somewhat less so for goods types NSTR1 and NSTR7.

Table 7.2. Coincidence ratios between the observed and predicted number of stops for the three different models, averaged over three runs per model.

Model	Coincidence ratio Number of stops
A	98.81%
B	98.98%
C	98.57%

Table 7.3. Coincidence ratios between the observed and predicted number of stops for the three different models, averaged over three runs per model. Calculated separately for tours visiting a distribution center and transporting different goods types.

Model	Coincidence ratio Number of stops								
	anyDC=0	anyDC=1	NSTR0	NSTR1	NSTR2-5	NSTR6	NSTR7	NSTR8	NSTR9
A	99.06%	96.56%	92.66%	69.57%	95.58%	96.44%	77.94%	99.53%	92.50%
B	98.95%	97.01%	93.74%	68.84%	95.23%	96.56%	78.40%	99.55%	93.05%
C	98.79%	97.26%	91.50%	70.56%	95.46%	98.00%	80.59%	99.47%	95.35%

Table 7.4 shows the percentage of direct tours (tours with 1-2 stops) and Figure 7.1 shows the percentage of multiple-stop tours with a specific number of stops for the observed data and predicted data with the three models. As tours with more than one shipment usually visit more than two stops, we can say that the ET first shipment is validated in Table 7.4 and the ET later shipments model is validated in Figure 7.1.

Table 7.4 shows that our tour formation model is able to replicate the observed percentage of direct tours very satisfactorily. In Figure 7.1, we see that the predicted multiple-stop tours follow the observed distribution of number of stops well too. However, an overestimation of the percentage of tours with 3-4 stops occurs, while an underestimation of the percentage of tours with 6-7 stops can be observed. This might be caused by the fact that the ET later shipments model is used for all ET decisions of tours with more than one shipment. A separate model for each consecutive shipment could improve this. In addition, more tours with 15+ stops are predicted than observed, which may be due to the probabilistic and iterative nature of this model, allowing the process of adding shipments to linger on too long.

Based on the coincidence ratios and visual comparison we can state that the model performance is very robust with respect to the two model parameters that are varied in models A to C: the choice set size and the proximity constraint.

Table 7.4. Observed and predicted percentage of direct tours, averaged over three runs per model.

Percentage of direct tours			
Observed	Predicted A	Predicted B	Predicted C
92.50%	92.51%	92.56%	93.01%

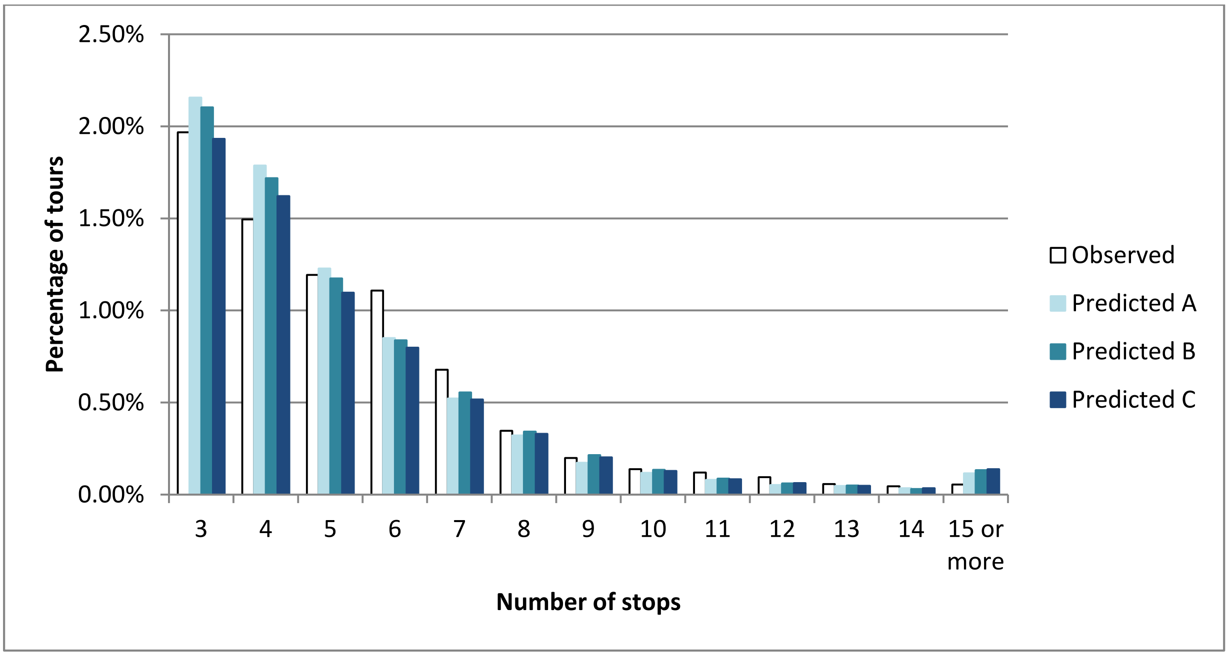

Figure 7.1. Observed and predicted percentage of tours by number of stops, averaged over three runs per model.

The developed tour formation model is able to reproduce differences between tours that visit a distribution center *(anyDC=1)* and tours that do not *(anyDC=0)*. In Table 7.5, we see that the model predicts correctly that tours tend to be direct less often when a distribution center is visited. The difference between the observed and predicted percentage of tours is very small for both *anyDC=0* and *anyDC=1*.

Table 7.5. Observed and predicted percentage of tours by number of stops, averaged over three runs per model. Data divided into tours that do not visit a distribution center (left) and tours that do (right).

Number of stops	anyDC = 0				anyDC = 1			
	Observed	Predicted A	Predicted B	Predicted C	Observed	Predicted A	Predicted B	Predicted C
1-2	96.94%	97.39%	97.45%	97.54%	86.97%	86.44%	86.47%	87.42%
3-5	2.54%	2.17%	2.11%	2.02%	7.29%	8.91%	8.58%	7.89%
6-10	0.42%	0.40%	0.38%	0.39%	0.27%	3.96%	4.20%	3.94%
>10	0.10%	0.04%	0.05%	0.05%	0.98%	0.69%	0.75%	0.75%

Most general tendencies of tours transporting different types of goods are reproduced by our model, as seen in Table 7.6 and Table 7.7. Tours transporting fuels/oils/metals (*2-5*), constructions materials (*6*), and chemical products (8) visit rarely more than two stops, and a larger share of tours with more than five stops is found for foodstuffs (*1*) than for agricultural products (*0*).

Some notable differences between observed and predicted tours can be identified. The predicted percentage of direct tours is too high for foodstuffs (*1*). Other variables such as vehicle type and location type may impact the distribution within this goods category and our model can combine shipments of different goods types in tours, although rather simplistically, which can also lead to differences between observed and predicted tours for different goods types. The fact that the share of predicted tours with more than 10 stops for manure (*7*) is too low can be attributed to the relatively low number of tours transporting manure (about 200 tours per run).

Table 7.6. Observed and predicted percentage of tours by number of stops, averaged over three runs per model. Data divided into tours transporting different goods types (NSTR0 to NSTR6).

Number of stops	NSTR0 *Agricultural products and livestock*		NSTR1 *Other foodstuffs and fodder*		NSTR2-5 *Fuels, oils, and metals*		NSTR6 *Construction materials*	
	Observed	Predicted A	Observed	Predicted A	Observed	Predicted A	Observed	Predicted A
1-2	72.51%	75.24%	64.80%	82.30%	97.89%	96.16%	97.49%	95.81%
3-5	26.01%	22.88%	23.87%	11.78%	1.32%	3.23%	2.05%	2.60%
6-10	1.46%	1.87%	9.48%	4.47%	0.79%	0.61%	0.33%	1.34%
>10	0.02%	0.01%	1.84%	1.45%	0.00%	0.00%	0.13%	0.25%

Table 7.7. Observed and predicted percentage of tours by number of stops, averaged over three runs per model. Data divided into tours transporting different goods types (NSTR0 to NSTR6).

Number of stops	NSTR7 *Manure*		NSTR8 *Chemical products*		NSTR9 *Vehicles, machines, and other goods*	
	Observed	Predicted A	Observed	Predicted A	Observed	Predicted A
1-2	77.88%	80.29%	99.31%	99.45%	84.14%	82.02%
3-5	4.15%	7.50%	0.61%	0.47%	8.20%	12.01%
6-10	4.15%	9.31%	0.08%	0.07%	6.78%	5.30%
>10	13.82%	2.90%	0.01%	0.02%	0.89%	0.66%

7.1.2 Tour distance

The developed tour formation model also reproduces the observed distribution of tour distance in a very satisfactory manner, with coincidence ratios approaching 90% (Table 7.8).

Table 7.8. Coincidence ratios between the observed and predicted tour distances for the three different models, averaged over three runs per model.

Model	Coincidence ratio Tour distance
A	89.30%
B	89.36%
C	89.54%

In Figure 7.2, we see that the general curve of tour distances is reproduced well. The most notable difference is that too many short-distance tours (<100 km) are predicted. This is very likely due to differences in the way observed tour distances are reported in the data and the way we calculate the predicted tour distances. Firstly, the observed tour distances can include kilometers driven for stops to lunch and get gas, while the predicted tour distances do not. Secondly, observed tour distances might be longer due to detours as a consequence of AM and PM peak congestion and truck road restrictions, while our predicted tour distances are based on shortest paths using off-peak skim matrices. Thirdly, many of the observed 'single-stop' tours in the observed data (i.e. one loading location, same unloading location) have a tour distance of several tens to hundreds of kilometers, while our skim matrix (naturally) assumes a very low tour distance for an intrazonal trip. Taking this into account, the similarity between the observed and predicted tour distance distribution in Figure 7.2 is very satisfactory.

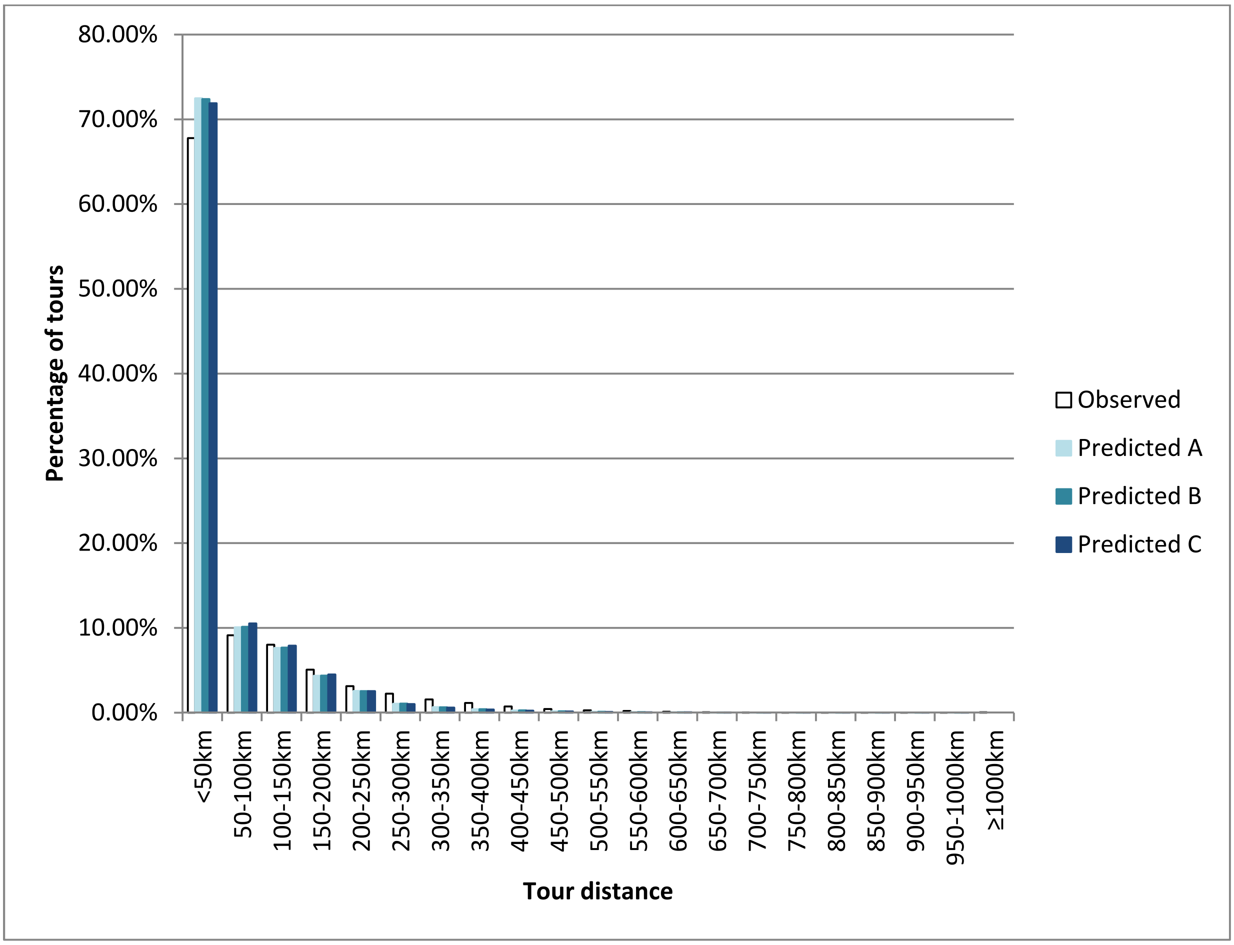

Figure 7.2. Observed and predicted percentage of tours by tour distance, averaged over three runs per model.

As a consequence of the three differences described above, the predicted total Vehicle Kilometers Traveled (VKT) is substantially lower than the observed VKT, also when we exclude single-stop tours and excessively long tours (≥1000 km) from the comparison (Table 7.9 and Table 7.10). Model C predicts the highest total VKT. Its more lenient proximity constraint allows for the addition of more distant shipments, leading to slightly longer tours. Due to the consistent overestimation of the share of short-distance tours of all models, the coincidence ratio of model C is slightly higher (Table 7.8). Model B predicts a lower total VKT than Model A. A larger choice set size increases the probability that a shipment with very little additional time is included in the choice set, which leads to shorter tours on average.

Table 7.9. The total Vehicle Kilometers Traveled of the observed tours in the validation data set.

Observed total VKT	
All tours	Only tours with > 1 stop and < 1000 km
13,709,344 km	11,321,899 km

Table 7.10. The total Vehicle Kilometers Traveled of the predicted tours in the validation data set, averaged over three runs per model.

Model	Predicted total VKT			
	All tours		Only tours with > 1 stop and < 1000 km	
	Absolute	Predicted/observed	Absolute	Predicted/observed
A	9,863,052	71.94%	9,854,677	87.04%
B	9,851,791	71.86%	9,843,364	86.94%
C	9,902,783	72.23%	9,894,038	87.39%

7.1.3 VALIDITY FOR OTHER CARRIERS

The shipments of fifty percent of the carriers, instead of day + carriers, were used to estimate the ET and SS models in Section 6.4. Here we report the results of the application of these estimated models on the shipments of the other fifty percent of the carriers.

In Table 7.11 and 7.12 the coincidence ratios of Model A estimated on a subset of days are compared with the coincidence ratios of Model A estimated on a subset of carriers. We see that in the second case, the model has more difficulty reproducing the observed distribution of the number of stops and tour distance, coincidence ratios are lower. Less diverse information (i.e. set of carriers) is available for estimation, and the estimation and validation data set consist of more dissimilar carriers and shipments (see Appendix I). Tables 7.13 and 7.14 also show larger differences between observed and predicted percentages of tours by number of stops and distance than we found in Section 7.1.1 and Section 7.1.2.

While the predictive performance is not as good when we use a model estimated on a subset of carriers to construct tours with shipments of other carriers, the results are still highly satisfactory. Coincidence ratios exceed 80% (Tables 7.11 and 7.12) and the observed and predicted distribution of number of stops and tour distance are very similar. Consequently, we conclude that our estimated model is applicable to model tour formation for other carriers, although these carriers should be similar to those that are in the XML microdata, i.e. Dutch third-party carriers with advanced planning systems.

Table 7.11. Coincidence ratios between the observed and predicted number of stops and tour distance, averaged over two models runs. Comparison between validation with data divided by day + carriers and by carriers.

Coincidence ratio			
Number of stops		Tour distance	
50% of day + carriers for estimation	50% of carriers for estimation	50% of carrier + days for estimation	50% of carriers for estimation
98.81%	96.92%	89.30%	84.19%

Table 7.12. Coincidence ratios between the observed and predicted number of stops for different location and goods types, averaged over two models runs. Comparison between validation with data divided by day + carriers and by carriers.

Data used for estimation	Coincidence ratio Number of stops								
	anyDC=0	anyDC=1	NSTR0	NSTR1	NSTR2-5	NSTR6	NSTR7	NSTR8	NSTR9
50% of day + carriers	99.06%	96.56%	92.66%	69.57%	95.58%	96.44%	77.94%	99.53%	92.50%
50% of carriers	98.61%	90.36%	95.83%	85.05%	94.39%	94.88%	88.99%	96.25%	90.82%

Table 7.13. Observed and predicted percentage of tours by number of stops, averaged over two model runs. Estimation and validation data divided by carrier.

Number of stops	Percentage of tours *50% of carriers for estimation, 50% of carriers for validation*	
	Observed	Predicted
1 to 2 (direct)	90.76%	89.22%
3	3.28%	3.49%
4	2.25%	2.76%
5	1.43%	1.51%
6	0.81%	0.92%
7	0.51%	0.54%
8	0.28%	0.38%
9	0.25%	0.23%
10	0.17%	0.18%
11	0.13%	0.14%
12	0.05%	0.12%
13	0.03%	0.08%
14	0.02%	0.07%
≥15	0.04%	0.36%

Table 7.14. Observed and predicted percentage of tours by tour distance, averaged over two model runs. Estimation and validation data divided by carrier.

Tour distance	Percentage of tours 50% of carriers for estimation, 50% of carriers for validation	
	Observed	Predicted
<50km	49.50%	58.09%
50-100km	19.75%	16.53%
100-150km	12.59%	11.41%
150-200km	7.75%	7.22%
200-250km	3.79%	2.93%
250-300km	2.12%	1.44%
300-350km	1.50%	0.84%
350-400km	1.15%	0.55%
400-450km	0.70%	0.33%
450-500km	0.38%	0.21%
500-550km	0.25%	0.16%
550-600km	0.13%	0.11%
600-650km	0.10%	0.07%
650-700km	0.05%	0.05%
700-750km	0.04%	0.03%
750-800km	0.02%	0.02%
800-850km	0.02%	0.01%
850-900km	0.01%	0.00%
900-950km	0.03%	0.00%
950-1000km	0.02%	0.00%
≥1000km	0.10%	0.00%

7.2 Sensitivity analysis

In this section, we report a sensitivity analysis, in order to further understand and validate the developed tour formation model.

We define simple scenarios with varying travel times in the network. Travel time is an aspect that is present in many parts of the developed tour formation model. Travel time is an explanatory variable in all three choice models: *tour duration* is found in both ET choice models and *additional generalized cost* is found in the SS choice model. Furthermore, travel time is considered both a constraint and an explanatory variable. It is highly plausible that travel time changes will occur in reality. Understanding the sensitivity to varying travel times can provide useful information for policymakers.

Four scenarios are defined: (1) +50% travel times, (2) +10% travel times, (3), -10% travel times, and (4) -50% travel times. Scenarios are implemented through multiplication of all cells in the travel time skim matrix, while the null scenario uses the original skim matrix. In reality, certain links are more likely to experience changes in travel time than others, caused by complex phenomena such as latent demand. As this sensitivity analysis is only a small part of this research, the defined scenarios are kept simplistic, which allows reactions of the model to be understood more easily. Note that changes in goods flows and vehicle type choices are also likely to occur, but these are outside of the scope of this research, we isolate the effects on tour formation.

Model A (proximity constraint of 100 km; choice set of six shipments) is used in this scenario analysis because it requires the shortest running time of the three models and only marginal differences in predictive performance between the three models were found. The estimates based on fifty percent of day + carriers are used, for these estimates showed better validation results than the estimates based on fifty percent of carriers in Section 7.1.3. Because of random components, the results in this section are averaged over two model runs. In Appendix H, the results of the sensitivity analysis are reported for both runs.

7.2.1 Number of stops

When travel times in the network increase (scenario +50% and +10%), we see a lower percentage of direct tours (Figure 7.3) and thus a higher percentage of multiple-stop tours. Those tours that do visit multiple stop locations, though, tend to have fewer stops in this scenario (Figure 7.4). Fewer direct tours are made because it takes a longer time to deliver a shipment in such a direct tour. Travel time savings through construction of multiple-stop tours are larger; therefore, a strategy of direct tour construction to reduce the complexity of the planning is adopted less often. This effect is captured by the negative *tour duration* parameter in the ET first choice model.

Multiple-stop tours tend to have fewer stops in this scenario, because a tour with the same set of shipments has a longer duration. As the tour gets longer, there is a higher probability that the driver cannot deliver all shipments within the constrained working day of nine to ten hours. As a consequence, carriers start constructing tours with fewer stops. This effect is captured by the positive *tour duration* parameter in the ET later shipments model and the tour duration constraint of nine hours. As travel times increase, shipments that are located nearby are chosen more often in the SS choice model due to the *additional generalized cost* variable and proximity constraint. This provides a small counterforce that limits the described decrease of number of stops. The sensitivity with regard to the number of stops appears to be linear for larger travel time changes. Compared to the 10% scenarios, the effects of the 50% scenarios appear approximately five times stronger in Figure 7.3 and Figure 7.4.

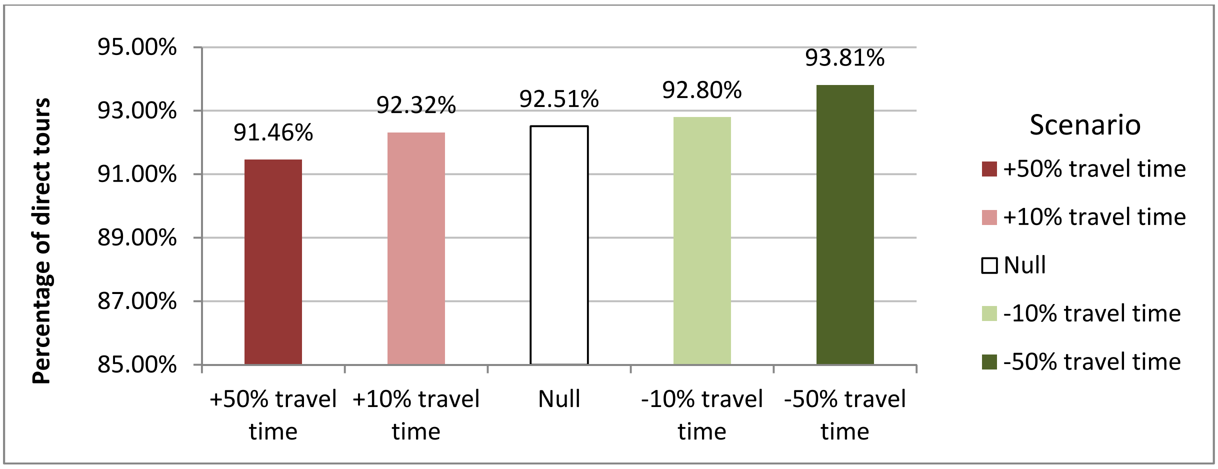

Figure 7.3. Percentage of predicted direct tours in different scenarios reflecting travel time changes, averaged over two model A runs per scenario.

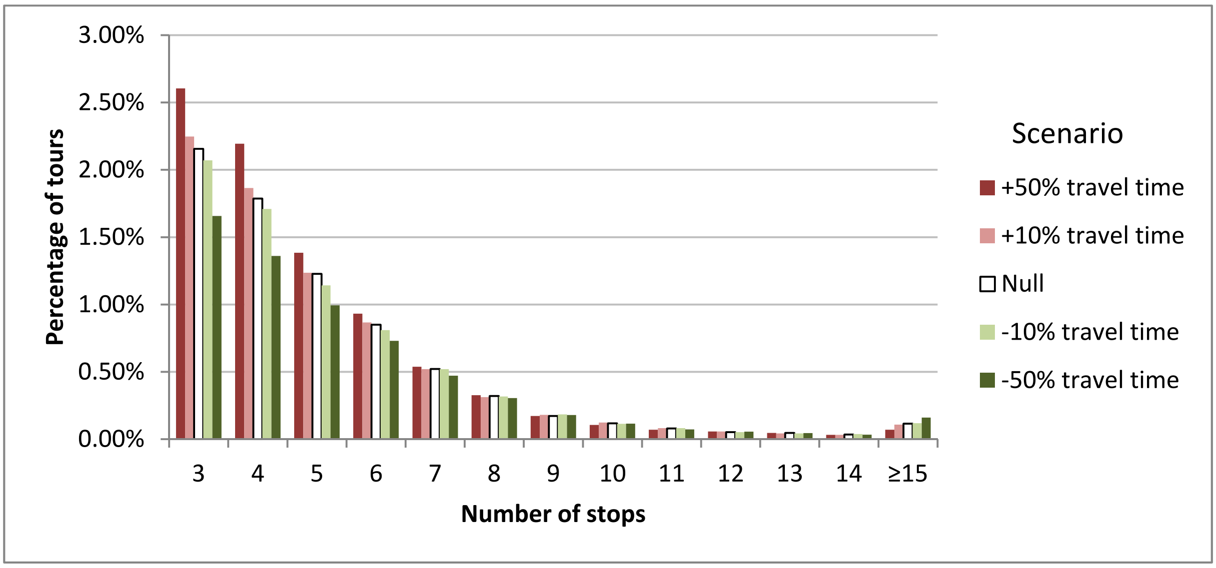

Figure 7.4. Percentage of predicted tours by number of stops, in different scenarios reflecting travel time changes, averaged over two model A runs per scenario.

7.2.2 Tour distance

The distribution of tour distances shows a less pronounced sensitivity than the distribution of number of stops. For this reason, we do not only show the distribution in a bar chart (Figure 7.5) but also in table format (Table 7.15).

Two main effects are at play: (1) tours with more or fewer stops are made, (2) shipments with more or less additional distance are chosen. As a tour includes more stops, more trips are made, leading to higher tour distances. As shipments with a higher additional distance are chosen, trips within a tour are longer, also leading to higher tour distances. These two effects are intertwined, as picking shipments with a lower additional distance also allows the carrier to make tours with more stops while respecting working hour constraints. As a consequence, positive and negative impacts can be cancelled out.

In Figure 7.3 and Figure 7.4, we observed that in the +50% scenario more tours with an intermediate number of stops (3-5) are made, while direct tours and tours with many stops are made less often. As tours with more stops tend to have a higher distance, we also observe that in this scenario the number of tours in the intermediate distance range (100-400 km) increases, while fewer very short (<50 km), short (50-100 km), long (400-700 km), and very long (≥700 km) tours are observed. Because shipments with a lower additional distance are chosen on average too, the effect is rather limited for such an extreme scenario (Figure 7.5).

In the -50% scenario, we see more tours that are less than 150 km long, fewer tours in the distance range 150-700km, and more very long tours (≥700 km). Very long tours are observed more often because more shipments can be grouped in a tour shorter than nine hours. A complex interaction of the following effects might explain the fluctuation of number of tours for different distance ranges: (1) more direct tours are made, (2) more tours with more than five stops are made, (3) shipments with a higher additional distance are chosen on average.

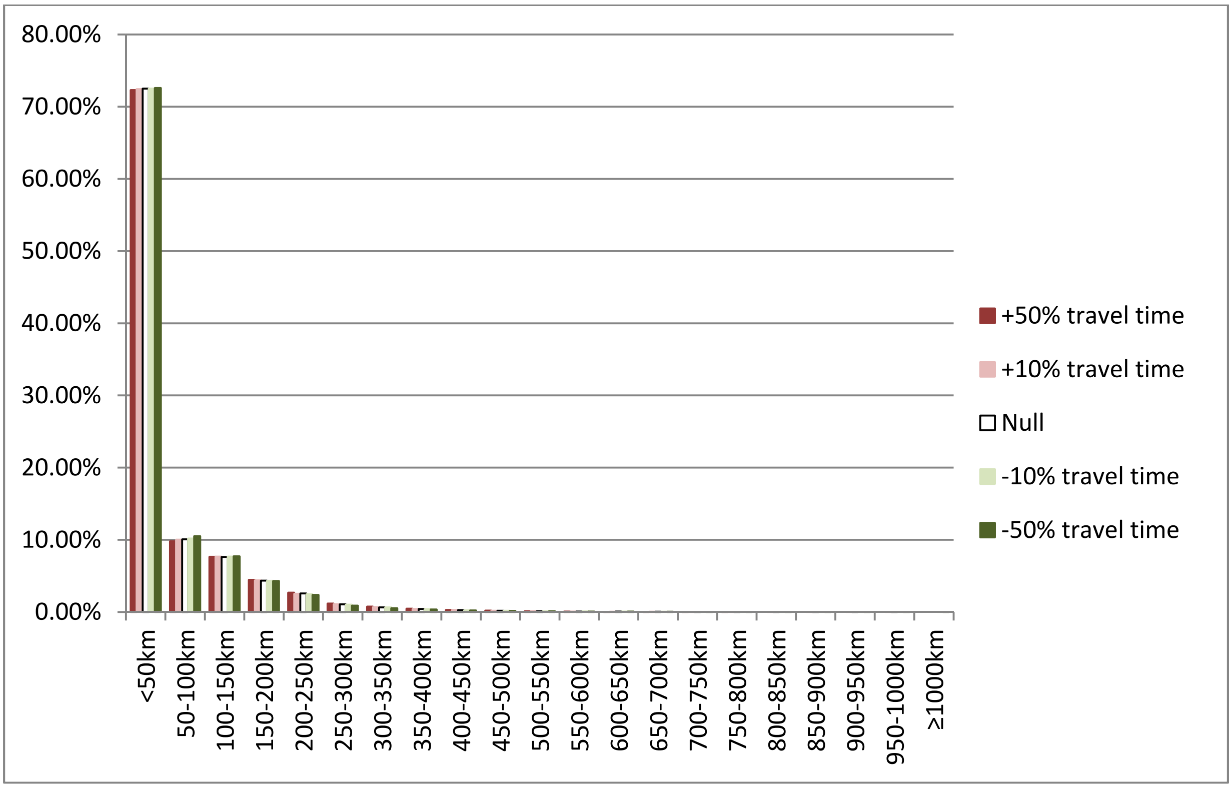

Figure 7.5. The distribution of tour distances in different travel time scenarios, averaged over two Model A runs per scenario.

Table 7.15. The distribution of tour distances in different travel time scenarios, averaged over two runs per scenario.

Tour distance	+50%	+10%	Null	-10%	-50%
<50km	72.30%	72.43%	72.50%	72.51%	72.59%
50-100km	9.81%	10.02%	10.07%	10.17%	10.51%
100-150km	7.66%	7.67%	7.63%	7.68%	7.71%
150-200km	4.46%	4.38%	4.35%	4.32%	4.29%
200-250km	2.67%	2.52%	2.57%	2.47%	2.36%
250-300km	1.17%	1.08%	1.06%	1.03%	0.90%
300-350km	0.75%	0.68%	0.65%	0.64%	0.54%
350-400km	0.47%	0.43%	0.41%	0.40%	0.35%
400-450km	0.28%	0.29%	0.27%	0.25%	0.22%
450-500km	0.19%	0.17%	0.17%	0.18%	0.15%
500-550km	0.12%	0.12%	0.12%	0.11%	0.10%
550-600km	0.07%	0.07%	0.08%	0.08%	0.07%
600-650km	0.03%	0.05%	0.06%	0.06%	0.06%
650-700km	0.01%	0.03%	0.04%	0.03%	0.04%
700-750km	0.00%	0.02%	0.02%	0.03%	0.03%
750-800km	0.00%	0.01%	0.01%	0.01%	0.02%
800-850km	0.00%	0.00%	0.01%	0.01%	0.02%
850-900km	0.00%	0.00%	0.00%	0.01%	0.01%
900-950km	0.00%	0.00%	0.00%	0.00%	0.01%
950-1000km	0.00%	0.00%	0.00%	0.00%	0.01%
≥1000km	0.00%	0.00%	0.00%	0.00%	0.01%

A crucial footnote to this analysis is that empty trips are not included. As a consequence, our model cannot capture sufficiently that construction of multiple-stop tours rather than direct tours might actually reduce the VKT required to transport the same set of shipments. Consolidation of shipments of different customers reduces the required number of tours (Figliozzi, 2007) and number of empty trips (Roorda et al., 2010). A direct tour can require an empty trip to and from the home base to deliver just one shipment, while a multiple-stop tour requires the same number of empty (home base) trips to deliver (possibly far) more shipments. Due to this omission of empty trips, we see that while in the scenarios with lower travel times the share of direct tours increases quite substantially (Figure 7.3), the total VKT actually decreases compared to the null scenario (Table 7.16). As no clear relationship can be defined between the number of empty trips and the total VKT (Figliozzi, 2007), it is difficult to formulate sound statements about the total VKT in this research and inclusion of empty trips is an absolute priority for future research.

Table 7.16. The total Vehicle Kilometers Traveled of the predicted tours in the different travel time scenarios, averaged over three runs per scenario.

Scenario	Predicted total VKT			
	All tours		Only tours with > 1 stop and < 1000 km	
	Absolute	Compared with null	Absolute	Compared with null
+50%	9,908,791	+0.46%	9,900,561	+0.47%
+10%	9,890,156	+0.27%	9,881,751	+0.27%
-10%	9,862,867	-0.00%	9,851,255	-0.03%
-50%	9,836,560	-0.27%	9,795,526	-0.60%

In general, the developed tour formation model shows very plausible effects when travel times in the network increase and decrease. The effects can be explained very well, especially with respect to the number of stops. This further underlines the validity of the model.

8 CONCLUSIONS AND RECOMMENDATIONS

The main research question and sub questions read as follows:

Can we develop a behavioral shipment-based tour formation model that reproduces observed tour patterns?

1. Which objectives, constraints, and other factors influence freight tour formation?
2. To what extent is the XML microdata useful for calibration of a freight tour formation model?
3. How can we structure the allocation of shipments to tours in such a model?
4. Which aspects of freight tour formation can we include in the model?
5. How well does the model reproduce observed tour patterns?

In this final chapter, we answer these research questions, after which we give recommendations for further improvement of the model, future research, policymakers, and data collection.

8.1 CONCLUSIONS

SUB QUESTION 1

Objectives in freight tour formation reflect those of the carrier, for this is the actor that allocates shipments to tours. Constraints can be imposed by senders and receivers, by the carrier, by regulations, and by simple common sense.

As carriers are usually private companies, the most obvious objective is profit maximization. Different strategies can be employed for profit maximization and carriers can have more subjective preferences; consequently, different types of objectives exist. The following types of objectives have been identified: (1) minimization of travel time/distance/cost, (2) minimization of the makespan, the total daily operating hours, (3) minimization of emissions, (4) maximization of punctuality, (5) maximization of the visual attractiveness and tractability of the constructed tour set, (6) minimization of the complexity of tour planning, (7) maximization of the level of balance of the durations of constructed tours, and (8) maximization of the utilized vehicle capacity.

Constraints can impact tour formation strongly, they limit which tours a carrier can construct. The following types of tour formation constraints have been identified, (1) the availability of shipments with which to construct tours (2) goods type compatibility, (3) time windows, (4) vehicle capacity, (5) precedence of loading and unloading locations, and (6) driver working hours.

Other variables that explain differences in observed tours can be divided into three categories: location, vehicle, and goods type. Through analysis of the number of visited stops, the following differences between tours visiting certain location types have been identified:

- Tours that visit a *port transshipment node* are more likely to be direct, regardless of whether goods are loaded or unloaded here. These tours are more likely to transport shipments originating from a producer. These shipments tend to have larger sizes, which leaves little room in a vehicle for other shipments (Friedrich et al., 2014).
- Tours that visit a *distribution center* to *load* goods have more stops. These shipments are likely to be transported to consumers instead of producers; therefore, they tend to be smaller in terms of volume and weight (Friedrich et al., 2014). Distribution centers also have large set of available shipments with similar loading points, and organize their vehicle loading and unloading activities such that more customers can be visited in the same tour (Khan & Machemehl, 2017). In addition, larger vehicles are used on average at distribution centers (van Duin et al., 2012).
- Tours that visit a *distribution center* to *unload* goods have fewer stops. These shipments are more likely to originate from producers and have greater shipment sizes (Friedrich et al., 2014).

- Tours that visit an *urban zone* or *retail zone* have more stops. Potential customers are more concentrated, which stimulates the grouping of shipments of these customers in an efficient tour, especially since entering and leaving a city can be very time consuming.

Differences in tours by vehicle type can be related to capacity and ease of (un)loading. *Tractor + trailers* and *trucks* are used more often for tours with multiple stops than *truck + trailers* and *vans*. Loading/unloading of shipments of different customers is less practical with truck + trailers, since these always consist of two or more compartments where the last needs to be uncoupled to access goods in the first. Vans have a much lower vehicle capacity. When shipment sizes stay constant, much fewer shipments can be transported with a van in one tour.

Differences between goods types can relate to: shipment size, rigidity of goods type combination restrictions, ease of (un)loading, and dispersion of demand. A goods type clearly distinguishable from the rest is *cement/concrete*. Tours with more than one shipment are virtually never observed for cement/concrete due to large shipment sizes and a high time-sensitivity (Khan & Machemehl, 2017). Furthermore, tours transporting *oils*, *fuels*, *construction materials*, and *chemical products* make few stops, while tours transporting *agricultural products*, *foodstuffs*, *manure* and *other products* make more stops.

Sub question 2

The XML microdata has been a very useful source of data to calibrate a tour formation model in this research. Firstly, it provides a lot of insight into variables explaining tour differences. Secondly, since we know which shipments are part of each tour and loading and unloading locations of these shipments are reported, calibration of choice steps in an algorithm that allocates shipments to tours is possible. However, several peculiarities and definitions influence what information can be extracted from this data, for example:

- A reported tour is not started until goods are loaded into the vehicle, and the tour is ended when the vehicle turns empty or returns to the home base. Therefore, empty trips are not included in the data and a strikingly large portion of tours with only one shipment is found.
- The tour starting and end time are known, but the arrival and departure times at intermediate tour stops are not present. As a consequence, realized tour sequences and dwelling times are unknown.
- Shipment size is only available in a usable format with respect to weight, therefore, the volume utilization rate of the vehicle cannot be deduced satisfactorily.
- As implementation of the XML-interface requires carriers to have an advanced transport management system, a self-selection of large third-party carriers has taken place. The data is not representative for all Dutch freight carriers.

Enrichment of the XML microdata with other data sources is crucial to overcome some of these data peculiarities and to obtain meaningful tour statistics for estimation of a tour formation model. For example, a skim matrix of travel times and distances between zones is needed to construct tour sequences and to calculate additional costs of adding shipments.

Sub question 3

To be able to calibrate a tour formation model on the XML microdata, we structure the model as an algorithm that iteratively allocates an additional shipment to a tour until it is ended. This process is continued until all shipments that a carrier has to transport on a day are allocated to a tour, and then repeated for all days and carriers.

This algorithm has two steps where a choice is modeled, for which we estimate a utility function on the empirical data. We call these two choices the End Tour (ET) and the Select Shipment (SS). The ET choice is modeled with a Binary Logistic Regression. The dependent variable is binary; a value of 0 means that another shipment is added to the tour, a value of 1 means that the current tour is ended and a new tour is started. The

SS choice model is a Multinomial Logit model. If the decision to add another shipment is made in the ET choice model, the SS choice model is used to choose which shipment is added.

SUB QUESTION 4

The developed model considers many aspects of tour formation, most importantly:

- Of the identified tour formation objectives, the following are represented in our model: minimization of travel time/distance/cost, maximization of punctuality, minimization of complexity of tour planning, maximization of utilized vehicle capacity.
- The following identified constraints are included: availability of shipments, vehicle capacity, precedence of loading and unloading locations, driver working hours. Goods type restrictions are considered more implicitly through a preference for tours with one goods type.
- The model acknowledges that different types of tours are preferable and feasible for different location, vehicle, and goods types.
- The model is shipment-based and, therefore, considers that shipments form the fundamental level at which many decisions in freight transportation are made (de Bok & Tavasszy, 2018), and that the demand for freight transportation is derived from goods flows between firms or zones (Boerkamps & van Binsbergen, 1999; Wisetjindawat et al., 2006). Additionally, application of the tour formation model in a shipment-based simulation framework allow us to consider the impacts of specific policies and scenarios on tour formation, such as distribution centers, changes in delivery frequencies and shipment sizes, and increased cooperation of senders and carriers (Boerkamps & van Binsbergen, 1999).

There are, however, several components of tour formation that our model does not include or acknowledge:

- Empty trips to and from the home base are not included in the data and, therefore, not modeled. For this reason, the predicted total Vehicle Kilometers Traveled (VKT) in the network is underestimated, and the fact that construction of tours with more stops can reduce the number of empty trips (and thus total VKT) is not considered.
- The incremental fashion of constructing tours imposes a simplified structure on the complex tour formation process. Consequently, our model does not consider all interrelated decisions. We only relate the selection of the next shipment to the shipments that are already allocated to the tour and not to those that might be allocated after it.
- The following identified objectives are not considered in our model: makespan, emissions, visual attractiveness, and level of balance.
- Time windows are the only identified type of constraint that is not acknowledged by our model in any way. Furthermore, while vehicle capacity constraints are included in terms of transported weight, neither volume capacity nor vehicle fleet size is considered.
- While we do consider the impacts of congestion, these impacts are underestimated as we use off-peak skim matrices to determine travel times and distances. Detours to get lunch or gas or due to truck access restrictions are not included in these travel times and distances either.
- Tour durations do not include dwelling time to load and unload goods.

SUB QUESTION 5

Application of the estimated tour formation model to form tours with the shipments of carriers that did not provide data for estimation has shown that the model can reproduce observed tour statistics very satisfactorily:

- The predicted distributions of tour distance and number of stops are highly similar to the observed distributions for a given set of shipments.

- The model can reproduce differences in the distribution of number of stops between tours visiting different locations types and transporting various goods types.

In addition, plausible effects are predicted when travel times in the network increase . In this scenario, there is a stronger focus on the construction of efficient tours to minimize travel costs; therefore, fewer tours are predicted that transport only one shipment. Tours that visit a large number of customers are found less often too, as working hour constraints are violated more quickly.

It should be noted that the same set of reported shipments is transported in the observed and predicted tours. In a freight simulation framework these shipments need to be generated, in that case it might be harder to reproduce these statistics this accurately.

Because we do not construct empty trips and vehicle trips are not assigned to a traffic network yet, it is difficult to formulate informed statements about the extent to which this tour formation model can improve our ability to predict freight truck flows. Because of the surprisingly large share of observed direct tours, the improvement might be more limited than expected upfront. The extraordinary share of cement/concrete shipments in the data, results from previous studies, and the fact that tours in the data are ended when the vehicle turns empty indicate that this share might be lower in reality, though.

Main research question

Development of a model that forms tours out of shipments in a way that is similar to reality is indeed possible. Adopting an iterative shipment allocation structure allows for the calibration of choice models on empirical shipment data. Most of the identified aspects that explain tour formation can be included in such a model, which reproduces observed tour statistics very satisfactorily.

In addition, calibration of this model can be performed in a computationally efficient and statistical way. Construction of tours from a set of more than 200,000 shipments requires about half an hour, and straightforward measures of model fit can be obtained.

Our tour formation model including estimated parameters can be used in a freight simulation framework, although several conditions must be fulfilled. Firstly, the geographical scope of the framework should be freight transportation by road within the Netherlands. Factors such as work hour regulations and the spatial distribution of activities in the Netherlands influence the estimation results and constraints. Secondly, the model should be used only to construct tours performed by third-party carriers with advanced planning systems. We have shown that a model estimated on fifty percent of these carriers in our data can be used to construct tours of the other carriers in the data in a satisfactory manner. The exact bias of the XML microdata should be specified more clearly, though, to identify the population of carriers for which our model is applicable. Thirdly, off-peak skim matrices should be used to obtain travel impedances. Finally, shipments between firms need to be synthesized and assigned to carriers, and a vehicle type choice model must be estimated before tours can be constructed.

8.2 Recommendations

Model improvements

While the developed tour formation model is able to consider many influences and reproduce observed tour statistics excellently, there are technical ways to improve it:

- Both a Binary and Multinomial Logistic Regression assume independently and identically distributed error terms. This is not the case in the XML microdata, since many shipments are transported by the same carrier. Estimating the choice models with panel effect can account for this.

- A thorough analysis can improve our knowledge of the home base locations of carriers in the data. Robroeks (2016), for example, analyzed the number of shipments that have a particular loading or unloading location for each carrier to identify the home base location of more tours. Doing so would allow the incorporation of empty trips in the model. Alternatively, a 2-opt-algorithm may be applicable for construction of a logical set of empty trips between direct tours.
- For the last remaining shipment of a carrier on a day, we can develop a decision rule that checks whether it would fit well in a tour already constructed. This way a backwards flow of information is included, and we consider that the selection of shipments for a tour is a process of many interrelated decisions.
- A Nested Logit formulation can be used to model the ET and SS choice as a simultaneous one. The first nest would be simply ending the tour, the second nest would be continuing the tour, where the second nest has several shipment alternatives. This way we can incorporate in a more elegant fashion how the attractiveness of the set of remaining shipments influences the choice to end the tour.
- In the construction of the tour sequence, the first visited location is the loading location of the first allocated shipment. After further consideration, it may be more efficient and logical to start the tour sequence at a more peripheral location. Other algorithms may be used to construct the tour sequence too. One example is the sweep algorithm, which can be used to construct a sequence of the tour locations by sorting the locations based on their polar angle to the tour starting location (Suthikarnnarunai, 2008).
- A more elaborate capacity utilization can be used that recognizes that not all shipments of a tour need to be present in the vehicle simultaneously.
- Usage of different skim matrices for tours departing in the AM peak or PM peak may improve the prediction of tour distances and durations, as the effects of congestion are included. A challenging aspect is that tours can stretch out over multiple parts of the day; therefore, different parts of a tour may require different skim matrices.

FUTURE RESEARCH

More generally, there are many ways in which to extend this research to understand and predict freight transportation better.

- Logistic decisions are highly interrelated (Khan & Machemehl, 2017; de Bok et al., 2018). Tour formation cannot be seen completely isolated from the choice for vehicle type and shipment size. If a carrier wants to add a shipment that would cause an exceedance of the vehicle capacity, it might be possible to choose a larger vehicle. Another decision closely related to tour formation is the departure time choice. Incorporation of this tour formation model in a larger simulation framework with other interrelated choices provides many interesting new challenges to freight modelers.
- A carefully synthesized set of shipments is of vital importance to obtain good results with the tour formation model developed in this research. Analysis of the spatial distribution of shipments and relationships between shipment attributes can be of great use to synthesize a realistic set of shipments in a freight simulation framework. In addition, the synthesized shipments must be assigned to carriers in an appropriate way.
- Along with such a shipment synthesizer and a traffic assignment module, observed and predicted traffic link flows can be compared, which would provide far more insight into the extent to which a tour formation model like ours can improve the predictive performance of a freight simulation framework.
- As we identified in the interview with a transportation planner that tour formation decisions are often constrained by the vehicle fleet size, it is useful to include this in a freight simulation framework. While this is a more strategic decision than the tactical tour formation decisions, assigning a realistic vehicle fleet size to each carrier would allow us to consider this important constraint for tour formation.

- Time windows are a crucial feature of freight transportation that impact which tours can be constructed (Figliozzi, 2007). Inclusion of time windows would greatly improve the behavioral foundation of a tour formation model like this.
- The current model results are not representative for all freight carriers in the Netherlands. Its use for application in another geographical context is also questionable. In such a small and dense country as the Netherlands, parameters such as *tour duration* and *additional generalized cost* are expected to have different values than elsewhere. Therefore, we recommend to estimate similar tour formation models paying special attention to representative data and perhaps with data of other parts of the world.
- The general hypothesis is that third-party carriers tend to construct tours with more stops, for they have a larger set of shipments to combine efficiently (McCabe et al., 2006; Roorda et al., 2010; Nuzzolo et al., 2012; Beziat et al., n.d.). However, an interview with the transportation planner of an own-account carrier showed that they make more stops on average than any of the third-party carriers in our data. We may have to revise this hypothesis and look for new explanations, such as vehicle fleet size differences. A better understanding of the differences between third-party and own-account carriers would greatly assist in a more accurate prediction of the tours made by all carriers in the Netherlands. Using data about own-account carriers, our tour formation model may be estimated for these carriers too.

POLICYMAKERS

While this research has a strong methodological focus, it has also provided useful insights for policymakers.

- In certain segments of freight transportation almost only direct tours are observed. Tours transporting concrete, fuels, oils, metals, construction materials, and shipments (un)loaded at ports are the clearest examples. In these segments, sensitivity with respect to tour formation is strongly limited by large shipment sizes, dispersed demand, and high time-sensitivity. For policymakers that wish to reduce the negative external impacts in these segments, developing policies that influence choices other than tour formation might be more effective, such as the procurement of cleaner vehicles.
- Travel times on the Dutch road network are increasing due to congestion. In the first half year of 2018, the gravity[3] of traffic jams increased by 20% compared to the same period in 2017 (ANWB, 2018). With higher travel times in the network, our results indicate that multiple customers are visited in one tour more often, as there is a stronger focus on travel time savings. Fewer tours (Figliozzi, 2007) and possibly fewer empty trips are made in that case (Roorda et al., 2010). Fewer empty trips might also lead to a lower Vehicle Kilometers Traveled (VKT) and may provide a counterforce to the increased intensity of traffic on the Dutch roads. For solid conclusions, further analyses with an empty trip model are required though.
- Previous research has shown that distribution centers can reduce the negative external effects of freight transportation (van Duin et al., 2012). The findings in this research indicate indeed that distribution centers might facilitate more efficient tour formation. Tours with more stops are observed and predicted originating from distribution centers. When tours with more stops are constructed, less empty trips are required (Roorda et al., 2010).

DATA COLLECTION

Finally, several recommendations can be made with regard to data collection in order to improve our understanding of and ability to predict freight transportation. These recommendations relate specifically to the XML microdata of the CBS, but are valuable to any new freight tour data collection effort.

[3] Jam length multiplied with duration

- Empty trips to and from the home base are not included in the XML microdata, and are difficult to deduce as home base locations are filled out poorly. Empty trips have been found to make up 20% of urban trips and 30-40% of inter-city trips (Sánchez-Díaz et al., 2015). They are a crucial feature of freight transportation that needs to be understood to predict truck flows accurately. We recommend either to (1) include a variable that lists the origin and destination of a possible empty start and return trip, or to (2) make it easier for respondents to fill in the home base location automatically.
- The interview and our interpretation of model parameters lead us to conclude that vehicle capacity in terms of volume is a very important constraint in freight tour formation. Currently, volume can be filled in with different units, leading to a variable that is difficult to decipher. It would be advantageous to have a consistent measure of volume that can be filled out more easily by respondents.
- If the arrival and departure time are filled out for each visited location, many other behavioral aspects of freight tour formation can be unraveled, such as tour sequencing, dwelling times, and goal arrival times. By simply pressing a button on a display screen at arrival and departure, the truck driver can provide these times in high detail. Alternatively, GPS or engine data can be used to determine when the truck is standing still for a longer time.
- More information about the carriers can assist in exposing and explaining heterogeneous behavior. For example, the size of the vehicle fleet provides information about which constraints are experienced by different carriers.
- If the XML-interface can be adapted in such a way that it is easier to install for smaller and own-account carriers, we would obtain a larger amount of more diverse data. This allows us to understand the differences in behavior between third-party and own-account carriers, and to sample a data set representative for all freight carriers in the Netherlands.

BIBLIOGRAPHY

Abate, M. A., & Kveiborg, O. (2010). Capacity utilization of vehicles for road freight transportation. *12th World Conference on Transport Research.* Lisbon, Portugal.

Adler, T., & Ben-Akiva, M. (1979). A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B 13(3)*, 243-257.

Ahuja, R. K., & Orlin, J. B. (2001). Inverse optimization. *Operations Research 49(5)*, 771-783.

Alho, A., Bhavathrathan, B., Stinson, M., Gopalakrishnan, R., Le, D.-T., & Ben-Akiva, M. (2017). A multi-scale agent-based modelling framework for urban freight distribution. *Transportation Research Procedia 27*, 188-196.

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.

ANWB. (2018). *20 procent meer files op de Nederlandse wegen*. Retrieved from https://www.anwb.nl/verkeer/nieuws/nederland/2018/juli/20-procent-meer-files-op-de-nederlandse-wegen

Ben-Akiva, M., & Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA, USA: The MIT Press.

Beziat, A., Launay, P., & Toilier, F. (n.d.). *Analysis of different types of freight tours according to their logistics organization in the Paris Region.* Retrieved from https://www.metrans.org/sites/default/files/Beziat-freight%20tours%20paper_0.pdf

Bodin, L., Maniezzo, V., & Mingozzi, A. (2003). Street routing and scheduling problems. In R. Hall, *Handbook of Transportation Science.* Norwell, MA, USA: Kluwer Academic Publishers.

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.

Bovy, P. (2009). On modelling route choice sets in transportation networks: A synthesis. *Transport Reviews 29(1)*, 43-68.

CBS. (2008). *XML berichtspecificatie - Goederenvervoer over de weg - Versie 2.1.* Retrieved from https://www.cbs.nl/nl-nl/deelnemers-enquetes/deelnemers-enquetes/bedrijven/meer-over-cbs-enquetes/xml-vanuit-transportmanagementsysteem

CBS. (2015a). *Begrippenlijst Wegtransport en Logistiek.* Den Haag, the Netherlands: Centraal Bureau voor de Statistiek.

CBS. (2015b). *Toelichting Kerncijfers Wijken en Buurten 2015.* Retrieved from https://www.cbs.nl/-/media/cbs%20op%20maat/maatwerk/documents/2015/48/toelichting-variabelen-kwb-2015%20versie-2015-11-30.pdf

CBS. (2017a). *Nederlandse vrachtwagens vervoeren meer in eigen land*. Retrieved from https://www.cbs.nl/nl-nl/nieuws/2017/12/nederlandse-vrachtwagens-vervoeren-meer-in-eigen-land

CBS. (2017b). *Basisbestanden Goederenwegvervoer 2015.* Heerlen, the Netherlands: Centraal Bureau voor de Statistiek.

Chow, J. Y., & Recker, W. W. (2012). Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem. *Transportation Research Part B 46*, 463–479.

de Bok, M. (2007). *Infrastructure and firm dynamics: a microsimulation approach (doctoral dissertation).* Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3Af62e84bd-dd9e-4db2-b992-94146299100e

de Bok, M., & Tavasszy, L. (2018). An empirical agent-based simulation system for urban goods transport (MASS-GT). *Procedia Computer Science 130*, 126–133.

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.

de Jong, G., Kouwenhoven, M., Ruijs, K., van Houwe, P., & Borremans, D. (2016). A time-period choice model for road freight transport in Flanders based on stated preference data. *Transportation Research Part E 86*, 20-31.

de Oliveira, G. F., & de Oliveira, L. K. (2017). Stakeholder's perception about urban goods distribution solution: exploratory study in Belo Horizonte (Brazil). *Transportation Research Procedia 25*, 942-953.

Donnelly, R., Wigan, M., & Thompson, R. (2010). A hybrid microsimulation model of urban freight travel demand. *SHRP2 Innovations in Freight Demand Modeling and Data Symposium.* Washington, DC, USA.

Doustmohammadi, E., Sisiopiku, V. P., & Sullivan, A. (2016b). Modeling freight truck trips in Birmingham using tour-based approach. *Journal of Transportation Technologies 6*, 436-448.

Doustmohammadi, E., Sisiopiku, V. P., Anderson, M. D., Doustmohammadi, M., & Sullivan, A. (2016a). Comparison of freight demand forecasting models. *International Journal of Traffic and Transportation Engineering 5(1)*, 19-26.

Federal Highway Administration. (2007). *Quick Response Freight Manual II.* Washington, DC, USA.

Ferguson, M., Maoh, H., Ryan, J., Kanaroglou, P., & Rashidi, T. H. (2012). Transferability and enhancement of a microsimulation model for estimating urban commercial vehicle movements. *Journal of Transport Geography 24*, 358–369.

Figliozzi, M. A. (2007). Analysis of the efficiency of urban commercial vehicle tours: Data collection, methodology, and policy implications. *Transportation Research Part B 41*, 1014–1032.

Figliozzi, M. A., Kingdon, L., & Wilkitzki, A. (2007). Analysis of freight tours in a congested urban area using disaggregated data: characteristics and data collection challenges. *2nd Annual National Urban Freight Conference.* Long Beach, CA, USA.

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

Gilbert, N. (2008). *Agent-based models.* Thousand Oaks, CA, USA: Sage.

Groot, N., & Miete, O. (2016). *Ontwikkelingen strategisch goederenvervoermodel [Powerpoint slides].* Retrieved from https://slideplayer.nl/slide/10263195/

Hillier, F. S., & Lieberman, G. J. (2001). *Introduction to operations research (7th edition).* New York, NY, USA: McGraw-Hill.

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.

Hougardy, S., & Wilde, M. (2015). On the nearest neighbor rule for the metric traveling salesman problem. *Discrete Applied Mathematics 195*, 101-103.

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

Inspectie Leefomgeving en Transport. (2018, July 31). *Rijtijden*. Retrieved from https://www.ilent.nl/onderwerpen/rij-en-rusttijden/rijtijden

Irannezhad, E., & Hickman, M. (2016). Behavioural urban freight modelling: exploring effects of policies on an urban freight distribution system. *Australasian Transport Research Forum.* Melbourna, Australia.

ITF-OECD. (2015). *Global trade: international freight transport to quadruple by 2050.* Retrieved from https://www.itf-oecd.org/sites/default/files/docs/2015-01-27-outlook2015.pdf

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

Kim, H., & Park, D. (2017). Empirical comparison of tour- and trip-based truck travel demand models. *KSCE Journal of Civil Engineering 21(7)*, 2868-2878.

Kim, S., Park, D., Kim, S., & Park, H. (2014). Modeling courier vehicles' travel behavior: case of Seoul, South Korea. *Transportation Research Record (2410)*, 67-75.

Kizilateş, G., & Nuriyeva, F. (2013). On the nearest neighbor algorithms for the traveling salesman problem. In D. Nagamalai, A. Kumar, & A. Annamalai, *Advances in computational science, engineering and information technology. Advances in intelligent systems and computing 225.* Heidelberg, Germany: Springer.

Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Vallabhaneni, L., & Nippani, S. (2014). Development of a tour-based truck travel demand model using truck GPS data. *Transportation Research Board 93rd Annual Meeting.* Washington DC, USA.

Levine, N. (2010). *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3).* Houston, TX, USA; Washington, DC, USA: Ned Levine & Associates; National Institute of Justice.

Martínez, F., Aguila, F., & Hurtubia, R. (2009). The constrained multinomial logit: A semi-compensatory choice model. *Transportation Research Part B 43*, 365–377.

McCabe, S., Kwan, H., & Roorda, M. J. (2006). Freight transportation: who is the decision maker? *53rd Annual North American Meetings of the Regional Science Association International.* Toronto, Canada.

Mohammadian, K., Kawamura, K., Sturm, K., & Pourabdollahi, Z. (2013). *GPS based pilot survey of freight movements in the Midwest region.* Retrieved from https://rosap.ntl.bts.gov/view/dot/26173

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

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.

Ortúzar, J., & Willumsen, L. (2011). *Modelling transport.* West Sussex, UK: Wiley.

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.

Pluvinet, P., Gonzalez-Feliu, J., & Ambrosini, C. (2012). GPS data analysis for understanding urban goods movement. *Procedia - Social and Behavioral Sciences 39*, 450-462.

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

Poot, A., Kant, G., & Wagelmans, A. (2002). A savings based method for real-life vehicle routing problems. *Journal of the Operational Research Society 53*, 57-68.

Prato, C. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling 2(1)*, 65-100.

Quak, H. (2008). *Sustainability of urban freight transport: retail distribution and local regulations in cities (doctoral dissertation).* Retrieved from https://www.researchgate.net/publication/254805169

Recker, W., Duan, J., & Wang, H. (2008). Development of an estimation procedure for an activity-based travel demand model. *Computer-Aided Civil and Infrastructure Engineering 23*, 483–501.

Robroeks, J. (2016). *Large-scale operational company matching for horizontal collaboration in road transport: a commodity driven approach (MSc thesis).* Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3A81f42024-e5b6-413e-bcba-3b99133aedba

Roorda, M. J., Cavalcante, R., McCabe, S., & Kwan, H. (2010). A conceptual framework for agent-based modelling of logistics services. *Transportation Research Part E 46*, 18-31.

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.

Solomon, M. M. (1987). Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research 35(2)*, 254-265.

Stathapoulos, A., Valeri, E., & Marcucci, E. (2012). Stakeholder reactions to urban freight policy innovation. *Journal of Transport Geography 22*, 34-45.

Suthikarnnarunai, N. (2008). A Sweep Algorithm for the Mix Fleet Vehicle Routing Problem. *Proceedings of the International MultiConference of Engineers and Computer Scientists* , (pp. 1914-1919). Hong Kong.

Taniguchi, E., & Tamagawa, D. (2005). Evaluating city logistics measures considering the behavior of several stakeholders. *Journal of the Eastern Asia Society for Transportation Studies 6*, 3062-3076.

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

Tavasszy, L. (2008). Freight modeling: an overview of international experiences. In NASEM, *Freight demand modeling: tools for public-sector decision making* (pp. 47-55). Washington, DC, USA: The National Academies Press.

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.

Wang, L. (2009). Cutting plane algorithms for the inverse mixed integer linear programming problem. *Operations Research Letters 37*, 114-116.

Wang, Q. (2008). *Tour-based urban freight travel demand models (doctoral dissertation).* Retrieved from http://digitool.rpi.edu:8881/dtl_publish/28/11979.html

Wang, Q., & Holguín-Veras, J. (2009). Tour-based entropy maximization formulations of urban freight demand. *Transportation Research Board 88th Annual Meeting.* Washington, DC, USA.

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

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.

Xu, Z., Kang, J. E., & Chen, R. (2017). A random utility based estimation framework for the household activity pattern problem. *Transportation Research Procedia 23*, 809-826.

You, S. I. (2012). *Methodology for tour-based truck demand modeling (doctoral dissertation).* Retrieved from https://www.researchgate.net/publication/298376390

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.

Zhou, W., Chen, Q., & Lin, J. (2014). Empirical study Of commercial vehicle tour patterns in urban area in Texas. *Transportation Research Board 93rd Annual Meeting.* Washington, DC, USA.

Appendices

Appendix A: Interview with transportation planner

On July 18th, 2018, a semi-structured phone interview was held with Harrie Tissen, transportation planner at Rensa BV, a Dutch wholesaler of heating and ventilation products. This appendix provides a global transcription of this interview.

Q: Wat voor goederen vervoeren jullie?

A: De goederen die wij vervoeren hebben betrekking op verwarming en ventilatie.

Q: Doen jullie alleen aan eigen vervoer, zijn het jullie eigen goederen?

A: Ja, al vervoeren we inmiddels ook sanitair, maar dat is van een bedrijf dat we hebben overgenomen. Het is in principe eigen vervoer.

Q: Over welke afstand worden deze zendingen over het algemeen vervoerd?

A: Wij vervoeren in principe over heel Nederland, en een klein stukje over de grens in Duitsland en België. Niet veel verder dan 10-15 km over de grens.

Q: Zou u stapsgewijs kunnen beschrijven hoe u met de planning voor het transport op een dag komt?

A: We controleren gedurende de hele dag de orders die binnenkomen, op juistheid van volume en afmetingen. Dan krijgen wij van de afdeling verkoop orders door die op een bepaald tijdstip geleverd moeten worden, die tijdstippen hangen we ook aan de orders. Als we dat allemaal gedaan hebben, dan gaan we er een rittenplanning van maken. Dat zijn zowel trailers, bakwagens, en busjes.

Q: Maken jullie hierbij gebruik van planningssoftware?

A: Ja, wij plannen met het programma Smartour.

Q: In hoeverre is dat slechts een druk op de knop? Wat is de rol van deze software?

A: Wij hebben acht vestigingen in Nederland, op de meeste vestigingen staan twee auto's, in Moordrecht vier, en in Doetinchem staat dan de rest van de vloot. We plannen altijd eerst de ritten van de vestigingen vol, als we dat klaar hebben, dan gaan we vanuit Doetinchem de rest plannen. De planner doet dat zelf, en de software assisteert hem daarin in zoverre dat het de openingstijden en tijdafspraken bewaakt, en de optimale rit weergeeft. Maar het is de planner uiteindelijk die bepaalt wat er precies in de auto terechtkomt, om in ieder geval zo veel mogelijk zendingen op een optimale manier met een auto mee te geven.

Q: Is er veel speling in de dag waarop een zending kan worden afgeleverd?

A: Bij ons is het allemaal 24-uurslevering, alles wat vandaag besteld wordt, moet dus in principe morgen geleverd worden. Het kan wel eens zijn dat klanten zendingen eerder van tevoren bestellen, of dat er uitloop is en dat orders vertraging oplopen en dus uitgesteld worden, maar dat is meer uitzondering dan regel.

Q: Hoe verloopt de interactie met de chauffeur bij jullie? Overleggen jullie tijdens de rit om dingen aan te passen aan de planning?

A: Dat kan, in principe als er wensen zijn van klanten onderweg, dan is het vaak de planning of de service die met de chauffeur contact heeft om dit aan te passen. We hebben inmiddels overal boardcomputers in zitten, en de volgende stap is eigenlijk dat we via de boardcomputers communiceren met de chauffeur, omdat daar de opdrachten in komen.

Q: Wat zijn voor jullie de voornaamste doelen in het plannen van ritten? Wanneer vindt u het een goede/sterke planning voor een dag?

A: We proberen met zo weinig mogelijk zo veel mogelijk goederen weg te brengen, dus zo weinig mogelijk kilometers en uren te maken. Maar wel zo dat in het magazijn de auto's fatsoenlijk beladen kunnen worden, de zendingen niet beschadigd raken, en dat we er ook zeker van zijn dat de chauffeur de andere dag de gelegenheid krijgt om de goederen te lossen. Dus dat we ze niet zo zwaar belasten dat ze aan het eind van de dag moeten zeggen dat ze een aantal stops niet gehaald hebben. Onze chauffeurs werken in principe 45 uur per week, dus dat zijn ritten van 9 uur per dag. Dat heeft ook weer verband met wetgeving rondom de Rij- en Rusttijdenwet. Door dat in de planning goed te doen, zitten we altijd aan de veilige kant.

Q: Dus jullie proberen zendingen in een vrachtwagen samen te voegen, op zo'n manier dat de kans klein is dat het niet lukt om ze allemaal af te leveren in de shift van de chauffeur.

A: Ja, en dat de chauffeur zich aan de Rij- en Rusttijdenwet kan houden, en dat we niet iedere dag hoeven te kijken welke chauffeur de volgende dag meer of minder uren moet gaan maken.

Q: In welk aandeel van de ritten die jullie plannen wordt maar 1 klant aangedaan?

A: Minder dan 1% procent. Soms rijdt een chauffeur wel naar een klant, en dan komt hij daarna terug om andere klanten te doen. Dus dat is maar heel zelden. Als dat wel zo is, dan geven we hem meestal extern weg, tegen een basisprijs, dat we daar in ieder geval zelf geen last van hebben, en de auto's bij onszelf dus optimaal kunnen blijven presteren. Dus we zitten eigenlijk wel gemiddeld 18-20 stops voor een auto te plannen. Het komt ook wel voor dat auto's te ver weg gaan en dat we maar 6 stops plannen. Maar dichtbij huis zie je ook wel veel 30-35 stops.

Q: Is dat dan het aantal stops dat over de hele dag wordt gemaakt? Of meer van vertrek van standplaats tot terugkomst?

A: Dus van vertrekken van standplaats tot eindigen op de standplaats.

Q: Hebben jullie in principe een voorkeur voor simpele ritten, dus ritten met minder stops?

A: Nee, daar wordt in principe niet naar gekeken. We proberen zo veel mogelijk die 9 werkuren vol te plannen in de rit. En dat dit dusdanig gepland kan worden dat we qua werkuren, kilometers, en CO_2 uitstoot niet te veel in doen.

Q: Heeft u een idee in welke gevallen er vaker ritten met meer stops voorkomen? Is dat vaker bij zendingen over lange afstand bijvoorbeeld, of bepaalde goederentypes?

A: Het zijn de auto's die kort bij de standplaats rondrijden die de meeste stops wegbrengen, die zijn binnen 20 minuten bij hun eerste klant. Die maken dan een rondrit, die kunnen er dan wel 30-35 doen. En een auto die eerst 100-150 km moet rijden, die is qua rij-uren al 2 uur bezig voor die kan gaan lossen, en dat moet hij ook nog terugrijden. Dan ben je dus een halve dag aan het rijden en een halve dag aan het lossen. Vaak proberen we er wel naar te plannen dat andere klanten in de buurt ook aangedaan worden, maar als je met volumes zit dan wil dat niet altijd lukken. Dan heb je wel eens chauffeurs die met 4 stops de hele dag vol hebben.

Q: En bijvoorbeeld bij winkelbevoorrading, vinden daar over het algemeen meer of minder stops plaats?

A: Ze lopen gewoon in het programma mee, maar het is vooral in de winkelgebieden en binnensteden waar er rekening mee moet worden gehouden dat de chauffeur er voor een bepaald tijd daar moet zijn. Die informatie staat allemaal weer in het plansysteem. Daar kun je dan heel makkelijk mee werken, je ziet al die tijden, dus je gaat het ook niet dwars door die tijden plannen. Dan geeft hij al aan dat een openingstijd niet kan worden gehaald en dat dus de rit niet zo kan worden gereden. Dit planningssysteem is dus meer een ondersteuning voor de planner.

Q: Hoe wordt nu die keuze gemaakt over welke zendingen samengevoegd worden in een rit?

A: Dat is vaak een kwestie van de levertijden plus het volume. Dan proberen we ook de chauffeurs zo veel mogelijk bij hun vaste klanten en vaste ritten te houden. Omdat ze daar goed te weg weten, maar ook weten welke wensen de klant heeft, en hoe ze daar moeten laden of lossen. Dat is ook een extra service naar klant, dat die vaak hetzelfde gezicht ziet. Dat die ook weet, ik moet achterom lossen. Dat zijn allemaal van die zaken die je niet in het plansysteem kunt vangen, die kennis ligt bij de chauffeur. Een planningssysteem is een ondersteuning, maar de rest van de informatie ligt toch meestal bij de chauffeur en de planner.

Q: U sprak over volume, heeft u het idee dat volume vaker een rol speelt dan gewicht in wat mogelijk is?

A: Ja, bij ons hier zeker, omdat het gewicht vaak lager is dan het volume. Dan moet je denken aan vloerisolatie, grote boilers, hoge zonnepanelen. Ze staan op een pallet, en het gewicht is misschien 100kg, maar ze nemen wel het volume van een pallet in.

Q: Staan de zendingen bij jullie vaak op pallets?

A: De meeste goederen staan op pallets. Wat wij aan kleine goederen versturen, dat wordt via een geautomatiseerd systeem verzameld, en dat wordt dan op een pallets gestapeld in de juiste volgorde van het lossen. Dus ook die zendingen staan dan vaak op een pallet.

Q: Dus er staan soms ook zendingen van meerdere klanten op een pallet?

A: Ja, dan heeft de chauffeur een lijst bij met wat er allemaal op die pallet staat, en hoeveel doosjes. Dan weet hij ook precies, ik moet van die pallets zoveel van die doosjes pakken voor die klant. Die zitten ook nog in de goede stopvolgorde, dus dan maak je het op die manier makkelijker voor de chauffeur om de juiste spullen te pakken. Dan wordt de kans op fouten kleiner, en heb je meer tevreden klanten.

Q: Verschilt de hoeveelheid zendingen die jullie per dag hebben veel?

A: Ja op de maandag en dinsdag is er veel volume en is het vrij druk, en dat wordt in de loop van de week weer wat minder. Wij leveren veel aan bouwplaatsen, die mannen willen dan aan het begin van de week hun spullen hebben, en daar gaan ze dan de hele week mee aan de slag. En wat er dan nog in de loop van de week ontbreekt, wordt er nog bijbesteld. Dus het meeste volume gaat er aan het begin van de week uit.

Q: Is het dan ook zo dat jullie ritten plannen met meer stops in die eerste twee dagen, omdat je gewoon meer klanten hebt?

A: Ja. We huren dan vaak ook nog weleens wat bedrijven in die spullen voor ons wegbrengen, daar hebben we een drietal partners voor die dat voor ons regelen. We hebben zo’n beetje 45 auto’s van onszelf, en dan zitten er dagelijks zo’n 10-15 inhuurauto’s bij, dus dan kom je op ongeveer 60 ritten per dag.

APPENDIX B: MULTICOLLINEARITY STATISTICS

The Tolerance and VIF statistics indicate that there are no multicollinearity issues in the ET first shipment and ET later shipments model (Table B.1 and Table B.2). The Tolerance is much higher than 0.10 and the VIF is much lower than 10 for each parameter. Proximity was removed from the ET first shipment model. In Table B.3 we see that inclusion of this parameter leads to a VIF that approaches 10, and Table 6.1 shows that inclusion leads to large changes in the 'tour duration' parameter.

Table B.1. Multicollinearity statistics of the ET first shipment model (specification F).

ET first shipment		
Variable	Tolerance	VIF
$\sqrt{\textit{Tour duration [h]}}$	0.944	1.06
(Weight/capacity)^2	0.564	1.772
any port	0.715	1.398
any loading DC	0.822	1.217
any unloading DC	0.854	1.171
any urban zone	0.956	1.046
vehicle type [truck]	0.918	1.089
[truck + trailer]	0.817	1.225
[tractor + trailer]	-	-
[other/special]	-	-
NSTR tour [0: agricultural]	0.916	1.092
[1: foodstuffs]	0.869	1.151
[2-5: fuels, oils, metals]	0.986	1.014
[6: construction materials]	0.570	1.753
[7: manure/fertilizers]	0.992	1.008
[8: chemical products]	0.855	1.169
[9: machinery and other]	-	-

Table B.2. Multicollinearity statistics of the ET later shipments model (specification F).

ET later shipments		
Variable	Tolerance	VIF
Tour duration [h]	0.550	1.818
Weight/capacity	0.783	1.277
Proximity nearest shipment [km]	0.817	1.224
Ln(# stops)	0.535	1.870
any port	0.849	1.178
any loading DC	0.618	1.617
any unloading DC	0.594	1.683
any urban zone	0.856	1.168
vehicle type [truck]	0.849	1.177
[truck + trailer]	0.935	1.069
[tractor + trailer]	-	-
[other/special]	-	-
NSTR tour [0: agricultural]	0.829	1.206
[1: foodstuffs]	0.709	1.41
[2-5: fuels, oils, metals]	-	-
[6: construction materials]	0.939	1.065
[7: manure/fertilizers]	0.969	1.031
[8: chemical products]	0.795	1.259
[9: machinery and other]	-	-

Table B.3. Multicollinearity statistics when 'proximity' is added to the ET first shipment model (specification F).

ET first shipment		
Variable	Tolerance	VIF
$\sqrt{\textit{Tour duration [h]}}$	0.157	6.357
(Weight/capacity)^2	0.966	1.035
Proximity nearest shipment [km]	0.159	6.279

Appendix C: Different runs Select Shipment estimation

Since the choice set formation is a process that includes random components, we analyzed the robustness of the parameter estimates of the SS choice model with an extra run for models E to L. In Table C.1, we can see that all parameters are very robust when the estimation is repeated with each model specification.

Table C.1 Comparison of SS estimation results for different runs.

Specification (run)	E(1)	E(2)	E(3)	F(1)	F(2)	G(1)	G(2)	H(1)	H(2)	I(1)	I(2)	J(1)	J(2)	K(1)	K(2)	L(1)	L(2)
$R^2_{McFadden}$	0.189	0.187	0.186	0.170	0.171	0.277	0.276	0.251	0.248	0.187	0.187	0.169	0.171	0.277	0.276	0.249	0.250
LL	-63109	-63258	-63291	-73839	-73815	-59403	-59471	-73614	-73928	-63256	-63271	-73929	-73799	-59413	-59459	-73834	-73752
N	43409	43409	43409	37112	37112	45851	45851	41001	41001	43409	43409	37112	37112	45851	45851	41001	41001
additional generalized cost [€]	-0.005	-0.005	-0.005	-0.005	-0.005	-0.010	-0.010	-0.009	-0.009	-0.005	-0.005	-0.005	-0.005	-0.010	-0.010	-0.010	-0.009
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
additional number of stops	-1.043	-1.027	-1.031	-1.187	-1.092	-1.150	-1.139	-1.189	-1.172	-1.039	-1.038	-1.088	-1.091	-1.160	-1.148	-1.176	-1.181
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)
Same NSTR	2.332	2.338	2.322	2.759	2.727	2.235	2.221	2.657	2.648	2.313	2.326	2.712	2.715	2.218	2.227	2.627	2.646
	(0.038)	(0.038)	(0.038)	(0.042)	(0.042)	(0.038)	(0.038)	(0.041)	(0.041)	(0.038)	(0.038)	(0.042)	(0.042)	(0.038)	(0.038)	(0.041)	(0.041)

Table C.2. Description of tested model specifications for the SS model

Specification	Choice set size	Tour sequence algorithm	'Weight/capacity > 1.1' excluded	'Proximity > X' excluded	Concrete excluded
A	2	1	No	No	No
B	6	1	No	No	No
C	6	2	No	No	No
D	6	2	Yes	No	No
E	6	2	Yes	Yes (X=100)	No
F	11	2	Yes	Yes (X=100)	No
G	6	2	Yes	Yes (X=150)	No
H	11	2	Yes	Yes (X=150)	No
I	6	2	Yes	Yes (X=100)	Yes
J	11	2	Yes	Yes (X=100)	Yes
K	6	2	Yes	Yes (X=150)	Yes
L	11	2	Yes	Yes (X=150)	Yes

APPENDIX D: RETAIL ZONES

Retail zones are distinguished based on the percentage and number of retail establishments in each zone, reported in the CBS Kerncijfers Wijken en Buurten 2015. Retail establishments fall under the SBI categories G and I (business and food service). The goal is to distinguish zones where we see a particularly large and dense cluster of retail activities. Therefore we choose to mark a zone as a retail zone, if the percentage of retail establishments is above 40% and the number of retail establishments is above 100. In Figure D.1 and Figure D.2 it can be seen that zones with a percentage and number of retail establishments higher than these thresholds are rather rare. Setting these thresholds at a high end of the distribution like this, allows us to obtain a list of clearly distinguishable zones with a lot of retail activities. As seen in Figure 4.17, still quite a substantial share of tours visits a retail zone with this definition.

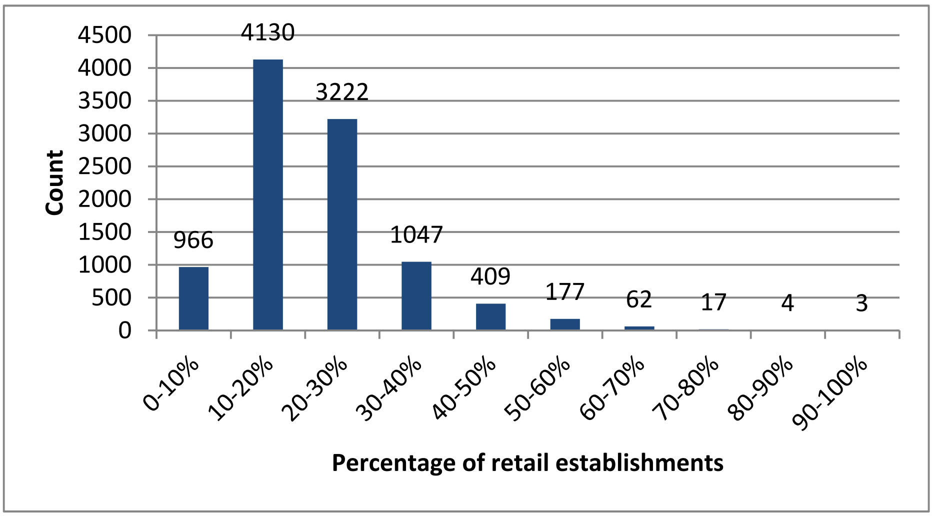

Figure D.1. Frequency distribution of the number of zones by percentage of retail establishments.

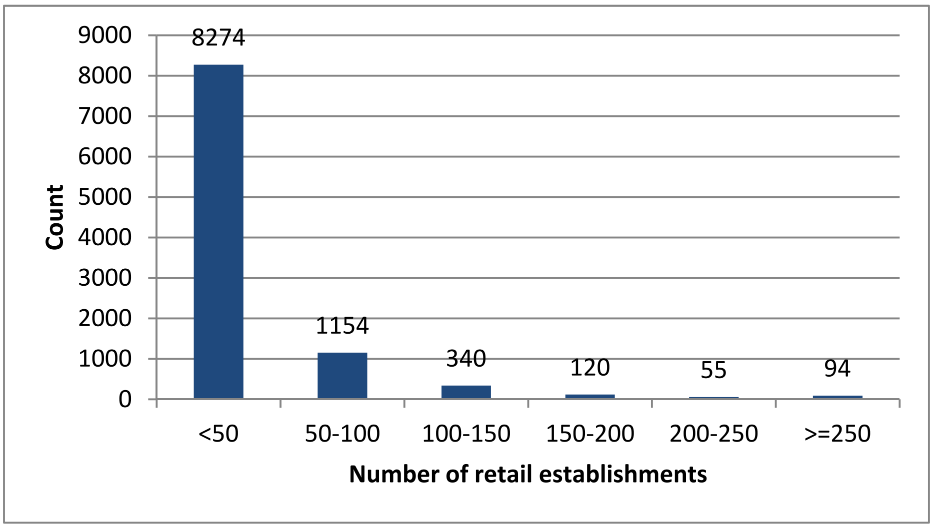

Figure D.2. Frequency distribution of the number of zones by number of retail establishments.

For 2290 of the zones the number of retail establishments is unknown. Further analyses showed that all of these zones have a total number of establishments of any kind that is lower than or equal to twenty. These are clearly not zones with a large and dense cluster of retail activities. Therefore these zones are not distinguished as a retail zone.

APPENDIX E: DISTRIBUTIONS IN THE END TOUR CHOICE DATA

In this appendix, the distributions of proximity, tour duration, capacity utilization, and tour distance are reported for the data used to estimate the ET choice models.

Table E.1. Frequency distribution of ET choice observations by proximity.

Proximity nearest shipment	ET choice			
	Add shipment		End tour	
	Count	Cum. % column	Count	Cum. % column
<10km	24739	40.87%	93601	48.09%
10-20km	10232	57.78%	23105	59.96%
20-30km	5402	66.70%	15650	68.00%
30-40km	4336	73.86%	12025	74.18%
40-50km	3203	79.16%	5205	76.86%
50-60km	2005	82.47%	4102	78.96%
60-70km	1924	85.65%	5080	81.58%
70-80km	2016	88.98%	3768	83.51%
80-90km	1046	90.71%	4218	85.68%
90-100km	1055	92.45%	2646	87.04%
100-110km	948	94.02%	2829	88.49%
110-120km	543	94.91%	3452	90.27%
120-130km	813	96.26%	2796	91.70%
130-140km	464	97.02%	2595	93.03%
140-150km	331	97.57%	2530	94.33%
150-160km	197	97.90%	1781	95.25%
160-170km	230	98.28%	1549	96.05%
170-180km	222	98.64%	1210	96.67%
180-190km	217	99.00%	1059	97.21%
190-200km	104	99.17%	1458	97.96%
>= 200km	501	100.00%	3969	100.00%

Table E.2. Frequency distribution of ET choice observations by tour duration.

Tour duration	ET choice			
	Add shipment		End tour	
	Count	Cum. % column	Count	Cum. % column
<1h.	27277	45.07%	157809	80.05%
1-2h	19419	77.15%	27070	93.78%
2-3h	8969	91.97%	7912	97.79%
3-4h	3141	97.16%	2520	99.07%
4-5h	1017	98.84%	1347	99.75%
5-6h	297	99.33%	330	99.92%
6-7h	141	99.56%	89	99.97%
7-8h	95	99.72%	32	99.98%
8-9h	43	99.79%	16	99.99%
more than 9h	129	100.00%	20	100.00%

Table E.3. Frequency distribution of ET choice observations by capacity utilization.

Capacity utilization	ET choice			
	Add shipment		End tour	
	Count	Cum. % column	Count	Cum. % column
0-10%	16527	27.30%	33163	16.82%
10-20%	8317	41.05%	12902	23.37%
20-30%	8736	55.48%	8130	27.49%
30-40%	3894	61.91%	7340	31.21%
40-50%	3429	67.58%	7275	34.90%
50-60%	4146	74.43%	25176	47.67%
60-70%	2862	79.16%	7327	51.39%
70-80%	2559	83.38%	8764	55.84%
80-90%	1139	85.26%	21711	66.85%
90-100%	1215	87.27%	25726	79.90%
100-110%	1276	89.38%	28419	94.31%
110-120%	638	90.43%	2218	95.44%
120-130%	824	91.80%	1796	96.35%
130-140%	557	92.72%	1342	97.03%
140-150%	369	93.33%	1007	97.54%
150+ %	4040	100.00%	4849	100.00%

Table E.4. Frequency distribution of ET choice observations by tour distance.

Tour distance	ET choice			
	Add shipment		End tour	
	Count	Cum. % column	Count	Cum. % column
<100km	32745	54.10%	160724	81.53%
100-200km	18213	84.19%	25278	94.35%
200-300km	6752	95.34%	7883	98.35%
300-400km	1927	98.53%	2181	99.45%
400-500km	492	99.34%	827	99.87%
>= 500km	399	100.00%	252	100.00%

APPENDIX F: DISTRIBUTIONS IN THE SELECT SHIPMENT CHOICE DATA

In this appendix, the distributions of the additional travel time, distance, number of stops, and generalized travel cost are reported for the data used to estimate the SS choice model.

Table F.1. Frequency distribution of SS choice observations by additional travel time.

Additional travel time	**SS choice**			
	Chosen shipment		Sampled unchosen shipment	
	Count	Cum. % column	Count	Cum. % column
<0.5h	36831	66.51%	11073	20.00%
0.5-1h	11464	87.21%	10103	38.24%
1-1.5h	4860	95.99%	11343	58.72%
1.5-2h	1418	98.55%	9564	76.00%
2-2.5h	625	99.68%	6431	87.61%
2.5-3h	143	99.94%	3401	93.75%
>=3h	35	100.00%	3461	100.00%

Table F.2. Frequency distribution of SS choice observations by additional travel distance.

Additional travel distance	**SS choice**			
	Chosen shipment		Sampled unchosen shipment	
	Count	Cum. % column	Count	Cum. % column
<50km	40021	72.27%	13212	23.86%
50-100km	9328	89.12%	10839	43.43%
100-150km	4104	96.53%	10828	62.99%
150-200km	1305	98.88%	8650	78.61%
200-250kmh	474	99.74%	5633	88.78%
250-300km	106	99.93%	3013	94.22%
>=300km	38	100.00%	3201	100.00%

Table F.3. Frequency distribution of SS choice observations by additional number of stops.

Additional number of stops	**SS choice**			
	Chosen shipment		Sampled unchosen shipment	
	Count	Cum. % column	Count	Cum. % column
0	13874	25.05%	4593	8.29%
1	36098	90.24%	13056	31.87%
2	5404	100.00%	37727	100.00%

Table F.4. Frequency distribution of SS choice observations by additional generalized cost.

Additional generalized travel cost	**SS choice**			
	Chosen shipment		Sampled unchosen shipment	
	Count	Cum. % column	Count	Cum. % column
<50€	40206	72.61%	13022	23.52%
50-100€	10018	90.70%	11969	45.13%
100-150€	3647	97.28%	11997	66.79%
150-200€	1123	99.31%	8916	82.90%
200-250€	310	99.87%	5016	91.95%
>=250€	72	100.00%	4456	100.00%

Appendix G: Validation data for separate runs

In Section 7.1, predicted distributions are averaged over three model runs because the tour formation model includes probabilistic and random components. In this appendix, the predicted distributions of number of stops and tour distances are reported separately for each of the three runs.

Table G.1. Observed and predicted percentage tours by number of stops.

Model	Run	Number of stops													
		1 to 2	3	4	5	6	7	8	9	10	11	12	13	14	≥15
A	1st	92.49%	2.18%	1.80%	1.21%	0.84%	0.54%	0.32%	0.18%	0.11%	0.08%	0.05%	0.05%	0.04%	0.11%
	2nd	92.57%	2.08%	1.78%	1.27%	0.84%	0.52%	0.33%	0.16%	0.12%	0.08%	0.05%	0.05%	0.03%	0.12%
	3rd	92.49%	2.21%	1.78%	1.21%	0.87%	0.51%	0.32%	0.18%	0.13%	0.08%	0.05%	0.04%	0.03%	0.11%
B	1st	92.52%	2.09%	1.74%	1.17%	0.85%	0.57%	0.35%	0.22%	0.14%	0.08%	0.06%	0.05%	0.03%	0.14%
	2nd	92.59%	2.11%	1.73%	1.16%	0.83%	0.55%	0.33%	0.20%	0.14%	0.09%	0.06%	0.05%	0.03%	0.13%
	3rd	92.57%	2.11%	1.69%	1.18%	0.84%	0.55%	0.35%	0.22%	0.13%	0.09%	0.06%	0.05%	0.03%	0.13%
C	1st	93.02%	1.90%	1.61%	1.13%	0.77%	0.51%	0.35%	0.20%	0.14%	0.08%	0.06%	0.05%	0.02%	0.14%
	2nd	93.03%	1.96%	1.64%	1.06%	0.79%	0.52%	0.31%	0.20%	0.13%	0.08%	0.07%	0.05%	0.03%	0.14%
	3rd	92.97%	1.94%	1.61%	1.10%	0.83%	0.52%	0.33%	0.20%	0.12%	0.08%	0.06%	0.05%	0.05%	0.13%

Table G.2. Observed and predicted percentage of tours by tour distance.

Tour distance	Model and run								
	A			B			C		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
<50km	72.53%	72.50%	72.45%	72.37%	72.41%	72.41%	71.92%	71.93%	71.84%
50-100km	10.04%	10.08%	10.08%	10.09%	10.16%	10.16%	10.49%	10.52%	10.57%
100-150km	7.62%	7.62%	7.65%	7.72%	7.67%	7.67%	7.90%	7.94%	7.89%
150-200km	4.34%	4.33%	4.38%	4.38%	4.40%	4.37%	4.51%	4.49%	4.51%
200-250km	2.56%	2.57%	2.57%	2.54%	2.53%	2.53%	2.53%	2.54%	2.55%
250-300km	1.06%	1.06%	1.05%	1.11%	1.03%	1.06%	1.01%	0.99%	1.02%
300-350km	0.65%	0.65%	0.64%	0.64%	0.66%	0.64%	0.61%	0.59%	0.59%
350-400km	0.43%	0.41%	0.40%	0.39%	0.42%	0.41%	0.37%	0.37%	0.37%
400-450km	0.26%	0.28%	0.26%	0.30%	0.26%	0.27%	0.24%	0.23%	0.23%
450-500km	0.17%	0.18%	0.17%	0.16%	0.16%	0.16%	0.15%	0.14%	0.15%
500-550km	0.12%	0.12%	0.11%	0.11%	0.12%	0.11%	0.10%	0.09%	0.09%
550-600km	0.08%	0.07%	0.08%	0.09%	0.06%	0.08%	0.06%	0.05%	0.06%
600-650km	0.05%	0.06%	0.06%	0.04%	0.05%	0.06%	0.04%	0.05%	0.04%
650-700km	0.04%	0.03%	0.04%	0.03%	0.04%	0.02%	0.03%	0.03%	0.03%
700-750km	0.02%	0.03%	0.02%	0.03%	0.02%	0.02%	0.02%	0.02%	0.02%
750-800km	0.01%	0.01%	0.01%	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%
800-850km	0.01%	0.01%	0.01%	0.01%	0.00%	0.01%	0.01%	0.01%	0.01%
850-900km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
900-950km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
950-1000km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
≥1000km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

APPENDIX H: SENSITIVITY DATA FOR SEPARATE RUNS

In Section 7.2, predicted distributions for different travel time scenarios are averaged over two model runs because the tour formation includes probabilistic and random components. In this appendix, the predicted distributions of number of stops and tour distance are reported separately for both runs.

Table H.1. Observed and predicted percentage of tours by number of stops in the four travel time scenarios.

Scenario	Run	Number of stops													
		1-2	3	4	5	6	7	8	9	10	11	12	13	14	≥15
+50%	1st	91.46%	2.63%	2.17%	1.36%	0.93%	0.56%	0.34%	0.17%	0.10%	0.07%	0.06%	0.04%	0.04%	0.07%
	2nd	91.46%	2.58%	2.22%	1.41%	0.93%	0.52%	0.32%	0.17%	0.11%	0.07%	0.05%	0.05%	0.03%	0.07%
+10%	1st	92.31%	2.27%	1.86%	1.25%	0.86%	0.51%	0.32%	0.18%	0.12%	0.08%	0.06%	0.04%	0.03%	0.11%
	2nd	92.33%	2.22%	1.87%	1.22%	0.87%	0.53%	0.30%	0.19%	0.13%	0.08%	0.06%	0.05%	0.03%	0.11%
-10%	1st	92.76%	2.09%	1.73%	1.16%	0.80%	0.52%	0.30%	0.18%	0.11%	0.08%	0.05%	0.04%	0.04%	0.11%
	2nd	92.83%	2.05%	1.69%	1.12%	0.82%	0.52%	0.33%	0.19%	0.11%	0.08%	0.05%	0.04%	0.04%	0.12%
-50%	1st	93.77%	1.66%	1.38%	1.01%	0.73%	0.47%	0.31%	0.19%	0.11%	0.08%	0.05%	0.04%	0.04%	0.17%
	2nd	93.86%	1.65%	1.34%	0.98%	0.73%	0.48%	0.30%	0.17%	0.12%	0.07%	0.06%	0.05%	0.03%	0.15%

Table H.2. Observed and predicted percentage of tours by tour distance in the four travel time scenarios.

Tour distance	Scenario and run							
	+50%		+10%		-10%		**-50%**	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
<50km	72.32%	72.28%	72.48%	72.39%	72.50%	72.53%	72.58%	72.60%
50-100km	9.79%	9.83%	10.03%	10.02%	10.17%	10.18%	10.45%	10.58%
100-150km	7.64%	7.67%	7.68%	7.66%	7.70%	7.66%	7.74%	7.68%
150-200km	4.50%	4.43%	4.32%	4.45%	4.31%	4.32%	4.31%	4.27%
200-250km	2.67%	2.68%	2.51%	2.53%	2.44%	2.50%	2.40%	2.33%
250-300km	1.15%	1.19%	1.09%	1.08%	1.05%	1.01%	0.90%	0.90%
300-350km	0.74%	0.75%	0.69%	0.67%	0.64%	0.63%	0.53%	0.54%
350-400km	0.47%	0.48%	0.42%	0.44%	0.42%	0.39%	0.35%	0.35%
400-450km	0.29%	0.28%	0.30%	0.28%	0.24%	0.26%	0.20%	0.23%
450-500km	0.20%	0.18%	0.18%	0.17%	0.17%	0.18%	0.16%	0.15%
500-550km	0.11%	0.12%	0.12%	0.13%	0.11%	0.12%	0.11%	0.10%
550-600km	0.07%	0.07%	0.07%	0.07%	0.08%	0.09%	0.08%	0.07%
600-650km	0.03%	0.02%	0.05%	0.06%	0.06%	0.06%	0.06%	0.06%
650-700km	0.01%	0.01%	0.04%	0.03%	0.04%	0.03%	0.03%	0.04%
700-750km	0.00%	0.00%	0.02%	0.02%	0.03%	0.02%	0.03%	0.03%
750-800km	0.00%	0.00%	0.01%	0.01%	0.01%	0.01%	0.02%	0.03%
800-850km	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.02%
850-900km	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.02%	0.01%
900-950km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%
950-1000km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%
≥1000km	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.02%

Appendix I: Comparison estimation and validation data

We divide the data for estimation of model parameters and the data for validation of the model. In this appendix, we report the distribution of the number of stops, goods type, and vehicle type in both data sets. Two types of data divisions were made: by day + carrier and by carrier. With the first division, data of all carriers is found in both the estimation and validation data, in contrast to the second division. In Tables I.1 to I.3, we see that the estimation and validation data set are more heterogeneous when we divide the data by carriers.

Table I.1. Percentage of tours by number of stops in the estimation and validation data set.

Number of stops	Data divided by **day + carriers**		Data divided by **carriers**	
	Estimation data	Validation data	Estimation data	Validation data
1 to 2	92.52%	92.37%	90.76%	92.82%
3 to 5	4.65%	4.72%	6.96%	4.17%
6 to 10	2.46%	2.59%	2.02%	2.64%
≥10	0.37%	0.32%	0.26%	0.36%

Table I.2. Percentage of tours by NSTR goods type in the estimation and validation data set.

NSTR goods type	Data divided by **day + carriers**		Data divided by **carriers**	
	Estimation data	Validation data	Estimation data	Validation data
0: agricultural	2.38%	2.47%	1.52%	2.63%
1: foodstuffs	5.09%	5.25%	15.84%	2.76%
2-5: fuels, oils, metals	0.19%	0.19%	0.68%	0.08%
6: construction materials	11.14%	11.74%	4.92%	12.92%
7: manure/fertilizers	0.11%	0.10%	0.15%	0.10%
8: chemical products	53.63%	52.58%	26.98%	59.02%
9: machinery and other	27.45%	27.67%	49.90%	22.50%

Table I.3. Percentage of tours by vehicle type in the estimation and validation data set.

Vehicle type	Data divided by **day + carriers**		Data divided by **carriers**	
	Estimation data	Validation data	Estimation data	Validation data
Truck	49.91%	48.83%	11.35%	58.00%
Truck + trailer	9.35%	9.73%	11.03%	9.20%
Tractor + trailer	40.19%	40.91%	77.33%	32.19%
Other/special vehicle	0.55%	0.53%	0.28%	0.60%

APPENDIX J: SUMMARY OF RESEARCH AS A DRAFT SCIENTIFIC PAPER

A BEHAVIORAL SHIPMENT-BASED MODEL OF FREIGHT TOUR FORMATION

S. Thoen

Abstract
Tour formation is a distinguishing feature of freight transportation. An increasing amount of research is dedicated to the consideration of tour formation in freight simulation models, but these tour formation models lack statistical calibration on empirical data, do not represent shipments explicitly, or focus on a narrow segment of freight transportation. This paper presents a novel shipment-based approach to model tour formation behavior. We developed an algorithm that constructs tours through iterative allocation of an additional shipment. Two choice models provide an empirical foundation to this algorithm. Parameters of these choice models are estimated on a dataset with information regarding over two million shipments. This information is gathered automatically from the planning systems of carriers transporting goods in the Netherlands. With this model we are able to reproduce observed tour patterns excellently for a given set of shipments. In addition, the model considers many objectives and constraints that determine the tour formation process and acknowledges differences between goods, vehicle, and location types. For example, shipments at ports tend to be transported with direct tours, while tours starting at a distribution center have more stops. This model can be applied in a shipment-based freight simulation framework to construct tours for large third-party carriers in the Netherlands.

Keywords: Tour formation, freight transportation, behavioral model, big data, third-party carriers

1. INTRODUCTION

Freight transportation is crucial for economic development, yet freight trucks have many negative external impacts on society. Compared to passenger cars, freight trucks pay disproportionally large contributions to congestion, air pollution, traffic accidents, and pavement wear (Hunt & Stefan, 2007; Quak, 2008; Kim et al., 2014). As freight volumes are expected to increase only further (ITF-OECD, 2015), governments are developing policies that reduce the negative impacts of freight transportation. Simulations models are a common tool for the evaluation of these policies (de Bok et al., 2018).

One of the distinguishing features of road freight transportation is tour formation, multiple shipments are delivered often in a single tour to save transportation costs (Hunt & Stefan, 2007; Sánchez-Díaz et al., 2015). Trucks do not need to drive directly from the pick-up location to the delivery location of a shipment but may pick-up/deliver other shipments in between. Consequently, omitting tour formation in a freight simulation model causes inaccuracy of traffic flow predictions (Holguín-Veras et al., 2014).

Shipment-based models describe the tour formation process more accurately than models that predict vehicle trips directly, because shipments determine the possibilities and constraints for tour formation. Consequently, differences in the economic characteristics, constraints, and geographical distribution of goods types can be considered in shipment-based models (Holguín-Veras et al., 2014). Additionally, shipment-based models allow us to analyze the impacts of detailed policies and scenarios, such as new distribution centers and changes in delivery frequencies (Boerkamps & van Binsbergen, 1999).

In this research, we develop a behavioral tour formation model that is shipment-based and calibrated on empirical data. Calibration is performed using a data set that covers the total road freight demand in the

Netherlands, consists of 2.6 million transported shipments, and contains highly detailed information on shipment attributes and (un)loading locations. Shipment data are collected with an innovative method that allows carriers to use an XML-interface to complete and deliver surveys automatically.

This paper is organized as follows. Section 2 provides a literature overview on tour formation modeling and defines the knowledge gap that our study fills. Section 3 presents the structure of the developed tour formation model, while Section 4 describes the data used for calibration of this model. In Section 5 the estimated parameters of the model are presented and interpreted. Section 6 reports a validation study and a sensitivity analysis performed with the model. Conclusions and recommendations are formulated in Section 7.

2. Literature overview

Different approaches exist to model tour formation, we distinguish the following three: (1) tour-based entropy maximization, (2) mathematical optimization, (3) and tour construction based on Random Utility Modeling (RUM).

Wang & Holguín-Veras (2009) developed the tour-based entropy maximization, in which an Origin Destination (OD) matrix is estimated by finding the most likely set of tours that respects constraints such as zonal trip productions and attractions. Sánchez-Diáz et al. (2015) extended this concept to match traffic counts and included a time-of-day component. While this approach can connect microscopic tours to traffic counts and macroscopic OD estimation, it is not shipment-based and a tour formation model is still needed to generate the tours as input.

Mathematical optimization techniques can be used to form tours from a set of shipments. While the Vehicle Routing Problem (VRP) was developed to prescribe optimal behavior to a decision-maker, VRPs may be solved to represent observed tour formation too (e.g. Boerkamps & van Binsbergen, 1999; Taniguchi & van der Heijden, 2000; Wisetjindawat et al., 2006; Polimeni et al., 2010; Anand et al., 2014). Others apply more heuristic algorithms for behavioral tour formation modeling (e.g. Alho et al., 2017). As carriers sometimes use these optimization techniques for tour formation, it is an appropriate and straightforward modeling choice if no empirical data is available (You, 2012). To be able to base parameters relating to the objectives and constraints in a VRP on empirical data, though, You et al. (2016) applied inverse optimization to GPS truck diary data of the San Pedro Bay Ports in California, USA. This model is able to consider time windows, an important constraint in tour formation, but it is not shipment-based, it is computationally too heavy for application in a regional model, and its parameters are not calibrated in a statistical way.

RUM provides the statistical methods to calibrate parameters. These statistical methods allow us to test hypothesized effects empirically, generalize findings to a population, and control for the correlation between variables. The difficulty of tour formation, however, is that it is not possible to narrow it down to a single choice. For this reason, different algorithms have been developed that form tours using several choice steps. Hunt & Stefan (2007) pioneered this approach to behavioral tour formation modeling and applied it to the city of Calgary, Canada. Firstly, the number of tours originating in each zone is estimated. Secondly, the vehicle type and tour purpose are chosen. Thirdly, the tour is grown iteratively by choosing the next stop location until the choice is made to return to the home base.

None of the aforementioned studies are both shipment-based and statistically calibrated on empirical data. To the best of the author's knowledge, Nuzzolo et al. (2012) and Outwater et al. (2013) are the only two studies that satisfy both criteria. Nuzzolo et al. (2012) developed a model for restocking tours performed for and by retailers in Rome, Italy. Tour formation starts by deciding for each shipment the number of trips of the tour that it will be part of, after which tours are constructed with a 'next stop location' choice model. For their framework in Chicago, IL, USA, Outwater et al. (2013) model the choice for the number of stops and the tour pattern, i.e. the number of tours required to deliver all shipments. Geographically close shipments with the same tour pattern and number of stops are grouped into tours with a hierarchical clustering, 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.

While Nuzzolo et al. (2012) and Outwater (2013) have developed tour formation models that are shipment-based and statistically calibrated, their scope of application is rather limited: retailer restocking 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. Modeling the tour formation process as such is more appropriate.

3. Methodology

Our tour formation model allocates shipments to tours. Tours are grown iteratively through allocation of one additional shipment. After each shipment allocation, we consider ending the tour. If the tour is ended, then a new tour is constructed. If the tour is not ended, one shipment is selected and added to the tour and we consider ending the tour again. Two choice models can be identified in our tour formation model, the End Tour (ET) model and the Select Shipment (SS) model. The ET model has a binary dependent variable with the categories ‘0 = continue adding shipments to tour’ and ‘1 = end tour’. The SS model is a Multinomial Logit (MNL) with a choice set of γ shipments that may be added to the tour.

This incremental structure is similar to the tour-based microsimulation of Hunt & Stefan (2007). However, Hunt & Stefan (2007) chain vehicle trips into a tour, while we group shipments into a tour. Additionally, we construct a sequence of visiting locations with a separate tour sequence algorithm, while Hunt & Stefan (2007) let this sequence depend on the iteration in which trips were added to the tour.

The ‘choices’ in the ET and SS models should not be seen as actual choices made by transportation planners. Representing each choice made in tour formation is not a realistic goal, because tour formation is a complex process with many unobserved choices, for which transportation planners might use different methods and software. Instead, the two models should be seen as statistical equations that allow us to consider observed behavior in a tour formation algorithm. To represent this behavior accurately, though, it is important that meaningful effects can explain the dependent variable. For example, because the number of stops is not a choice like in Nuzzolo et al. (2012) and Outwater et al. (2013) but an outcome of the process instead, we can consider explicitly that the geographical proximity of the set of available shipments influences not only which shipments, but also how many shipments are transported in the same tour.

Figure 3.1 shows the tour formation model in a flow diagram. In Section 3.1 and Section 3.2 we discuss the building blocks of this flow diagram in more detail. Section 3.1 focuses on the part that is used to make the End Tour decision, while Section 3.2 focuses on the Select Shipment procedure.

3.1 Ending the tour

Before we can start constructing tours, it is necessary to define the universal set of shipments that we construct tours with. The universal set consist of all shipments transported on the same day by the same carrier. We use this definition of the universal set for it is not reasonable to assume that carriers can add shipments of other carriers to their tours, and an interview with a transportation planner indicated that the delivery date of a shipment is often rigid.

A random non-allocated shipment in the universal set is selected as the first shipment of a new tour. We select the first shipment at random because the tour does not have any shipments yet, there are no characteristics to base this choice on.

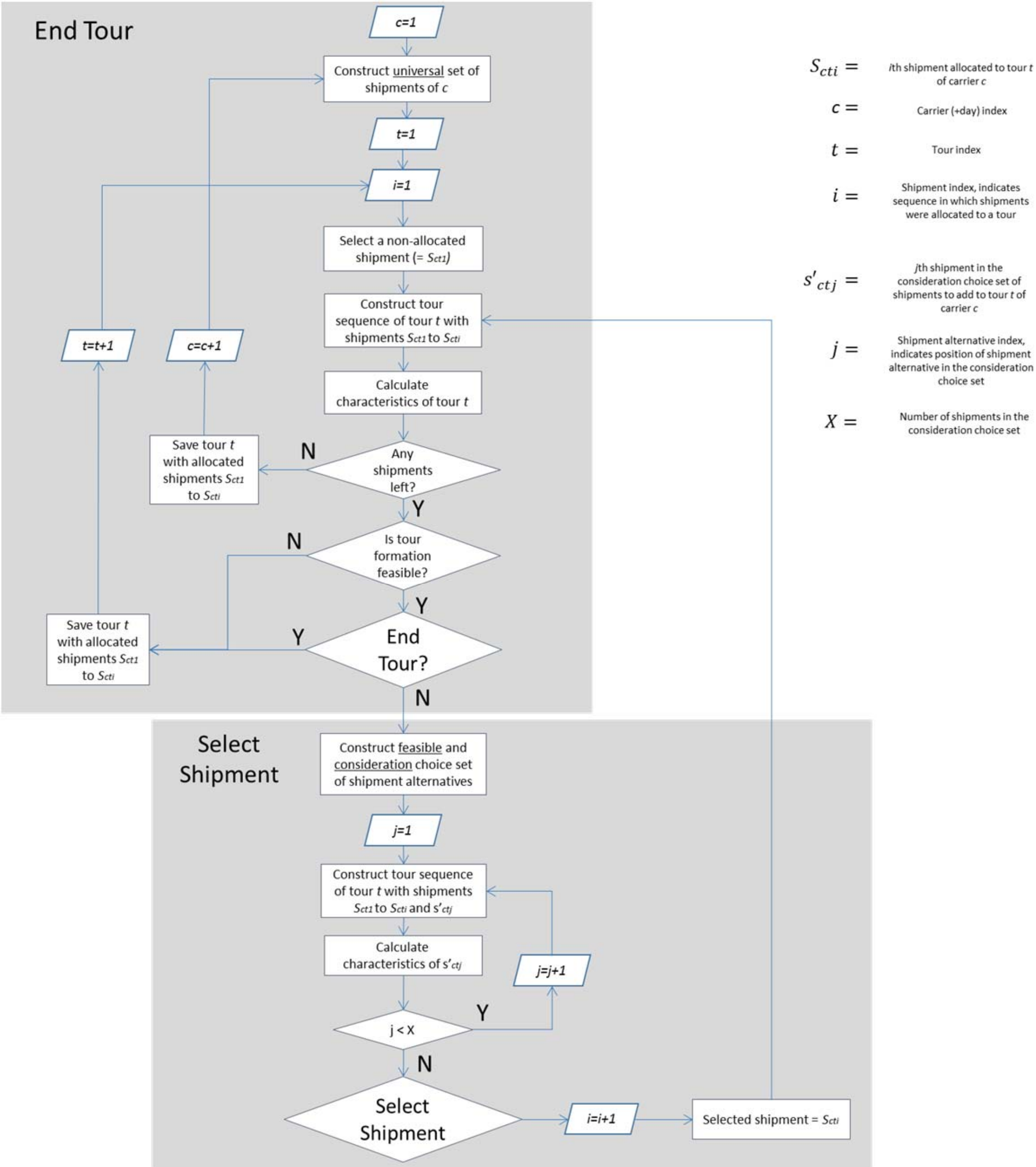

Figure 3.1 The iterative process of shipment allocation in our tour formation model.

Next, we construct the sequence of visiting the loading and unloading locations of all shipments that have been allocated so far to the tour. This sequence allows us to calculate the tour duration and, additionally, would allow us to assign vehicle trips to a network. To construct the sequence, we have developed two tour sequence algorithms. Both algorithms use a nearest neighbor search; after each location, the nearest remaining location is visited. The first algorithm visits all loading locations before unloading locations are visited, while the second algorithm visits alternately loading and unloading locations. Using more advanced algorithms to solve a Traveling Salesman Problem would lead to shorter and more logical sequences (AlSalibi et al., 2013). However, the computational efficiency of the nearest neighbor search is of large importance in this framework, as the sequence needs to be constructed each time we allocate a shipment to a tour, and sometimes for a large number of locations (>10). With our nearest neighbor search, construction of tours from 200,000+ shipments takes approximately half an hour.

Characteristics of the tour are calculated to obtain the probability that the tour is ended with the ET choice model and to check constraints. In Section 5, we report and interpret the variables that explain the ET choice. When constraints are violated, the tour is ended regardless of the probability calculated with the ET choice model. Four types of constraints are specified: (1) proximity, (2) concrete/cement, (3) vehicle capacity, and (4) work shift constraints.

Firstly, if there are no non-allocated shipments left with a proximity lower than α km to the tour as constructed so far, then we end the tour because all non-allocated shipments would require a lot of additional time. Proximity is calculated as shown in Figure 3.2. Secondly, we always end a tour with a concrete/cement shipment, for we observed virtually only direct tours (i.e. with one shipment) transporting concrete/cement, which can be explained by large shipment sizes and a high time-sensitivity (Khan & Machemehl, 2017). Thirdly, because of regulations and physical limitations, the total transported weight may not exceed the vehicle capacity. Fourthly, the tour is ended after nine hours because truck drivers are not allowed to drive for more than nine hours on more than two days of the week in the Netherlands.

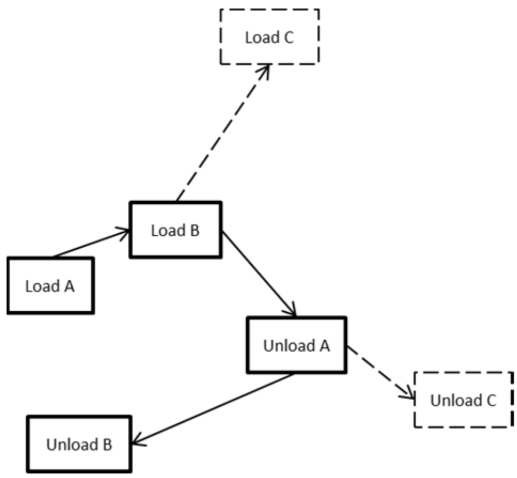

Figure 3.2 If the current tour consists of shipments A and B, then the 'proximity' of shipment C is the sum of the distance of the two dashed arrows.

3.2 Selecting an additional shipment

If the tour is not ended, we choose which shipment to add to the tour. For this purpose, we distinguish three choice sets of shipments that may be added to the tour: the universal choice set (UC), the feasible choice set (FC), and the consideration choice set (CC). We have defined the UC in Section 3.1: all shipments of the same carrier and day. The FC is a subset of the UC that respects constraints, while the CC is a randomly sampled subset of the FC.

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) concrete/cement, and (3) vehicle capacity. Shipments are removed when the proximity is larger than α km, when the shipment is concrete/cement, and when the shipment causes the total transported weight to exceed the vehicle capacity. The tour duration constraint is not checked in formation of the FC for it requires the construction of the tour sequence, which would get computationally very heavy, considering that the UC can consist of hundreds of shipments.

4. SHIPMENT DATA COLLECTION

For the development of our tour formation model, we have access to the aforementioned dataset, called the XML microdata, that is collected by the Dutch Central Bureau of Statistics (CBS). The staggering amount of data, 2.6 million shipments, has two main reasons. Firstly, CBS draws a random sample of trucks of the total Dutch fleet; for each truck that the CBS selects, carriers are obliged to report all shipments transported for a week. Secondly, carriers can install an XML-interface which allows them to send the data for the survey in a (partially) automated way from their transport management systems (de Bok et al., 2018).

A consequence of the automated data collection is that a self-selection of third-party carriers with advanced planning systems has taken place. Third-party carriers are firms that transport goods for other parties, in contrast to own-account carriers, who transport their own goods.

The data are listed as separate shipments and we know which shipments were transported in the same tour. 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 this shipment data, we use land use data[4], employment data[5], and skim matrices[6]. Land use data is used to distinguish urban and retail zones, while employment data provides the information to determine which zones have port transshipment and goods distribution activities. We distinguish zones at the level of 'buurten', a Dutch administrative unit of approximately 12,000 zones. As (un)loading zones and other variables are often not filled out, 515,810 shipment records of the approximately 2.6 million records remain for our analyses.

In Table 4.1 we see that a strikingly large portion of tours are direct (92.44%). Khan & Machemehl (2017), for example, found only 34.02% 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 concrete/cement shipments in our dataset. Due to large shipment sizes and a high time-sensitivity, multiple-stop tours are often not feasible for concrete/cement (Khan & Machemehl, 2017). Additionally, relatively short distances are observed because we only analyze tours that stay within the Netherlands. Table 4.2 shows for which goods, vehicles, and locations, direct tours are observed most often. The analyses in Table 4.2 have guided our search for explanatory variables in our model, which is why we interpret these effects in Section 5.

[4] CBS Kerncijfers wijken en buurten 2015
[5] CBS Algemeen Bedrijven Register
[6] Off-peak travel times and distances based on the shortest path with a traffic assignment of the calibrated NRM-West, a Dutch regional transportation model

Table 4.1. Descriptive tour statistics.

Tour characteristics	Frequency (tours)
Number of stops	
1-2 (direct)	365,905 (92.44%)
3-5	18,538 (4.68%)
6-10	10,008 (2.53%)
>10	1,361 (0.34%)
Tour distance [km]	
0-20	172,341 (43.54%)
20-40	82,995 (20.97%)
40-60	19,508 (4.93%)
60-80	14,976 (3.78%)
80-100	13,747 (3.47%)
100-120	14,383 (3.63%)
120-140	12,578 (3.18%)
140-160	11,749 (2.97%)
160-180	6,944 (1.75%)
180-200	6,748 (1.70%)
≥200	39,843 (10.07%)
Concrete/cement	179,468 (45.35%)
NSTR goods type[7]	
0: agricultural	9,541 (2.41%)
1: foodstuffs	20,617 (5.21%)
2-5: fuels, oils, metals	746 (0.19%)
6: construction materials	45,279 (11.44%)
7: manure/fertilizers	457 (0.12%)
8: chemical products	210,151 (53.09%)
9: machinery and other	109,021 (27.54%)
Vehicle type	
Truck	194,875 (49,37%)
Truck + trailer	37,660 (9,54%)
Tractor + trailer	160,049 (40,55%)
Other/special vehicle	2,134 (0,54%)
Any location visited in tour	
Port	102,679 (25,94%)
DC	176,249 (44,53%)
Urban zone	146,098 (36.91%)
Retail zone	48,164 (12.17%)

Table 4.2. The percentage of direct tours for different goods, vehicles, and locations.

Tour characteristics	Percentage of direct tours
Average	92.44%
Average (excl. concrete/cement)	86.18%
Concrete/cement	100.00%
NSTR goods type[7,8]	
0: agricultural	73.05%
1: foodstuffs	64.11%
2-5: fuels, oils, metals	96.92%
6: construction materials	97.67%
7: manure/fertilizers	77.02%
8: chemical products	95.25%
9: machinery and other	84.14%
Vehicle type[8]	
Truck	72.92%
Truck + trailer	96.44%
Tractor + trailer	85.26%
Other/special vehicle	97.29%
Any location visited in tour[8]	
Port loading	96.35%
Port unloading	96.01%
DC loading	68.45%
DC unloading	70.35%
Urban zone	60.97%
Retail zone	72.03%

5. ESTIMATION RESULTS

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

Estimation results with four model variations (A to D) are reported (Table 5.1). Observations that violate our tour constraints are not used for estimation; in model application we do not use the choice model either when constraints are violated. 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), for these methodological model specifications are more difficult to define intuitively than operational constraints such as vehicle capacity utilization. Fifty percent of the data (based on date) is used to estimate Model A to C. All carriers provide shipments for the estimation data sets of

[7] NSTR goods type with the highest transported weight in the tour
[8] Analysis excluding concrete/cement shipments

Model A to C. Model D tests how results differ when data regarding fifty percent of the carriers is used for estimation instead.

Table 5.1. The specifications 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 Estimates End Tour choice model

Tables 5.2 and 5.3 present respectively the estimation results of the ET first shipment model and ET later shipments model. A positive parameter implies a higher probability of ending the tour.

If the first shipment of a tour requires a longer travel time from loading to unloading, the probability of ending the tour is lower. 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 to construct direct tours to reduce the complexity of planning. Additionally, the travel time savings of grouping shipments might be smaller for these shipments. The square root indicates a stronger effect for lower tour durations; the attractiveness of a direct tour does not decrease as strongly for longer tour durations.

The probability of ending the tour increases with a larger share of the vehicle capacity (in terms of weight) used. This reflects the strategy of transportation planners to fill vehicles optimally to save transportation costs. The quadratic component 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.

Tours that visit a port transshipment zone 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. When a distribution center is visited, though, 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 than when they are unloaded. Shipments unloaded at a distribution center might be flows originating from a producer more often.

The probability of ending the tour after the first shipment is lower when an urbanized zone is visited. 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 can save a lot of time to reduce the number of trips in and out of the city.

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 of 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 on combination with other goods, and dispersion of supply/demand. Restaurants with a demand for foodstuffs (*NSTR 1*) might be concentrated in a city center, while gas stations (*NSTR 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 4.2.

Table 5.2. Estimation results ET first shipment. Cells present the 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
Constant	1.684 (0.029)	1.473 (0.027)	1.681 (0.024)
$\sqrt{\textit{Tour duration [h]}}$	-1.698 (0.037)	-1.112 (0.030)	-2.403 (0.034)
(Weight/capacity)^2	5.471 (0.102)	6.022 (0.098)	5.258 (0.088)
any port	1.588 (0.037)	1.484 (0.035)	2.354 (0.037)
any loading DC	-0.578 (0.026	-0.517 (0.024)	-0.942 (0.025)
any unloading DC	-0.475 (0.026)	-0.450 (0.024)	-0.765 (0.025)
any urban zone	-0.461 (0.038)	-0.605 (0.036)	-0.499 (0.037)
vehicle type [truck]	-1.295 (0.039)	-1.370 (0.037)	-1.684 (0.039)
[truck + trailer]	1.850 (0.049)	1.980 (0.045)	2.508 (0.047)
[tractor + trailer]	-	-	-
[other/special]	-	-	-
NSTR tour [0: agricultural]	-0.736 (0.047)	-0.881 (0.044)	-0.271 (0.037)
[1: foodstuffs]	-0.659 (0.032)	-0.808 (0.029)	-0.672 (0.037)
[2-5: fuels, oils, metals]	1.495 (0.337)	1.298 (0.324)	1.121 (0.311)
[6: construction materials]	1.452 (0.058)	1.472 (0.051)	2.253 (0.048)
[7: manure/fertilizers]	0.713 (0.253)	-	0.878 (0.237)
[8: chemical products]	0.583 (0.045)	0.530 (0.042)	1.821 (0.053)
[9: machinery and other]	-	-	-

Table 5.3. Estimation results ET later shipments. Cells present the Beta and standard error.

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
Constant	-2.526 (0.062)	-2.547 (0.060)	-2.516 (0.054)
Tour duration [h]	0.386 (0.014)	0.364 (0.014)	0.449 (0.012)
Weight/capacity	3.286 (0.057)	3.285 (0.055)	3.122 (0.048)
Proximity nearest shipment [km]	0.009 (0.001)	0.008 (0.000)	0.008 (0.000)
Ln(# stops)	-0.911 (0.042)	-0.841 (0.041)	-0.828 (0.036)
any port	0.526 (0.047)	0.545 (0.046)	0.450 (0.040)
any loading DC	-0.191 (0.036)	-0.179 (0.035)	-0.281 (0.031)
any unloading DC	0.094 (0.036)	0.078 (0.035)	-0.150 (0.031)
any urban zone	-0.145 (0.032)	-0.175 (0.032)	-0.036 (0.027)
vehicle type [truck]	-1.968 (0.061)	-1.968 (0.058)	-2.354 (0.059)
[truck + trailer]	-0.954 (0.088)	-1.003 (0.086)	-0.845 (0.085)
[tractor + trailer]	-	-	-
[other/special]	-	-	-
NSTR tour [0: agricultural]	2.226 (0.059)	2.203 (0.056)	2.182 (0.045)
[1: foodstuffs]	0.871 (0.035)	0.873 (0.033)	0.546 (0.031)
[2-5: fuels, oils, metals]	-	-	-
[6: construction materials]	0.556 (0.081)	0.538 (0.078)	0.396 (0.068)
[7: manure/fertilizers]	-1.105 (0.327)	-1.702 (0.289)	-0.888 (0.244)
[8: chemical products]	1.517 (0.063)	1.468 (0.060)	1.168 (0.062)
[9: machinery and other]	-	-	-

Most effects are similar for the ET later shipments model, but three key differences are found with the ET first shipment model: (1) the sign of tour duration switches from negative to positive, and (2) proximity nearest shipment and (3) number of stops are additional variables.

The probability of ending the tour increases with a higher tour duration 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 either customers experience a delay of a day, or the driver must work overtime.

If the nearest non-allocated shipment is closer to the tour as constructed so far, 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 number of stops shows a negative parameter. When the tour as constructed so far 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 and fifteen are considered more similar than tours with three and four stops.

Models A and B show the same estimation results, the choice set size only impacts the shipment selection, it does not influence the choice to end the tour. A more lenient proximity constraint (α=150km) has only minor impacts on the estimated parameters. Model D, estimated on a subset of the carriers, leads to larger differences with Models A to C. The only sign that changes direction is that of the 'any unloading DC' parameter in the ET later shipments model; however, in accordance with the parameters for Model A to C, the 'any unloading DC' parameter is still lower than that of 'any loading DC' in Model D.

5.2 Estimates Select Shipment choice model

Table 5.4 presents the estimation results of the SS model. A positive parameter increases the probability that an alternative (i.e. a non-allocated shipment) is selected as the additional shipment to a tour. All three variables can be considered instrumental.

The additional generalized cost is a weighted sum of the travel time (€45.12/h) and the distance (€0.45/km) the 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 make for each driven hour or kilometer (Significance, 2018). A shipment with a higher additional cost has a lower probability of being selected, carriers wish to minimize transportation costs by constructing efficient tours.

As a shipment has only two stops, one for loading and one for unloading, the additional number of stops of a shipment can be zero, one, or two. A shipment that adds more stops to the tour (or: has 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).

In 93.17% of the cases that multiple shipments are transported in a tour, we observe that all shipments have the same NSTR goods classification. Consequently, in the SS model the probability of selecting a shipment is higher if it has the same goods type as the other shipments in the tour. This may be explained with restricted goods combinations.

Table 5.4. Estimation results Select Shipment 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
additional generalized cost [€]	-0.005 (0.000)	-0.005 (0.010)	-0.010 (0.000)	-0.006 (0.000)
additional number of stops	-1.039 (0.010)	-1.088 (0.010)	-1.176 (0.010)	-0.918 (0.008)
Same NSTR	2.313 (0.038)	2.712 (0.042)	2.627 (0.041)	2.176 (0.031)

Estimation results are relatively stable for Models A to D. The pseudo-R^2 of Model C is higher and the 'additional generalized cost' 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, we have a less attractive choice set, it includes more distant shipments. Correctly predicting the observed choice is easier in such a choice set, which improves the pseudo-R^2. Distant shipments have a higher additional cost, these higher values influence the 'additional generalized cost' parameter.

6. VALIDATION AND SENSITIVITY ANALYSIS

Estimation of the ET and SS models does not provide sufficient information to judge the performance of the tour formation model. Other model aspects, such as constraints and choice set formation, influence how tours are constructed. For this reason, we test the model performance by constructing tours with the shipments in the validation data set (i.e. the data not used for estimation). Section 6.1 reports the validation study while Section 6.2 presents a sensitivity analysis.

6.1 VALIDATION

The model performance is assessed by comparing the observed tours with our predicted tours. 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 validation of 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 6.1), we can conclude that our model reproduces aggregate tour statistics excellently for a given set of shipments. In addition, the distribution of number of stops is reproduced satisfactorily for different location and goods types (Table 6.2). The model acknowledges that tours that visit a distribution center tend to have more stops. For unknown reasons, though, too many direct tours are predicted for foodstuffs (*NSTR1*).

Table 6.1. Coincidence ratio between observed and predicted distributions of number of stops and distance. Averaged over three models runs for A-C and two model runs for D.

Model	Coincidence ratio	
	Number of stops	Tour distance
A	98.81%	89.30%
B	98.98%	89.36%
C	98.57%	89.54%
D	96.92%	84.19%

Table 6.2. Coincidence ratio between observed and predicted distributions of number of stops by location and goods type. Averaged over three models runs for A-C and two model runs for D.

Model	Coincidence ratio Number of stops								
	DC visited	no DC visited	NSTR0	NSTR1	NSTR2-5	NSTR6	NSTR7	NSTR8	NSTR9
A	99.06%	96.56%	92.66%	69.57%	95.58%	96.44%	77.94%	99.53%	92.50%
B	98.95%	97.01%	93.74%	68.84%	95.23%	96.56%	78.40%	99.55%	93.05%
C	98.79%	97.26%	91.50%	70.56%	95.46%	98.00%	80.59%	99.47%	95.35%
D	98.61%	90.36%	95.83%	85.05%	94.39%	94.88%	88.99%	96.25%	90.82%

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 (A to B) and the rigidity of the proximity constraint (B to C). Model D shows lower coincidence ratios overall than 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 our data shows a strong self-selection of large third-party carriers, though, the estimated model is not representative for any other set of carriers. To be able to specify the population to which our model results apply, we need to know better what the exact bias of the XML microdata is.

While observed distributions are reproduced well, Models A to C overestimate slightly the percentage of tours with three or four stops and underestimate the percentage of tours with six or seven stops (Table 6.3). 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. third shipment, 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 6.4). We expect this to be caused by measurement differences between observed and predicted tour distances. Observed tour distances are filled out 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 get gas, have lunch, and evade a congested AM or PM peak highway; kilometers that our predicted tour distance does not include.

Table 6.3. The observed and predicted distribution of number of stops. Averaged over three models runs for A-C and two model runs for D.

Number of stops	Percentage of tours					
	50% of *days* for estimation				50% of *carriers* for estimation	
	Observed	Predicted (A)	Predicted (B)	Predicted (C)	Observed	Predicted (D)
1-2 (direct)	92.50%	92.51%	92.56%	93.01%	90.76%	89.22%
3	1.97%	2.16%	2.10%	1.93%	3.28%	3.49%
4	1.49%	1.79%	1.72%	1.62%	2.25%	2.76%
5	1.19%	1.23%	1.17%	1.10%	1.43%	1.51%
6	1.11%	0.85%	0.84%	0.80%	0.81%	0.92%
7	0.68%	0.52%	0.55%	0.52%	0.51%	0.54%
8	0.35%	0.32%	0.34%	0.33%	0.28%	0.38%
9	0.20%	0.17%	0.21%	0.20%	0.25%	0.23%
10	0.14%	0.12%	0.13%	0.13%	0.17%	0.18%
11	0.12%	0.08%	0.09%	0.08%	0.13%	0.14%
12	0.09%	0.05%	0.06%	0.06%	0.05%	0.12%
13	0.06%	0.05%	0.05%	0.05%	0.03%	0.08%
14	0.05%	0.03%	0.03%	0.03%	0.02%	0.07%
≥15	0.05%	0.12%	0.13%	0.14%	0.04%	0.36%

Table 6.4. The observed and predicted distribution of tour distance. Averaged over three models runs for A-C and two model runs for D.

Tour distance [km]	Percentage of tours					
	50% of *days* for estimation				50% of *carriers* for estimation	
	Observed	Predicted (A)	Predicted (B)	Predicted (C)	Observed	Predicted (D)
<50	67.78%	72.50%	72.39%	71.90%	49.50%	58.09%
50-100	9.13%	10.07%	10.13%	10.53%	19.75%	16.53%
100-150	8.02%	7.63%	7.69%	7.91%	12.59%	11.41%
150-200	5.07%	4.35%	4.38%	4.50%	7.75%	7.22%
200-250	3.13%	2.57%	2.53%	2.54%	3.79%	2.93%
250-300	2.23%	1.06%	1.07%	1.01%	2.12%	1.44%
300-350	1.56%	0.65%	0.65%	0.60%	1.50%	0.84%
350-400	1.14%	0.41%	0.41%	0.37%	1.15%	0.55%
400-450	0.72%	0.27%	0.28%	0.23%	0.70%	0.33%
450-500	0.43%	0.17%	0.16%	0.15%	0.38%	0.21%
500-550	0.27%	0.12%	0.11%	0.09%	0.25%	0.16%
550-600	0.19%	0.08%	0.08%	0.06%	0.13%	0.11%
600-650	0.11%	0.06%	0.05%	0.04%	0.10%	0.07%
650-700	0.06%	0.04%	0.03%	0.03%	0.05%	0.05%

700-750	0.03%	0.02%	0.02%	0.02%	0.04%	0.03%
750-800	0.02%	0.01%	0.01%	0.01%	0.02%	0.02%
800-850	0.02%	0.01%	0.01%	0.01%	0.02%	0.01%
850-900	0.01%	0.00%	0.00%	0.00%	0.01%	0.00%
900-950	0.01%	0.00%	0.00%	0.00%	0.03%	0.00%
950-1000	0.01%	0.00%	0.00%	0.00%	0.02%	0.00%
≥1000	0.05%	0.00%	0.00%	0.00%	0.10%	0.00%

We should note that the predicted tours are constructed with the same set of shipments as the observed tours. This explains at least partially why observed tour statistics are reproduced so 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 the sensitivity to travel time changes. Four simple scenarios are defined in which all OD pairs experience the same increase or decrease in travel time. In reality, feedback mechanisms (e.g. latent demand) and spatial demand patterns influence which links experience larger increases/decreases, and behavior other than tour formation might be impacted (e.g. vehicle type/shipment size choice). However, large model elaborations are required to consider this accurately and the impacts of our simple scenarios are easier to interpret.

When travel times in the network increase, fewer direct tours (Figure 6.1) and fewer tours with 15+ stops are predicted (Figure 6.2). 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.

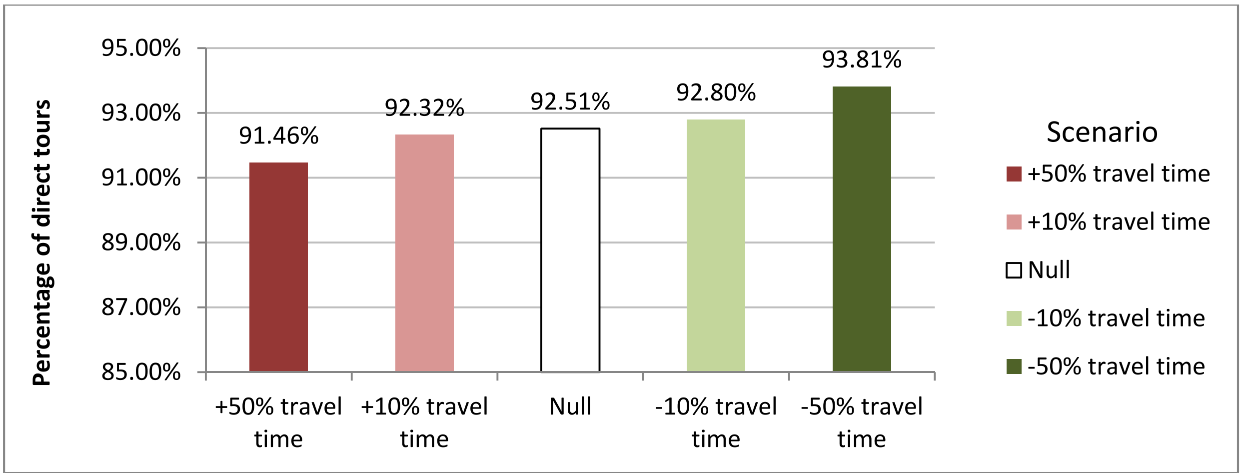

Figure 6.1. The percentage of direct tours under varying travel time scenarios. Averaged over two runs with Model A.

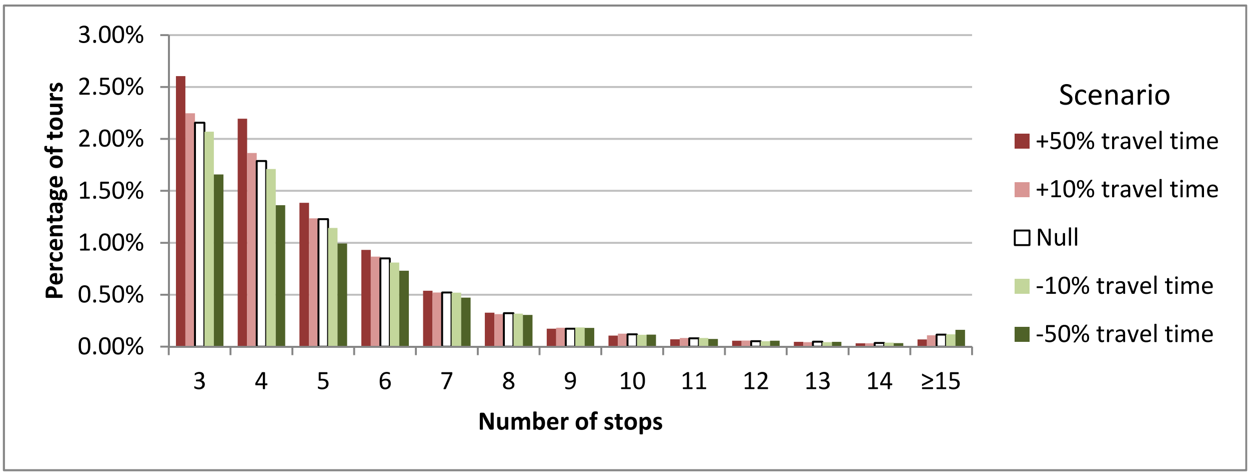

Figure 6.2. The percentage of tours with multiple stops under varying travel time scenarios. Averaged over two runs with Model A.

7. CONCLUSIONS

In this research, we developed a behavioral tour formation model that is shipment-based and statistically calibrated on empirical data. Tours are grown iteratively by allocating one additional shipment until the choice is made to end the tour. Parameters of choice models were estimated using a large and inclusive database that covers road freight transportation performed by third-party carriers in the Netherlands.

Relevant aspects of our tour formation model include, but are not limited to, the following:

- The model is shipment-based; therefore, we can represent the possibilities and constraints that guide the tour formation process more accurately than in tour-based models, we can consider the heterogeneity of goods types (Holguín-Veras et al., 2014), and we can evaluate the impacts of more detailed policies and scenarios (Boerkamps & van Binsbergen, 1999).
- Parameters of the model are statistically calibrated on empirical data that is not limited to a specific segment (e.g. retail restocking tours) but covers all freight demand in the Netherlands.
- Realistic considerations influence the tour formation process. The model acknowledges that carriers construct efficient multiple-stop tours to minimize transportation costs but prefer to construct direct tours to reduce the complexity of planning (Nuzzolo et al., 2012) when efficiency gains are a small. Constraints related to vehicle capacity (in terms of weight) and working shift regulations are respected in the model.
- We can consider that tours visiting distribution centers and urbanized areas are more likely to visit multiple stops, while direct tours are more common when a port transshipment zone is visited.

Validation results showed a highly satisfactory reproduction of observed statistics regarding tour distance and number of stops, with coincidence ratios exceeding 90% when the model is used to construct tours with shipments of other carriers than the carriers that provide data for estimation. Both the model estimates and performance are robust for varying choice set sizes and shipment selection rules. In addition, increases in travel times lead to plausible results; fewer direct tours and fewer tours with more than fifteen stops are predicted, because there is a stronger focus on travel time savings and working shifts are filled with fewer stops.

Consequently, we conclude that this model provides a valid representation of tour formation and can be applied in a shipment-based freight simulation framework. However, four conditions must be met. Firstly, because the Netherlands is a particularly dense and small country, and patterns of freight transportation differ highly between regions (Zhou et al., 2014), application of the model in other countries requires estimation of new parameters and possibly specification of new constraints. Secondly, in a framework covering the Netherlands, the model should be applied only to transportation performed by large third-party carriers, because a self-selection of these carriers has taken place in the dataset. More research is needed about the exact bias of the dataset, or parameters should be estimated with representative data. Thirdly, off-peak skim matrices should be used, as those were used in estimation. Finally, both a synthesized set of shipments assigned to carriers and a vehicle type choice model are needed before tours can be constructed.

Several features that might be added to or improved about the model include the following:

- 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 departure time choice model would allow us to consider that traffic flows and travel times vary throughout the day.
- Estimation of a separate End Tour model for each consecutive shipment in a tour is expected to improve the reproduction of statistics regarding the number of stops.

A more general direction that we recommend for future research relates to the application of a tour formation model in a larger freight simulation framework. As mentioned, we need a vehicle type choice model and a module that generates shipments assigned to carriers. Both are currently under development for the agent-based MASS-GT framework of the agglomeration of Rotterdam, the Netherlands (de Bok et al., 2018). Analysis of the relationships between shipment attributes and the geographical distribution of supply and demand would allow for the synthesis of a realistic set of shipment that leads to accurate traffic predictions. The integration of tour formation and vehicle type choice requires more attention. Our model assumes the vehicle type to be given, but it might be more appropriate to allow feedback of information, e.g. choosing a larger vehicle because capacity is nearly reached. Additionally, a traffic assignment allows us to compare predicted traffic flows with observed traffic counts, which would give much greater insight into the extent to which this tour formation model improves freight traffic predictions.

BIBLIOGRAPHY

Alho, A., Bhavathrathan, B., Stinson, M., Gopalakrishnan, R., Le, D.-T., & Ben-Akiva, M. (2017). A multi-scale agent-based modelling framework for urban freight distribution. *Transportation Research Procedia 27*, 188-196.

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.

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., 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.

ITF-OECD. (2015). *Global trade: international freight transport to quadruple by 2050.* Retrieved from https://www.itf-oecd.org/sites/default/files/docs/2015-01-27-outlook2015.pdf

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

Kim, S., Park, D., Kim, S., & Park, H. (2014). Modeling courier vehicles' travel behavior: case of Seoul, South Korea. *Transportation Research Record (2410)*, 67-75.

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

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.

Quak, H. (2008). *Sustainability of urban freight transport: retail distribution and local regulations in cities (doctoral dissertation).* Retrieved from https://www.researchgate.net/publication/254805169

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.

Wang, Q., & Holguín-Veras, J. (2009). Tour-based entropy maximization formulations of urban freight demand. *Transportation Research Board 88th Annual Meeting.* Washington, DC, USA.

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. (2012). *Methodology for tour-based truck demand modeling (doctoral dissertation).* Retrieved from https://www.researchgate.net/publication/298376390

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.

Zhou, W., Chen, Q., & Lin, J. (2014). Empirical study Of commercial vehicle tour patterns in urban area in Texas. *Transportation Research Board 93rd Annual Meeting.* Washington, DC, USA.