

# Identifying and Modeling Urban Truck Daily Tour-Chaining Patterns

by

Peiyu Jing

B.Eng. in Transportation, Shipping, and Logistics

Shanghai Jiao Tong University, 2016

Submitted to the Department of Civil and Environmental Engineering

in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

© 2018 Massachusetts Institute of Technology. All rights reserved.

Signature of Author: \_\_\_\_\_ **Signature redacted** \_\_\_\_\_  
Department of Civil and Environmental Engineering

May 14, 2018

Certified by: \_\_\_\_\_ **Signature redacted** \_\_\_\_\_  
Moshe E. Ben-Akiva

Edmund K. Turner Professor of Civil and Environmental Engineering

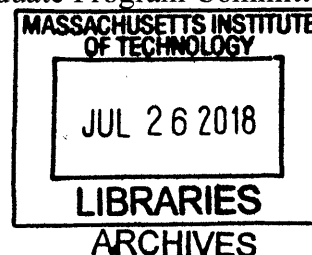
**Signature redacted** Thesis Supervisor

Accepted by: \_\_\_\_\_  
Jesse Kroll

Professor of Civil and Environmental Engineering

Chair, Graduate Program Committee

1





77 Massachusetts Avenue  
Cambridge, MA 02139  
<http://libraries.mit.edu/ask>

## **DISCLAIMER NOTICE**

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

**The images contained in this document are of the best quality available.**



# Identifying and Modeling Urban Truck Daily Tour-Chaining Patterns

by

Peiyu Jing

Submitted to the Department of Civil and Environmental Engineering

on May 14, 2018 in partial fulfillment of the requirements for

the degree of Master of Science in Transportation

## ABSTRACT

The main goal of this research is to better understand truck tour patterns in an urban setting and develop models that can describe daily tour-chaining patterns. This research uses truck activity data collected for the *Urban Freight Heavy Vehicle Study* ongoing in Singapore, which is an advancement in freight data collection studies. The data contain individual truck's Global Positioning System (GPS) traces and rich behavioral details including the activities at stops and operator's characteristics that were processed and verified through a freight data collection platform. Based on the initiative of using post-processed GPS data for tour identification, this paper refines the definition of tour and tour chain to explicitly reflect stop purpose, stop duration, and time of stop. Tour types and daily tour-chaining patterns in the dataset are identified. Further, this paper presents discrete choice models developed to explore factors that influence daily tour-chaining patterns. Identified important factors are: the difference between the number of distinct pickup and delivery locations, geographical spread of distinct pickup and delivery locations, shipment type, time to start work, employment type, land use type, and truck type. The major contributions of the paper are: 1) identifying limitations of the conventional definitions of tour and tour chain and proposing new approaches to reflect logistics practices; 2) explaining the tour-chaining patterns of heavy goods trucks in Singapore; 3) developing tour-chaining pattern choice models that aims serving agent-based simulation platforms.

**Thesis Supervisor: Moshe E. Ben-Akiva**

**Title: Edmund K. Turner Professor of Civil and Environmental Engineering**

## Acknowledgement

First, I want to express my sincerest gratitude to my faculty advisor Prof. Moshe Ben-Akiva and my post-doctoral supervisor Dr. Kyungsoo Jeong and Dr. André Romano Alho. They gave me a lot of inspirations and insights in this research. They directed me to the right track when I was confused with technical details and conceptual designs. I would like to thank them for their expertise, time, and efforts.

Next, I want to thank the members in the freight data collection team, including the students in the US and researchers in Singapore. The data collection studies could not have been possible without joint dedications of the whole team. My sincere thanks go to: Prof. Lynette Cheah, Dr. Fang Zhao, Dr. Linlin You, Dr. Takanori Sakai, Dr. Jing Ding-Mastera, Monique Stinson, Fangping Lu, Ziyue Chen, Surya Ravikumar, William Wong, Eric Manzi, and all the software engineers based in Singapore-MIT Alliance for Research and Technology. In addition, I want to sincerely thank Katherine Rosa and Eunice Kim for your professional assistance in the data collection project and for taking care of all of us.

I also want to thank other members in the Intelligent Transportation Systems Lab at MIT, who provided me with important suggestions in this research. I feel fortunate to have such good friends around me during my experience at MIT. I would like to thank research associates: Dr. Carlos Azevedo, Prof. Bilge Atasoy, Dr. Jimi Oke, Dr. Andrea Araldo, Dr. Arun Akkinpally, and also my fellow students: Yundi Zhang, Xiang Song, Isabel Hemerly Viegas de Lima, Yihang Sui, Ajinkya Gup, Mazen Danaf, Haizheng Zhang, Eytan Gross, and many others. My special thanks go to Yundi Zhang and Xiang Song who gave me a number of suggestions in great details.

Last but not least, I sincerely thank my parents Yijun Jing and Xiuli Qin for their support and love. They have always been my source of confidence and power when I am faced with challenges in life.

## Thesis Errata Sheet

Author Peiyu Jing  
Primary Dept. Civil and Environmental Engineering  
Degree Master of Science in Transportation Graduation date June 8, 2018

### Thesis title

Identifying and Modeling Urban Truck Daily Tour-Chaining Patterns

### Brief description of errata sheet

A paragraph of disclaimer is added in the acknowledgement part.

Number of pages 2 (11 maximum, including this page)

► **Author:** I request that the attached errata sheet be added to my thesis. I have attached two copies prepared as prescribed by the current *Specifications for Thesis Preparation*.

Signature of author Signature redacted Date July 22, 2018

► **Thesis Supervisor or Dept. Chair:** I approve the attached errata sheet and recommend its addition to the student's thesis.

Signature Signature redacted Date Aug 8, 2018

Name Moshe Ben-Akiva  Thesis supervisor  Dept. Chair

► **Vice Chancellor or his/her designee:**

I approve the attached errata sheet and direct the Institute Archives to insert it into all copies of the student's thesis held by the MIT Libraries, both print and electronic.

Signature Signature redacted Date 8/20/18  
Name Ian Waitz

**Add the following paragraph at line 31 of page 4. Line 31 has no text in the original paper.**

This research is supported in part by the Singapore Ministry of National Development and the National Research Foundation, Prime Minister's Office under the Land and Liveability National Innovation Challenge (L2 NIC) Research Programme (L2NIC Award No L2 NICTDF1-2016-1.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of National Development and National Research Foundation, Prime Minister's Office, Singapore. We thank the Urban Redevelopment Authority of Singapore, JTC Corporation and Land Transport Authority of Singapore for their support.

**Revise the sentence at line 19-21 of page 22.**

The original text is: "We did not use data of the first and second batches from the Singapore study mainly for the assurance and consistency of data quality."

Revise the text as: "We did not use data of the first and second batches from the Singapore study mainly for the consistency of data."

# Contents

<b>Acknowledgement .....</b>	<b>4</b>
<b>1 Introduction.....</b>	<b>7</b>
1.1 Overview of freight behavioral modeling.....	7
1.2 Tour-based and tour-chain-based approaches .....	8
1.3 Thesis motivation.....	9
1.4 Thesis outline.....	10
<b>2 Literature review.....</b>	<b>11</b>
2.1 Classification of tour types and tour-chaining patterns .....	11
2.2 Tour type and tour-chaining pattern choice models .....	14
<b>3 Data Descriptions .....</b>	<b>17</b>
3.1 Data collection.....	17
3.2 Data descriptions .....	20
<b>4 Identifying Tour Type and Tour-Chaining Pattern.....</b>	<b>23</b>
4.1 Enhancement to the conventional definition.....	23
4.2 Definition of tour and tour chain .....	24
4.3 Features of tours and tour-chaining patterns.....	29
<b>5 Modeling tour-chaining pattern choice .....</b>	<b>33</b>
5.1 Candidate explanatory variables.....	33
5.2 Model development .....	42
<b>6 Results and Analysis.....</b>	<b>45</b>
6.1 Estimation results.....	45
6.2 Findings.....	48
<b>7 Conclusions and Future Work.....</b>	<b>51</b>
7.1 Conclusions .....	51
7.2 Future work .....	52



## List of Tables

Table 1: Daily tour-chaining patterns corresponding to the number of 4 tour types made in the day .....	27
Table 2: Number of 4 tour types .....	29
Table 3: Number of stops in 4 tour types .....	29
Table 4: Statistical features of tour-chaining patterns .....	30
Table 5: Number of distinct pickup and delivery locations .....	34
Table 6: Dwell time at distinct stop locations .....	35
Table 7: Average spread ratio of pickup and delivery locations .....	36
Table 8: Ratio of choosing an available pattern in the entire dataset .....	37
Table 9: Ratio of choosing an available pattern in the entire dataset .....	38
Table 10: Ratio of choosing an available pattern in the entire dataset .....	39
Table 11: Ratio of choosing an available pattern in the entire dataset .....	40
Table 12: Ratio of choosing an available pattern in the entire dataset .....	41
Table 13: Availability of tour-chaining patterns given the number of distinct pickup and delivery locations .....	42
Table 14: Estimation results of the MNL model .....	45
Table 15: Estimation results of the NL model with a direct nest and a peddling nest	46
Table 16: Estimation results of Estimation results of the CNL model with 1P1D, 1PmD, mP1D nests .....	47

## List of Figures

Figure 1: FMS – Freight architecture .....	19
Figure 2: Web interface of FMS - Freight .....	19
Figure 3: Examples of tour-chaining patterns .....	28
Figure 4: Number of tours in tour-chaining patterns .....	31
Figure 5: Lorry wooden truck .....	41
Figure 6: Recovery vehicle .....	41

# 1 Introduction

## 1.1 Overview of freight behavioral modeling

Freight movements play very important roles in the transportation system. Modeling freight movements is of significance because of its great influence on traffic performances and economic activities. Apart from giving better empirical explanations of travel patterns, understanding behaviors related to freight operations at a disaggregate level is crucial for agent-based simulations, which can be applied to predict freight vehicle and commodity movements under different scenarios such as special events, new public policies, and changes in logistics practices of the freight activity operator.

Freight behavioral modeling distinguishes itself from conventional passenger behavioral modeling due to several well acknowledged reasons (Ogden, 1992; Holguin-Veras and Patil, 2005, Gliebe et al., 2007). First, the revealed freight activity pattern reflects the decision jointly made by multiple agents, such as shippers, carriers, and receivers. Their roles in decision-making are difficult to identify and thus difficult to be incorporated by a systematic approach in modeling. Second, the variations in logistics practices add to complexity in modeling. This challenge is also related to the first point, since different decision-making agents have different objectives and constraints in their practices. For example, shippers are more concerned about the shipping costs if they are the cost-bearers, while the primary concern of receivers (or customers) may well be to minimize delivery delays. There exist trade-offs between the conflicting objectives and constraints, which requires the identification of key measurements contributing to the revealed behavior. Third, commodities and commercial vehicles have large variations in characteristics, so travel patterns of different freight operators are less homogenous than passengers. Fourth, commodity movements and vehicle movements are two separate but related components that

should be studied jointly. A holistic viewpoint is required to model the two components together.

Freight behavioral modeling is faced with challenges due to the lack of high quality data rich in travel behavioral and logistics planning details. Ideal data required to develop freight behavioral models should include commodity movements, freight vehicle movements, logistics practices including decision makers and their roles, and objectives and constraints the different stake-holders have. The advancement in data collection methodology and technology has enabled this research. Utilizing freight data collection studies that provide processed and driver-annotated Global Positioning System (GPS) traces of freight vehicles and multiple characteristics of freight activity operators, we will develop freight behavior models.

## **1.2 Tour-based and tour-chain-based approaches**

In general, freight behavioral modeling can be classified as two directions: commodity-based and freight-vehicle-based. Although as aforementioned these two components should be studied jointly, this research is primarily focusing on the latter one. To study freight vehicle movements, we need to identify the basic unit of analysis. A trip is the movement between two consecutive stops. A tour is a chain of trips beginning and ending at a base location, which can be the home or work location in passenger activity modeling, or a depot or overnight parking location in the freight context. A tour chain is a chain of tours made within certain time frame, which is usually a day for research interest.

Urban commercial vehicles deliver both services and cargoes within a metropolitan area and are characterized by making multiple stops in one tour in daily operations. Tour or trip-chaining is an important feature in urban commercial vehicle movements (Wang and Holguin, 2008). It has been a recent research interest to study the typical types of tours and how tours are chained together in daily operations. The study of

freight vehicle tour-chaining patterns is of significance because it can provide insights into supply chain agent decision-making and behavior (Figillozzi, 2007). It is also a cornerstone in agent-based simulation which can be used to predict freight vehicle movements for scenario analysis. The tour-chaining strategy is at the upper level of many important decisions including next destination choice, route choice, etc., and thus a good prediction of the strategy enhances the overall prediction accuracy. Based on previous work, we want to improve the understanding of urban truck daily tour-chaining patterns.

### **1.3 Thesis motivation**

With an advancement of data collection study (Cheah et al., 2017), most recently Alho et al. (2018a, 2018b) proposed tour and tour-chaining pattern identification approaches and identified tour-chaining behaviors in the dataset. Based on this initiative, this study is motivated by 2 major aspects.

First, the definitions of tour and tour-chain may be further refined so as to better reflect logistics planning of freight vehicle movements. The popularly adopted definition of a tour in freight studies is primarily the same as in passenger studies, yet we question the rationality of such an approach. In addition, instead of containing a large set of possible possibilities, each pattern should be clearly defined and distinguished from each other.

Second, we want to develop the tour-chaining pattern choice models. For the first point, previous literature did not take into consideration the correlations between alternative tour-chaining patterns or concluded that structures that explicitly reflect correlations between alternatives were not advantageous, however we want to test this conclusion under a different definition of tour-chaining. Next, previous work used many observed attributes of the actual choice as explanatory variables for all alternatives. This is only applicable for the purpose of identifying key influencing

factors, but models derived as such cannot well fit into agent-based simulators. For an example, previous literature used travel distance between stops as an explanatory variable. However, the distance is the calculated based on the sequence of stops revealed in the day, which actually relies on the tour-chaining pattern chosen. In that sense, the calculated distance can only be viewed as an attribute of the chosen (observed) pattern, but it cannot be used to describe the attributes of other patterns not chosen. We will only use attributes and characteristics not relying on any observations and find generic attributes that can function as proxies of the observed attributes, so that the model can be used for predictions and incorporated into agent-based simulators. Last, Lin and Zhou (2013) indicated that tour-chaining patterns show large cross-regional variations, so we want to develop a model specifically suitable for the Singapore region which has not been studied before.

#### **1.4 Thesis outline**

The remainder of the thesis is organized as follows. Section 2 summarizes literature on two topics – the classification and modeling of tour type and tour-chaining pattern. Section 3 presents details of the data collection studies and descriptions of the data. Section 4 presents a new approach to define tour types and tour-chaining patterns and shows statistics of identified patterns in the dataset. Section 5 details the development of discrete choice models of tour-chaining patterns. Section 6 presents the model estimation results and discusses the insights derived from the results. Finally, conclusions and future work are summarized in Section 7.

## **2 Literature review**

In this section, we first summarize the common classifications of tour types and tour-chaining patterns of commercial vehicles in urban settings. Then we show literature discussing the advantages of using a tour or a tour chain as a basic unit of freight analysis. Finally, we show various models that investigate the choice of tour types and tour-chaining patterns.

### **2.1 Classification of tour types and tour-chaining patterns**

A trip is the movement between two consecutive stops. A tour is a chain of trips beginning and ending at a base location. A tour chain is a chain of tours made within certain time frame.

Insights for freight activity modeling come from passenger activity modeling, which has been well studied (Miller et al., 2005; Ben-Akiva and Bowman, 1998; Bowman and Ben-Akiva, 2001). Bowman and Ben-Akiva (2001) developed a system consisting of multiple choice models to forecast an individual's daily activities and travel schedules. The activity pattern includes three main components: 1) the primary activity, 2) the type of tour of the primary activity, and 3) the number and purpose of secondary activities. Tour models include the choice of time of day, destination and mode of travel. An individual's daily activities and travels are analyzed on the level of tour and tour decisions are conditioned by the choice of activity pattern. There are a number of other literature suggesting that approaches based on "tours" and "daily activity patterns" are advantageous than the trip-based approaches.

In freight behavioral modeling, many researchers used tour as a basic unit for analyzing urban commercial vehicle movements. (Wisetjindawat et al., 2006, Hunt and Stephan, 2007; Wang and Holguin-Veras, 2008; Ruan et al. 2012; Kim et al. 2014). Hensher and Figliozzi (2007) addressed the importance of studying trip-

chaining behaviors in freight transportation. In urban freight settings, a tour is a more appropriate unit of analysis than a trip, because it considers a series of freight activities or schedules as a whole and thus contains important interactions between trips. In terms of the impact on congestions, choice of tour type has a strong influence on vehicle miles traveled (VMT) (Figillozzi, 2007) so the study on tour type contributes to the prediction of VMT.

To conduct tour-based analysis, the classification of freight vehicle tour types is of importance. In the freight setting, tours are primarily realizations of cargo distribution strategies. Burns et al. (1985) first classified two basic types of cargo distribution strategies of urban commercial vehicles. One strategy is named as “direct”, where the vehicle serves only one customer directly in one vehicle load. The other strategy is “peddling”, where the vehicle ships cargoes to multiple customers per vehicle load before it returns to the base location. The authors found that the peddling strategy is more cost-saving when customer demands are smaller, cargoes are of higher values, and customer locations are concentrated geographically. Battelle Memorial Institute (1995) extended the definition of peddling to include pickup activities at intermediate stops. Liu et al. (2003) studied the features of two urban delivery systems. The first one is “direct shipment”, in which the suppliers ship cargoes to their customer(s) directly. Depending on shipping costs and capacity constraints, one or more customers may be served within one vehicle tour. The second system is “hub-and-spoke system”, in which cargoes from multiple suppliers are consolidated before being distributed to one or more customers. In terms of delivery to customers, the two systems both consist of the direct and the peddling strategy and in addition they highlight the different strategies of picking up cargoes. Chopra (2003) summarized six common types of distribution network designs based on the delivery strategy, the pickup strategy, and the land use type of the base location. The designs are: retail storage with customer pickup, manufacturer storage with direct shipping, manufacturer storage with in-transit merge, manufacturer storage with pickup,

distributor storage with package carrier delivery, and distributor storage with last mile delivery.

Further, Holguin-Veras and Patil (2005) analyzed features of urban commercial vehicle movements and found that about 25% of all commercial vehicles make more than one tour daily, thus the tour-chaining behavior is of interest. A tour-chaining pattern of a freight vehicle can be viewed as the activity pattern of an individual in a day, which is a corner-stone of activity-based modeling. In the study of individual activity patterns, Ben-Akiva and Bowman (1998) pointed out that tour-based modeling fails to consider the temporal and spatial constraints between tours, which is less advantageous than tour-chain-based approaches capturing multi-facet of the decision-making process. In a related work, Bowman (1998) held the similar argument yet admitted the challenges of developing tour-chain-based modeling for a day (or more) due to the huge size of activity schedule alternatives and complexity in factors influencing the decisions.

The first attempt to study tour-chaining behaviors of freight vehicles is by Ruan et al. (2012). They first found that freight tours are logically chained based on shipping cost, customer demand, number of customers, type of service, etc. The rationale is that the objective of freight activity decision-makers is to minimize logistics costs in daily operations, thus the individual tours are considered jointly to complete daily tasks. Tour-chain-based approach is advantageous than tour-based approach because it considers the interconnections among linked tours and thus incorporates multiple logistics considerations in a holistic framework. They first defined 5 types of tour-chaining patterns on the basis of classifications by Burns et al. (1985) and Battelle Memorial Institute (1995). A single direct tour starts at a base location, serves only one customer or intermediate stop per vehicle load, and ends at the base location. A single peddling tour starts at a base location, serves more than one customer stops or intermediate stops per vehicle load, and ends at the base location. The five types of daily tour-chaining patterns are: single direct, single peddling, multiple direct,



multiple peddling, and mixed. If the vehicle only makes one tour in a day, the daily pattern is the same as the tour type (single direct or single peddling). If the vehicle makes multiple single direct tours a day, then the daily pattern is multiple direct. By the same logic, if multiple single peddling tours are made, then the daily pattern is multiple peddling. If single direct and single peddling tours coexist in the day, then the daily pattern is mixed. A tour is defined by a base location but a tour-chain may have many base locations.

## **2.2 Tour type and tour-chaining pattern choice models**

Ruan et al. (2012) developed logit models (a multinomial logit model (MNL), a nested logit model (NL), a mixed model) to study the choice of 5 tour-chaining strategies they defined as aforementioned. The dataset is the Texas Commercial Vehicle Survey in 2005 and 2006 (Nepal et al., 2007a, 2007b, 2007c) which collects plentiful travel details of commercial vehicles in 5 counties of Texas. Although the nesting structure is not specified in the paper, theoretically an NL model partitions the choice set to several nests and assigns correlated alternatives in one nest (Ben-Akiva and Lerman, 1985). A mixed model is a generalization of the MNL (or binary logit) model by means of estimating distributed coefficients, transforming variables to a non-linear form, etc. They concluded that stop purpose (i.e. whether stop is a pickup stop), travel distance, dwell time, cargo type, land use type, industry type, pickup/delivery cargo weight, truck type, and other zonal features influence the choice of tour-chaining patterns. The NL did not show significant structural advantages over the MNL and the mixed model did not yield stable results probably due to the small sample sizes of some patterns for some study areas. Further, they compared the classic tour-based model with tour-chain-based model and concluded that the latter performs better in capturing critical distribution decisions.

Lin and Zhou (2013) estimated a binary logit model for Texas and Idaho respectively to investigate the choice between the direct and the peddling tour strategies of

commercial vehicles. Two datasets from two similarly structured commercial vehicle surveys are used for analysis. One is the Texas Commercial Vehicle Surveys in 2005 and 2006, and the other is Community Planning Association of Southwest Idaho Commercial Vehicle Surveys. They showed that cargo type, travel purpose, travel time, dwell time, and tour destination are factors influencing the choice. They also revealed large cross-regional variations in commercial vehicle movement patterns between Texas and Idaho.

Following previous work, Zhou et al. (2014) developed an MNL model to study commercial vehicle tour type based on direct and peddling strategy with different number of customer stops. The continuous component – number of customer stops is classified as 4 levels and implicitly considered as different alternatives, contributing to 5 types: direct tour, peddling tour with two customer stops, peddling tour with three to five customer stops, and peddling tour with more than five customer stops. The dataset is the Texas Commercial Vehicle Survey in 2005 and 2006. They tested an NL model but did not observe structural advantage. The results show that factors influencing the choice of tour type are: commodity type, land use type, loading/unloading cargo weight, and travel speed.

Khan and Machemehl (2017) adopted the multiple discrete continuous extreme value (MDCEV) model proposed by Bhat (2005, 2008) to jointly model the choice of tour-chaining pattern and the number of trips in a tour chain. They argued that the choice of a tour-chaining pattern is a multiple discrete-continuous choice because besides the discrete component – the choice of tour-chaining pattern, there is a continuous component – the number of trips made in a selected tour pattern. The two components in decision-making are correlated to each other and should thus be studied together. The alternative tour-chaining patterns considered are the same as the 5 patterns in Ruan et al. (2012). The dataset they used is the 2006 Austin Commercial Vehicle Survey. Various explanatory variables were used to explain the choice, including cargo type, land use type, and other shipment and urban characteristics.

To summarize, previous literature identified the most common tour types and tour-chaining patterns of commercial vehicles in urban settings. The previous studies primarily used responses to commercial vehicle surveys in the US and pointed out the need for better quality data with more details. Factors influencing the choice of tour type and tour-chaining patterns were investigated and different model structures were tested.

Most recently, Alho et al. (2018a, 2018b) developed new approaches for identifying tours and daily tour-chaining patterns. One approach defines tours based on the pickup and delivery stops. A tour starts at a pickup stop and ends before another pickup or a long rest stop. They incorporated the direct and peddling strategies at the delivery stops, and also distinguished two strategies – regular and irregular – based on whether the next pickup is at the same location as in the previous tour. Another approach is based on the vehicle load, where the criteria for terminating a tour is based on the available capacity in vehicle. They identified and analyzed the tour-chaining behavior of freight vehicles using data collected in the *Singapore Heavy Vehicle Parking Study* (Cheah et al., 2017).

## 3 Data Descriptions

This research is enabled by data collection studies that provided rich driver-annotated GPS traces and stop activity data as well as operators' characteristics. This section presents the data collection methodology and summary statistics.

### 3.1 Data collection

The identification and prediction of daily tour-chaining patterns of commercial vehicles require disaggregate vehicle movement data of high quality and rich behavioral details. We have been working on the development of state-of-the-art freight data collection methodology and technology. The name of the study is the *Future Freight and Logistics Survey* which targets at collecting freight vehicle and shipment data in the US. Another ongoing study in Singapore is the *Urban Freight and Heavy Vehicle Study* and researchers from two studies collaborate on survey designs. The two studies use very similar survey flow, data collection methods, and questionnaires with minor customizations. For simplicity, the former study is referred to as the US study and the latter is referred to as the Singapore study in this thesis. For more details about the US study, refer to Ding-Mastera et al. (2017) and for details of the Singapore study, refer to Cheah et al. (2017).

The US study consists of 3 Phases. Phase 1 is the freight vehicle survey pilot study which focuses on testing the feasibility of the proposed data collection methodology and technology. The pilot study (Nov 2016 – Jan 2017) has been completed successfully. Truck drivers based in the Greater Boston area were targeted survey respondents. Phase 2 is the ongoing shipment survey pilot study which is also a feasibility testing and is projected to be completed by the end of Sep 2018. Shippers and shipments from establishments having outgoing shipments are targeted in the survey. Shippers are expected to answer questions related the shipments tracked by the provided GPS trackers. Phase 3 is a large-scale integrated survey to be conducted

in 2019. It will integrate freight vehicle survey and shipment survey in a holistic framework and will collect freight data from a large number of freight activity operators.

The Singapore study has the similar design as the freight vehicle study in the US, while only truck drivers with specific characteristics are targeted – those who operate heavy vehicles with maximum laden weight over 5 tons and are assigned with an overnight parking location in government-owned parking lots. 4 batches have been completed until Apr 2018 and the 5<sup>th</sup> batch is ongoing. Survey methods have been improved in each batch from feedback and experience in previous batches.

In both surveys, the uniform data collection platform employed is Future Mobility Sensing – Freight (FMS – Freight) (Cottrill et al., 2013; Ding-Mastera et al., 2017). FMS – Freight utilizes advanced sensing and communication technologies coupled with machine learning algorithms to collect and process actual data on trips and activities. Figure 1 illustrates the architecture of FMS – Freight, consisting of 3 main components:

- 1) *Tracking device*: The platform is compatible with multiple types of tracking devices. The web-based survey utilizes GPS loggers while the app-based survey collects data directly from a smartphone or a tablet. In the US vehicle tracking pilot study and the first four batches of the Singapore study, GPS loggers and web-based survey were utilized for tracking freight vehicles.
- 2) *Backend post-processing*: The collected truck GPS data is sent to the server database in real-time for post-processing. Machine learning algorithms, including map matching and stop detection algorithms, are deployed in the backend. The processed GPS data provides of the routes used, stops made, travel and stop times.
- 3) *Mobile/Web Interface*: The processed data is presented back to the driver in the form of a daily questionnaire for verification. The driver can access the personal webpage on the FMS – Freight website to complete the questionnaire, while in the

app-based survey, the driver can directly finish the daily questionnaire in the mobile application. Figure 2 shows the freight vehicle driver's personal web interface. The left column shows the truck trace of the selected day on a map; the middle column shows stop locations and stop times detected by the stop detection algorithms; the right column are contextual questions specific to a detected stop.

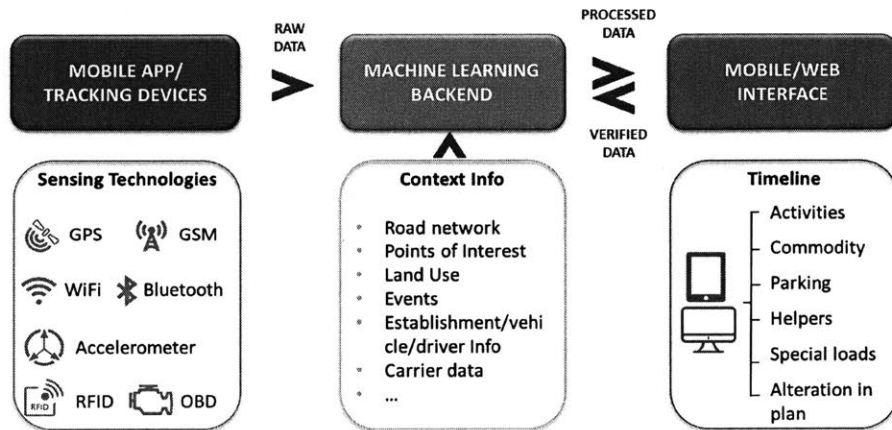


Figure 1: FMS – Freight architecture (Ding-Mastera et al., 2017)



Figure 2: Web interface of FMS - Freight

In both freight vehicle surveys, the survey respondents – freight vehicle drivers could create personal accounts on the FMS – Freight website. After registering basic

information about personal socioeconomic characteristics, the survey is initiated. The driver is asked to conduct the following three main steps, which all can be completed on the personal webpage.

- 1) *Operations questionnaire*: The driver is asked to answer a series of contextual questions regarding employment and operational details of the vehicle. The driver has the option to save frequently visited places and indicate the activities at the place, which is intended to reduce verification burdens in the following step 3.
- 2) *Vehicle questionnaire*: Next, the driver registers the operated freight vehicle by providing a unique vehicle identifier and answering questions regarding characteristics of the vehicle.
- 3) *Tracking and verification*: When the above two questionnaires are completed, the truck is tracked by a GPS logger for a period. At the end of each day, the driver is asked to verify freight activities at detected stops and report special events on the personal webpage.

However, only a low proportion of drivers self-verified daily freight activities on the webpage due to difficulties in accessing and using computers. Surveyors recorded the answers for the daily questionnaire from drivers via telephone interviews. In view of this burden, we are developing and testing the tablet version of FMS – Freight and planning to install tablets on freight vehicles for the following surveys, which will enable the drivers to complete all questionnaires in the tablet application.

### **3.2 Data descriptions**

From the freight vehicle survey, we collected 4 categories of data.

- 1) *Driver characteristics*: In the registration process, age and number of years working as a truck driver are collected. In the Singapore survey, the postal code of residential location is also collected to infer alternative overnight parking locations the driver may consider.

- 2) *Vehicle characteristics*: The vehicle questionnaire collects fuel type, vehicle body type, make and model, length of trailer (if trailer is used), length of the storage area, and maximum laden weight.
- 3) *Operations practices*: The first part collects information related to employment: self-employed or hired, drive for trucking company or non-trucking company, and payment terms. The second part asks about vehicle operational details: decision-maker of route and stops, criteria for route choice, sources of information for route planning, the time stops are determined, commodity types usually carried, industries usually served, special commodity types carried, and whether some costs are paid out of pocket by the driver. The Singapore survey also asks questions related to overnight parking.
- 4) *Annotated GPS data*: The processed GPS trace are presented to the driver for daily verification. The first two questions are purpose of the stop and land use type. Following questions are based on the answer to the purpose of the stop. If the purpose is “pick up cargo” or “drop off cargo”, then cargo types and amount of cargo are asked. If the purpose is “fueling”, then the question is amount of fuel. If the question is “pick up trailer”, then the question is the length of the trailer. In the Singapore study, the percentage of utilized capacity upon arrival is also asked at “pick up cargo” and “drop off cargo” type of stops. In addition, the driver can optionally report special events. Location, time, and type of special events are collected.

In summary, the data collected embrace driver-annotated GPS traces, logistics planning practices, as well as characteristics of the driver, vehicle, and commodities carried. GPS data coupling with driver verifications ensure the quality of data, and the questionnaires can facilitate the explanation of the decision-making process in freight vehicle movements. There is a great potential to develop agent-based and activity-based freight models using the data.



In the Phase 1 – freight vehicle survey pilot in the US study, 28 truck drivers were recruited in the Greater Boston area. They were tracked for a total of 442 days and 18 out of them were tracked for more than 20 days. Among all the drivers, 8 drivers were identified as driving only in urban areas of Boston, 15 drivers only made intercity travels, while the rest 5 made a mixture of urban and intercity travels.

Until Apr 2018, the Singapore study has collected data of 1610 vehicle\*days from 379 unique drivers/vehicles (Alho et al., 2018a). Each driver was tracked for consecutive 5 days from Monday through Friday, yet there exist days with no data observed. All of the truck operations are inside of Singapore (no cross-border travels) and thus are identified as urban trucks.

We used data from the 3<sup>rd</sup> batch (Oct 2017 – Nov 2017) of the Singapore study for the following analysis. The dataset has 7,289 verified stops and 561 vehicle\*days collected from 119 driver/vehicles. Each truck was operated by a single driver. We did not use data from the US study because of the small sample size and the existence of intercity travels, which brings difficulties in identifying daily patterns. It remains a research interest to study the tour and tour-chaining patterns of vehicles that do planning on the basis of multiple days. We did not use data of the first and second batches from the Singapore study mainly for the assurance and consistency of data quality.

## 4 Identifying Tour Type and Tour-Chaining Pattern

This section presents a refinement of definitions of tour type and tour-chaining patterns of freight vehicles. The definition incorporates multiple important factors in logistics practices. Using the proposed definition, we identified the tour type and daily tour-chaining patterns of the trucks in the dataset. The statistics are presented.

### 4.1 Enhancement to the conventional definition

The conventional definition of a tour is a chain of trips beginning and ending at the same base location. A daily tour chain is a chain of tours made in a day. Previous researchers developed this idea from passenger transportation modeling to study freight activities. However, there are several limitations in this definition.

#### 1) *Problems with using a base location as:*

The definition of a base location is vague. Unlike passengers who have very clear and stable base depots (e.g., home, work), freight vehicles may not have a clear base depot. First, the freight vehicle may have multiple frequently visited places which may change over time. These places may be where the commodity suppliers and customers are located, but it is difficult to determine which ones should be considered as the base depots and which ones are only temporary pickup and delivery stops. Second, in the data although the trucks have been allocated overnight parking locations, we observed that some of them are parked somewhere else for several days in a week. This indicates that the overnight parking location, usually viewed as a base depot, may not be frequently visited and thus its importance in logistics planning should not be overly emphasized. Third, if the base location is not defined properly, rich travel details within a tour may be lost. Consider the observation above, if we define the overnight parking location as the base depot, then the travel details are lost before the

truck returns to it. There is a risk of not capturing important decisions made within the tour if we use a base location to define a tour.

2) *Stop purpose needs highlighting:*

The conventional definition does not highlight the importance of pickup and delivery stops. Pickup and delivery stops are treated in the same way as stops of other purposes. A tour may contain multiple pickup stops and delivery stops, yet how they jointly impact the decision of a tour is unknown. However, in logistics planning, the travel pattern of freight vehicles is predominantly determined by the pickup and delivery activities, which can be viewed as the primary purposes of freight vehicle operations and should thus be separated from other types of stops in the analysis.

3) *Time frame of interest:*

Tour-chaining behavior is usually studied on a daily basis. The time frame is a calendar day from 0:00 am to 11:59 pm, yet in the dataset, a lot of freight activities were observed at night. In this sense, if a tour is made across two calendar days, it may be cut into two or partly discarded. Notably, we are limiting our focus on trucks traveling in urban areas in this research. Intercity trucks are significantly different from urban trucks in that their travels may last for several days and even up to a week, but they have less frequent stops. While urban truck movements are usually planned on a daily basis (so we want to investigate daily tour-chaining patterns), intercity truck movements are planned over a longer and varying time span depending on the origin and destination locations.

In view of the potential limitations, we wanted to develop a new approach to define tour and tour chain to reflect the strategic planning process of freight vehicle movements.

## **4.2 Definition of tour and tour chain**

The proposed tour type and tour-chaining pattern identification approach takes into account factors including stop purpose, stop duration, and time of stop, which together determine the boundary of a tour and tour chain.

1) *Stop purpose:*

Truck movements are predominantly affected by the demand and supply of commodities the truck carries. A complete tour must have pickup stops and delivery stops, and may also include intermediate stops for resting, refueling, maintenance, etc. Practically, we think that a consecutive sequence of pickups followed by a consecutive sequence of deliveries is a basic unit of tour planning. We define a tour as a consecutive sequence of pickups followed by a consecutive sequence of deliveries. A tour starts at a pickup location, and when the truck arrives at the next pickup location after delivering cargoes, the tour ends.

2) *Stop duration:*

As mentioned, a tour may have non-pickup and non-delivery intermediate stops. Some of such intermediate stops may not be planned ahead, e.g., a brief stop for restroom, refueling at a gas station found along the way; while others may be planned, e.g., lunch break at a predetermined location. Although driver's plan is unknown, we only considered long intermediate non-pickup and non-delivery stops (longer than the median: 19 minutes) as planned stops and view such a stop as a delivery followed by a pickup in tour pattern identification. Together with pickup and delivery stops, these long intermediate stops form the tours.

3) *Time of stop:*

If there are no pickup or delivery stops before stops verified as "start of my shift" or after "end of my shift", then all stops before "start of my shift" and after "end of my shift" are excluded because we do not want to include travels for personal purposes. By observing the data, we define 4:00 am – 3:59 am as a "day" because freight activities between 4:00 am and 5:00 am are the least frequent. In this way the least number of tours are cut into two. To avoid congestions and high toll rates, many drivers work after midnight, thus cutting a day by 11:59 pm is irrational.

By criteria above, we propose a different definition of tour types:

- 1) *Regular Direct (RD)*: 1 pickup followed by 1 delivery;
- 2) *Irregular Direct (ID)*: multiple consecutive pickups followed by 1 delivery;
- 3) *Regular Peddling (RP)*: 1 pickup followed by multiple consecutive deliveries;
- 4) *Irregular Peddling (IP)*: multiple consecutive pickups followed by multiple consecutive deliveries.

The “regular” type has only one pickup stop in a tour, while the “irregular” type has multiple pickups, which can be explained as collecting and consolidating commodities from different suppliers. For an example, a garbage truck driver may well prefer the irregular type when collecting garbage in residential areas. In terms of delivery stops, the “direct” type has only one delivery, while the “peddling” type has multiple ones, e.g., USPS trucks dropping off packages at multiple locations.

The daily tour-chaining pattern is identified based on the tour types identified in the day (4:00 am – 3:59 am). There is a total of 13 tour-chaining patterns defined by different combinations of tour types. The combinations of tour types and corresponding tour-chaining patterns are summarized in Table 1. Some drivers only make one tour daily, so their tour-chaining pattern is the same as the tour type: RD, RP, ID, or IP. Some drivers make multiple tours of the same type daily, we name the four patterns as mRD, mRP, mID, and mIP. Others mix 3 out of the 4 strategies (regular, irregular, direct, peddling), contributing to four patterns: RDRP, IDIP, RDID, and RPIP. The pattern mixing 4 strategies is named as mixed (MX). Another way to name the patterns is using the #T#P#D notation. #T denotes the number of tours in the day, so 1T denotes making only one tour and mT denotes multiple tours; #P is the number of pickup stops in tours made, so 1P means making 1 pickup in all tours, mP denotes making multiple consecutive pickups in all tours, and uP means that some tours have only 1 pickup while others have multiple ones; #D is the number of delivery stops in tours made, which shares the similar notation logic as #P. Some examples of tour-chaining patterns are shown in Figure 3.

Table 1: Daily tour-chaining patterns corresponding to the number of 4 tour types made in the day

	<b>RD</b>	<b>ID</b>	<b>RP</b>	<b>IP</b>	<b>mRD</b>	<b>mID</b>	<b>mRP</b>	<b>mIP</b>	<b>RDRP</b>	<b>IDIP</b>	<b>RDID</b>	<b>RPIP</b>	<b>MX</b>
<b>#T</b>	1T	1T	1T	1T	mT	mT	mT	mT	mT	mT	mT	mT	mT
<b>#P</b>	1P	mP	1P	mP	1P	mP	1P	mP	1P	mP	uP	uP	uP
<b>#D</b>	1D	1D	mD	mD	1D	1D	mD	mD	uD	uD	1D	mD	uD
<b>#RD</b>	1	0	0	0	>1	0	0	0	$\geq 1$	0	$\geq 1$	0	Other
<b>#ID</b>	0	1	0	0	0	>1	0	0	0	$\geq 1$	0	0	mix
<b>#RP</b>	0	0	1	0	0	0	>1	0	$\geq 1$	0	0	$\geq 1$	of
<b>#IP</b>	0	0	0	1	0	0	0	>1	0	$\geq 1$	$\geq 1$	$\geq 1$	types

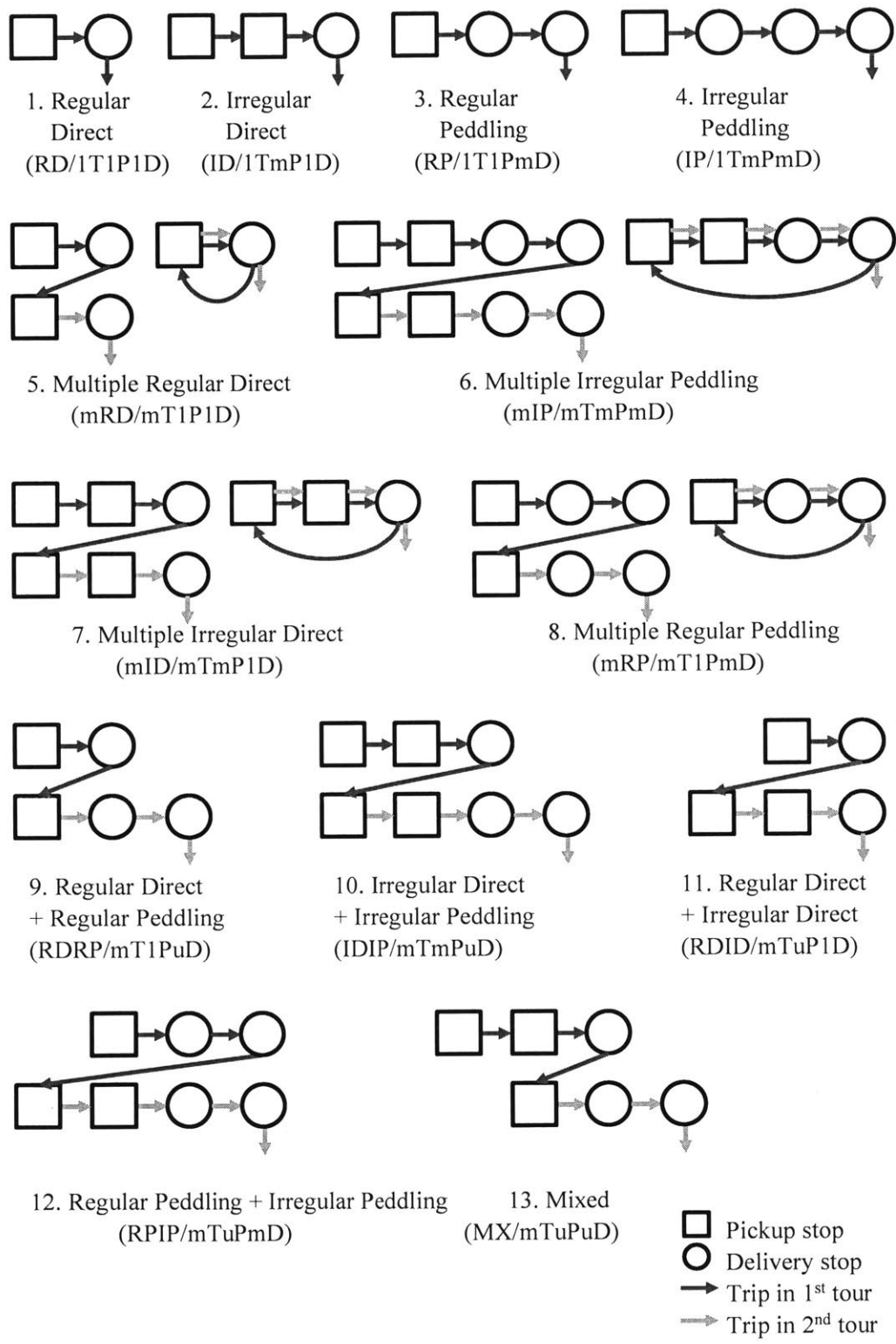


Figure 3: Examples of tour-chaining patterns

### 4.3 Features of tours and tour-chaining patterns

The dataset we used contains a total of 561 vehicle\*days. For the identification of tour and tour-chaining patterns, data processing includes parsing stop activities, converting stop data to daily stop sequences, filtering out days without tours, etc. After processing, a total of 504 vehicle\*days is kept for identifications. The type of each tour is identified and then the pattern for each day is identified. Table 2 shows the number and count of 4 tour types in the dataset. RD has the highest composition (85.7%), showing that a high proportion of tours do not involve consolidation and deconsolidation, but are directly from the supply to the demand. The percentage of ID and IP are both very low, meaning that in this dataset, the trucks do not frequently collect commodities from different suppliers before distributing to customers. 9.2% of the tours are of RP type, the percentage of which is relatively low. In terms of pickup strategies, regular (94.9%) takes the highest percentage, meaning that most trucks deliver commodities directly after pickup; irregular only shares a small portion (5.0%). As for delivery strategies, more direct (89.9%) are observed than peddling (10.1%), meaning that a huge portion of tours serve one customer only. Table 3 shows the number of stops (including pickup, delivery, and other long intermediate stops) in 4 tour types. In ID and RP types, the standard deviation of number of stops is high and minimum is both 3 while maximum is 10 and 11, respectively. The mean number of stops of all tour types is 2.24 and the median is 2, which is of the RD type.

Table 2: Number of 4 tour types

	All	RD	ID	RP	IP
Count	2,392	2,051	101	220	20
Percentage	100%	85.7%	4.2%	9.2%	0.8%

Table 3: Number of stops in 4 tour types

	All	RD	ID	RP	IP
Mean	2.24	2	3.55	3.77	4.25
Std	0.76	0	1.24	1.5	0.72
Median	2	2	3	3	4
Min	2	2	3	3	4
Max	11	2	10	11	7



Then the daily tour-chaining patterns are identified. Table 4 shows some statistical features of the patterns with more than 1 observation. In the dataset, only RD (36), RP (7), IP (1), mRD (250), mID (1), mRP (8), RDRP (103), RDID (51), RPIP (1), and MX (46) are observed (the number in parenthesis is the count), among which RD takes the highest proportion 49.5%. ID, mIP, and IDIP are not observed, which may be due to the small sample size as well as the truck types. All of the trucks are heavy goods vehicles or very heavy goods vehicles and 86.55% of them carry construction materials or minerals. Consistent with observations by surveyors that a great number of trucks transport materials from one side of a construction site to the other, patterns containing RD tour type (1 pickup and 1 delivery), including RD, mRD, RDRP, RDID, MX, make up a large part. Pure peddling (RP and mRP) is not frequent probably also due to the preference for full truckload operations when carrying commodities as construction materials and minerals. Irregular patterns (RDID, RPID, RPIP, IDIP, mID, IP, mIP) are not frequent. This may also be contributed to the monotony of commodity types carried in a day. If a vehicle carries multiple types of commodities, then it is more likely to collect the commodities from different suppliers before visiting customers who have varying needs. Figure 4 shows the average, minimal, maximum, and standard deviation of the number of tours in different tour-chaining patterns. The RD and RP patterns have exactly 1 tour daily while other patterns have at least 2 tours daily and at most 17-20 in mRD, RDRP, RDID, MX. Only 8 mRP patterns are observed, which do not vary significantly in the number of daily tours. As a pattern with 7 different combination of tour types, MX has the highest standard deviation in the number of daily tours.

Table 4: Statistical features of tour-chaining patterns

	All*	RD	RP	mRD	mRP	RDRP	RDID	MX
Count	501	36	7	250	8	103	51	46
Percentage (%)	99.2	7.1	1.4	49.5	1.6	20.4	10.1	9.1
# carry construction materials	232	14	0	113	1	60	22	22
# carry minerals	217	10	0	120	2	40	23	22

\* This does not include the 3 days identified as other patterns.

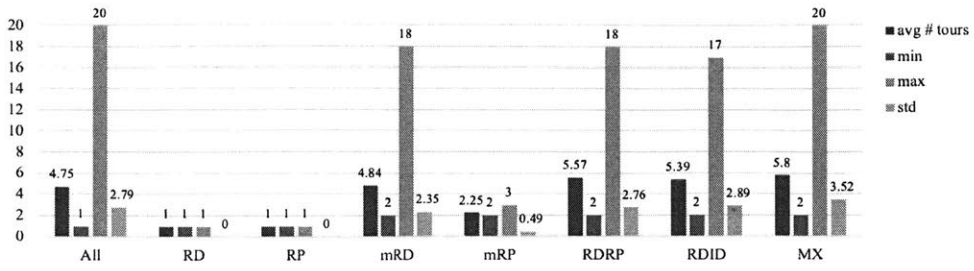


Figure 4: Number of tours in tour-chaining patterns



## 5 Modeling tour-chaining pattern choice

This section discusses the development of tour-chaining pattern choice models. First, we show how we selected candidate explanatory variables and give a priori beliefs of how these variables influence the choice. Next, we show the approaches and procedures in model estimation.

### 5.1 Candidate explanatory variables

To select the most appropriate explanatory variables out of a number of characteristics and attributes available from the dataset, we conducted a prior analysis on the data to examine whether some variables influence the tour-chaining pattern choice in a way consistent with intuitions and as suggested in literature. Based on the data and our a priori beliefs, 8 categories of candidate explanatory variables are specified. As a general guideline, the purpose of modeling tour-chaining patterns is for predicting (rather than only explaining) the choices, thus unlike previous work, we avoid the use of observed attributes belonging to the alternative chosen. Only information from an upper level (e.g., pickup and delivery locations, cargo types) of decision-making that the decision maker has good knowledge of is considered as candidate variables.

#### *1) Number of distinct pickup and delivery locations:*

A distinct stop has a unique location. Candidate variables in this category include the number of distinct pickup locations, the number of distinct delivery locations, the difference and the ratio between the two. The hypothesis is that as the number of distinct pickup locations increases relative to deliveries, irregular (mP) is more preferred than regular (1P); if it is on the contrary, peddling (mD) is more preferred than direct (1D). These preferences are obviously revealed by the definition of different strategies – irregular involves more pickup stops in a tour, while peddling involves more delivery stops than pickup stops in a tour. From Table 5, the data validates our hypotheses. The average numbers of distinct pickup locations of RDID

and MX, both containing the irregular strategy, are both higher than other patterns. The average numbers of distinct delivery locations of RP, mRP, RDRP and MX which contain the peddling strategy are all higher than other patterns. We did not use the average numbers of pickup and delivery stops as explanatory variables because they are affected by the choice of tour-chaining patterns. Yet we put those statistics in the table for comparison purposes. The average number of pickup stops is 5.14 in the entire dataset and RDRP, RDID, and MX have average number of pickup stops higher than 5.14. The average number of delivery stops is 5.47 in the entire dataset and mRP, RDRP, MX have more delivery stops than 5.47.

Table 5: Number of distinct pickup and delivery locations and number of pickup and delivery stops in a day

	All	RD	RP	mRD	mRP	RDRP	RDID	MX
Avg # distinct pickup locations	3.01	1	1	2.83	1.88	3.43	3.94	4.04
Avg # pickup stops	5.14	1	1	4.84	2.25	5.57	7.51	7.52
Avg # distinct delivery locations	3.87	1	4.71	3.39	6.25	5.09	3.76	5.52
Avg # delivery stops	5.47	1	4.71	4.84	7.5	7.51	5.39	7.7

2) *Dwell time:*

Dwell time is the duration at a stop, which is the difference in time between the arrival time and departure time. Dwell time is determined by many factors such as pickup or delivery amounts, cargo type, facility, and truck type. We want to use stop duration as a proxy of pickup or delivery amount, i.e. supply at a pickup location and demand at a delivery location. In the survey, the drivers were asked “How much cargoes did you picked up/deliver here?”, yet many entries are illogical or missing so that the answers are unusable. Instead of absolute values, we used the ratio between the average dwell time at distinct pickup locations and the average dwell time at distinct delivery locations as an approach to avoid its inaccuracy as a proxy of amount of cargo. The variable used is the logarithm of the ratio, which reduces the impact of some overly large values. The hypothesis is that when average dwell time at each distinct pickup location is longer relative to that at each distinct delivery location, peddling (mD) is more preferred than direct (1D) and regular (1P) is more preferred

than irregular (mP). The reason is that when on average the amount of cargo to pick up at one location is bigger than that to deliver at one location, the vehicle is likely collecting cargoes from a big supplier and then distributing to multiple customers. In Table 6, we observe that the average dwell time at each distinct pickup location ( $d_p$ ) is higher than that at each distinct delivery location ( $d_d$ ) for all patterns. This is somehow not validating our a priori hypotheses. One underlying reason may be that dwell time is not a good proxy of amount of pickup and delivery. Another reason is that while we did allocate the dwell time to each stop purpose equally if the driver indicated a multi-purpose stop (e.g., pickup and resting at the same stop), a pickup location is more likely a place where the driver rests than a delivery location but the driver may not have indicated the resting activity at pickup stops properly. The relative magnitude of the dwell time at pickup and delivery stops is  $\log(d_p) - \log(d_d)$ , which is higher than average of the entire dataset (0.70) among only 2 patterns containing the peddling strategy (RP and MX), and higher than 0.70 among 3 patterns containing the regular strategy (mRD, RDID, and MX). Based on a priori data analysis, this variable may not be able to explain the choice very well yet we still tested the validity of this variable.

Table 6: Dwell time at distinct stop locations

	All	RD	RP	mRD	mRP	RDRP	RDID	MX
Avg dwell time at each distinct pickup location ( $d_p$ )	52.86	42.22	108.00	55.42	37.27	46.23	57.18	52.54
Avg dwell time at each distinct delivery location ( $d_d$ )	29.06	39.94	31.37	30.32	24.19	26.95	26.10	23.08
$\log(d_p) - \log(d_d)$	0.70	0.32	1.34	0.76	0.47	0.52	0.93	0.79

### 3) *Spread ratio:*

The relative locations of distinctive pickups and deliveries play important roles in tour-chaining pattern strategies. For example, if the distinctive pickups are very concentrated geographically, then the irregular strategy is preferred given other constraints satisfied. In other words, given the geographical spread, the decision maker chooses a good strategy that saves distance to visit all locations. We define a

unitless factor “spread ratio” measuring how “dispersed” the network of some stops is, relative to the network of all stops. The formula of the spread ratio of pickup locations and delivery locations:

$$pSpreadRatio = \frac{\sqrt{\frac{\sum_{i \in Pickup} l_{ic_p}^2}{\# \text{ pickup locations}}}}{\sqrt{\frac{\sum_{j \in All \text{ Stops}} l_{jc_0}^2}{\# \text{ all pickup and delivery locations}}}} \quad (1)$$

$$dSpreadRatio = \frac{\sqrt{\frac{\sum_{i \in Delivery} l_{ic_d}^2}{\# \text{ delivery locations}}}}{\sqrt{\frac{\sum_{j \in All \text{ Stops}} l_{jc_0}^2}{\# \text{ all pickup and delivery locations}}}} \quad (2)$$

where  $c_p$ ,  $c_d$ ,  $c_0$  is the geographical centroid location of distinct pickup locations, distinct delivery locations, and all distinct pickup and delivery locations, respectively;  $l_{ij}$  is the Euclidean distance between stop location  $i$  and  $j$ .

The distances could be weighted by total supply/demand at each distinct stop to reflect the frequency of visit. The weight of each stop should be the supply or demand, i.e. the summation of all pickup or delivery amounts, but since such data is unusable, the total stop duration at each distinct stop location is used as the weight of the location.  $pSpreadRatioWeighted$  and  $dSpreadRatioWeighted$  are calculated accordingly. A priori belief is as  $pSpreadRatio$  and  $pSpreadRatioWeighted$  are large meaning the pickup locations are very dispersed, regular (1P) is preferred than irregular (mP); as  $dSpreadRatio$  and  $dSpreadRatioWeighted$  is large meaning the delivery locations are very dispersed, direct (1D) is preferred than peddling (mD). Table 7 shows the average weighted and unweighted spread ratio of pickup and delivery locations. The relative magnitudes of spread ratios validate a priori beliefs to some extent. The weighted spread ratios may not be very accurate due to the inaccuracy of dwell time as a proxy of amount of cargo, yet we still tested these variables.

Table 7: Average spread ratio of pickup and delivery locations

	All	RD	RP	mRD	mRP	RDRP	RDID	MX
Avg $pSpreadRatio$	0.41	0.00	0.00	0.44	0.27	0.44	0.53	0.42
Avg $dSpreadRatio$	0.67	0.00	0.41	0.65	1.00	0.77	0.85	0.81
Avg $pSpreadRatioWeighted$	0.37	0.00	0.00	0.40	0.17	0.40	0.51	0.37
Avg $dSpreadRatioWeighted$	0.68	0.00	0.24	0.66	0.98	0.80	0.89	0.79

4) *Commodity carried:*

Only construction materials and minerals take up more than 5% of the total number of commodity types carried, so we only considered the impacts of these two types. The dummy variable for minerals assumes 1 if the day involves at least one stop that supplies/demands minerals. Another dummy variable for construction materials is similarly defined. Hypotheses are vehicles carrying construction materials tend to make RD to transport between construction sites or mRD and RDRP within construction sites; vehicles carrying minerals less prefer irregular (mP) because in delivery locations, it is difficult to pick out different types of minerals from multiple suppliers.

Table 8 shows the potential impact of commodity carried on choosing each tour-chaining pattern. Specifically, the first row is the ratio of choosing a certain pattern given that pattern is available in the entire dataset. The second and third rows are the ratios of choosing a certain pattern given that pattern is available in the subset characterized by carrying minerals and construction materials, respectively. If in a subset, the ratio of choosing a pattern is higher than that in the entire dataset, then it means the characteristic of the subset positively influences the choice of the pattern. We observed that carrying minerals improves the ratios of choosing RD and RDRP, both using the regular (1P) strategy. Carrying construction materials improves the ratios of choosing RD and RDRP as expected, yet not mRD. We still tested the impact of carrying construction materials on choosing mRD.

Table 8: Ratio of choosing an available pattern in the entire dataset compared with that in the subset characterized by commodity carried

	# observations	RD	RP	mRD	mRP	RDRP	RDID	MX
All	501	83.7%	13.0%	49.6%	1.8%	22.9%	12.5%	11.7%
Carry minerals	216	66.7%	0.0%	55.1%	1.0%	20.1%	12.7%	12.4%
Carry construction materials	230	93.3%	0.0%	48.7%	0.5%	28.2%	11.5%	11.8%



5) *Land use type of stops visited:*

The land use types are indicated by the drivers in the daily verification. Some land use types are considered to have an impact on pattern choice: retail store, manufacturing, transfer terminal, and distribution center. As previous literature suggest, land use types are very important factors affecting the tour-chaining strategies. Dummy variables indicating whether the vehicle visits at least one stop of certain land use type are created. First hypothesis is that if in the day the vehicle visits retail stores, peddling (mD) is more likely chosen when the truck distributes commodities from one big supplier to customers with smaller demands, and RD is likely chosen if the supplier only serves a specific customer. For an example, the former is the case when Coca Cola distributes coke to retail stores and the latter is the usual practice when a small company replenishes the only franchisee in the region or one franchisee at one day. If the vehicle visits a distribution center, irregular (mP) and peddling (mD) are more likely chosen because the vehicle is likely collecting cargoes from or distributing cargoes to different distribution centers. If the vehicle visits a transfer terminal, which is the place where cargoes are transshipped from other transportation modes, regular (1P) and direct (1D) is more preferred because the cargo is likely in big amount (e.g. a container) and have a designated customer or a designated delivery location for deconsolidation before distributing to different customers. Table 9 shows how land use type of stops visited affect the choice of tour-chaining patterns, which are in general consistent with a priori beliefs. We observe that the ratio of choosing RD also increases slightly if a distribution center is visited, which is opposed to a priori belief but may reflect other logistics practices.

Table 9: Ratio of choosing an available pattern in the entire dataset compared with that in the subset characterized by land use type of stops visited

	# observations	RD	RP	mRD	mRP	RDRP	RDID	MX
All	501	83.7%	13.0%	49.6%	1.8%	22.9%	12.5%	11.7%
Visit distribution center	81	85.7%	12.5%	45.7%	2.9%	15.7%	22.7%	12.9%
Visit retail store	192	90.0%	27.3%	48.2%	3.4%	25.0%	13.0%	7.8%
Visit transfer terminal	77	100.0%	0.0%	62.3%	1.4%	19.7%	8.6%	2.9%

6) *Time to start work:*

Time to start work is assumed to come from upper level decisions. In the data it is reflected by the start time of the first pickup stop. Two dummy variables are created, one corresponding to start time before 8:00 am, and the other corresponding to start time after 10:00 am. This division is based on a priori analysis of data (49.5% start work before 8:00 am) and also reflects the impact of AM peak. Hypothesis is that if the vehicle starts working relatively late in the day, it is likely to only make 1 tour (RD and RP). The data in Table 10 shows that the ratios of choosing RD and RP increase if the driver starts working after 10:00 am as we expected.

Table 10: Ratio of choosing an available pattern in the entire dataset compared with that in the subset characterized by time to start work

	# observations	RD	RP	mRD	mRP	RDRP	RDID	MX
All	501	83.7%	13.0%	49.6%	1.8%	22.9%	12.5%	11.7%
Start after 10 am	94	85.7%	25.0%	51.6%	0.0%	17.1%	12.1%	9.7%

7) *Employment type:*

Two dimensions of employment type are available from the operations questionnaire: self-employed or hired; drive for a trucking company or a non-trucking company. Two dummy variables are introduced, one denoting whether the driver is self-employed, the other denoting whether the driver drives for a non-trucking company. Hypotheses are self-employed drivers tend to have MX (“Gypsy”) tour pattern; drivers from non-trucking companies more likely to choose regular (1P) than irregular (mP) and more likely to make only one tour. From Table 11, self-employed drivers are more likely to choose MX, RD, and mRP. The patterns using the regular strategy (RD, RP, mRP) are preferred by drivers of non-trucking companies and the ratios of choosing only one tour (RD and RP) increase. The preference for RD among self-employed drivers and drivers working for non-trucking companies may be explained as the driver is a supplier or works for an establishment supplying commodities, and

thus the only pickup location is the establishment. In addition, the demand for such commodity is from a specific customer so that can be accomplished in a tour directly.

Table 11: Ratio of choosing an available pattern in the entire dataset compared with that in the subset characterized by employment type

	# observations	RD	RP	mRD	mRP	RDRP	RDID	MX
All	501	83.7%	13.0%	49.6%	1.8%	22.9%	12.5%	11.7%
Self-employed	39	100.0%	14.3%	41.0%	5.4%	21.6%	10.0%	20.0%
Non-trucking company	203	100.0%	27.3%	48.3%	3.1%	19.4%	13.3%	13.6%

8) *Vehicle type:*

Some types of trucks in the dataset are considered to have an impact on the choice of tour-chaining pattern: garbage sanitary wagon, recovery vehicle, lorry wooden and lorry metal (together considered as lorry truck). It is believed that certain truck types imply the tasks undertaken. Figure 5 is a lorry wooden truck, the structure of which makes it difficult to place scattered or unbundled cargoes. Recovery vehicles (Figure 6) are typically used for transporting vehicles, so when a recovery vehicle tows an illegally-parked passenger vehicle, it is likely to make a RD tour. Hypotheses are garbage sanitary wagons tend to make irregular (mP) and direct (1D) to collect garbage from different stops and then dump at a garbage site; recovery vehicles are more likely to make RD and mRD to carry one passenger vehicle at each tour; lorry trucks are more likely to do RD and mRD. Dummy variables indicating whether the truck is of the corresponding types are created. In Table 12, we observe the impacts of garbage trucks and lorry trucks are generally in agreement with our a priori beliefs. However, garbage trucks also show an increased chance of choosing mRD and lorry trucks do not have a higher percentage in choosing mRD. Recovery vehicles do not influence the choice as we expected probably due to the small number of observations.

Table 12: Ratio of choosing an available pattern in the entire dataset compared with that in the subset characterized by vehicle type

	# observations	RD	RP	mRD	mRP	RDRP	RDID	MX
All	501	83.7%	13.0%	49.6%	1.8%	22.9%	12.5%	11.7%
Garbage Sanitary Wagon	14	NA	NA	57.1%	0.0%	14.3%	28.6%	0.0%
Recovery Vehicle	5	NA	NA	40.0%	0.0%	0.0%	40.0%	20.0%
Lorry Truck	37	90.9%	0.0%	43.2%	0.0%	16.0%	16.7%	13.0%



Figure 5: Lorry wooden truck<sup>1</sup>



Figure 6: Recovery vehicle<sup>2</sup>

<sup>1</sup> <https://www.gumtree.com/p/vehical-recovery-services>

<sup>2</sup> <http://www.foodtruck2u.com/services/truck-body>

## 5.2 Model development

We developed a tour-chaining pattern choice model. 3 patterns having only 1 observation are excluded in model estimation, so the total number of observations used is 501. 7 alternatives consist of the choice set: RD, RP, mRD, mRP, RDRP, RDID, MX. Under the proposed definition of tour-chaining patterns, the availability of an alternative is determined by the number of distinct pickup locations and distinct delivery locations. Availability is specified for each alternative in Table 13. If the day only has one distinct pickup location and one distinct delivery location, then only RD and mRD are feasible choices. If the day only has one distinct pickup location and multiple distinct delivery locations, RP, mRD, mRP, and RDRP are available choices. If the day has multiple distinct pickup locations but only one distinct delivery location, then mRD, RDID are available. Last, if the day has multiple distinct pickup locations and multiple distinct delivery locations, then only RD and RP are unavailable choices.

Table 13: Availability of tour-chaining patterns given the number of distinct pickup and delivery locations

	av_RD	av_RP	av_mRD	av_mRP	av_RDRP	av_RDID	av_MX
1P1D	1	0	1	0	0	0	0
1PmD	0	1	1	1	1	0	0
mP1D	0	0	1	0	0	1	0
mPmD	0	0	1	1	1	1	1

3 types of models are estimated and statistically tested.

A multinomial logit (MNL) model is first estimated. Based on a priori beliefs, we used generic coefficients for alternatives in the same category that are influenced by the variable in the same way. For example, if the variable increases the likelihood to choose peddling, then the alternatives involved with the peddling strategy are all given the same coefficient for this variable, the ones involved with the direct strategy are all given the other coefficient, while other alternative patterns having a mixture of peddling and direct strategies do not include this variable in their utility functions.

The candidate explanatory variables enter the model stepwise. At each step, one variable enters the utility functions, and is removed if it is insignificant at 20% level after estimation. If the introduction of one variable significantly changes the sign or magnitude of the coefficient of any other variable, only one of them is kept in the model based on the F-statistics. At last, the signs and magnitudes of the coefficients are examined and the variable is removed if its coefficient is counterintuitive.

Next, using the utility function specified in MNL model, 3 nested logit (NL) models of different nesting structures are tested. Due to the correlations between different patterns, it is highly likely that the error terms of the alternative are not independently and identically distributed (i.i.d.). The violation of i.i.d. makes NL models more structurally sound. The first model has 2 nests: Single, consisting of RD, RP; Multiple: mRD, mRP, RDRP, RDID, MX. This model intends to test if nesting by the number of daily tours has a structural advantage and if it improves the model estimation results. The second model has 2 nests: Regular, consisting of RD, RP, mRD, mRP, RDRP; Irregular, consisting of RDID, MX. This model assumes all alternatives only having tours with one pickup in each tour to be under the same nest and all other alternatives are under the other nest. The third model also has 2 nests: Direct, consisting of RD, mRD, RDID; Peddling: consisting of RP, mRP, RDRP, MX. The model assumes all alternatives only having tours with one delivery to be in the same nest and others are in the other nest.

Finally, 2 cross nested logit (CNL) models are estimated. In the NL formulation, one alternative can only belong to one nest. However, we could further partition the choice set to 4 nests and assign some alternatives to different nests. This assumes that these alternatives have different degrees of membership to different nests. In this situation, CNL models could be tested (Small, 1987). The correlation between nests in the NL model due to similarities in delivery strategies may make CNL models more favorable. The first model we tested has 4 nests each consisting of several alternatives:

- 1) Direct: RD, mRD, RDRP, RDID, MX;

- 2) Peddling: RP, mRP, RDRP, MX;
- 3) Regular: RD, RP, mRD, mRP, RDRP, RDID, MX;
- 4) Irregular: RDID, MX.

This model tests if nesting by logistics strategy has a structural advantage. RD, mRD, RP and mRP belong to 2 nests; RDRP and RDID belongs to 3 nests; MX belongs to all 4 nests. The second model has 4 different nests:

- 1) 1P1D: RD, mRD, RDRP, RDID, MX
- 2) 1PmD: RP, mRP, RDRP, MX
- 3) mP1D: RDID, MX
- 4) mPmD: MX

We also wanted to test the structure of nesting by tour types that consist of the tour-chaining pattern. RD and mRD only have the 1 pickup plus 1 delivery tour type, so they belong to the 1P1D nest. RP and mRP only have 1 pickup plus multiple delivery tour type, so they are under the 1PmD nest. RDRP belongs to 1P1D and 1PmD nests and RDID belongs to 1P1D and mP1D nests as they both have two different tour types. MX belongs to all 4 nests. Although not all MX patterns consist of 4 tour types, we cannot decompose easily the MX pattern.

## 6 Results and Analysis

### 6.1 Estimation results

The estimation results of the MNL are shown in Table 14. The last row “alternatives associated” lists the alternative that have the corresponding variables and coefficients in the utility functions. The initial log likelihood is -745.140 and the final log likelihood is -597.306. The adjusted rho squared is 0.181. Signs and relative magnitude of the coefficients are consistent with some of the a priori beliefs.

Table 14: Estimation results of the MNL model

Variable	Coefficient	Std	Alternatives associated
<i># distinct P location - # distinct D location</i>	0.39***	0.0934	RDID
	-0.163***	0.0538	RP, mRP, RDRP
<i>pSpreadRatio</i>	1.11***	0.187	RD, RP, mRD, mRP, RDRP
<i>dSpreadRatio</i>	0.576***	0.128	RD, mRD, RDID
	-2.47***	0.387	RP, mRP
<i>carryConstructionMaterials_dummy</i>	1.84*	1.07	RD
<i>carryMineral_dummy</i>	-1.16**	0.584	RD, RP
<i>visitTransferterminal_dummy</i>	1.37***	0.268	RD, mRD
<i>startAfter10_dummy</i>	1.51**	0.699	RD, RP
<i>selfEmployed_dummy</i>	0.617^	0.477	MX
<i>driveForNonTruckingCompany_dummy</i>	1.5***	0.574	RD, mRD, RDID
<i>GarbageSanitaryWagon_dummy</i>	1.18^	0.789	RD, mRD, RDID
<i>LorryVehicle_dummy</i>	1.59^	1.09	RD
# Observations	501		
Initial log likelihood	-745.14		
Final log likelihood	-597.306		
Adjusted rho-square	0.181		

^Significant at 20% \*Significant at 10% \*\* Significant at 5% \*\*\*Significant at 1%

The NL models do not show very good results because in three nesting structures, the scaling factor of nests are very high, meaning that the correlations of alternatives



across nests is higher than that within the nest. The first nest structure consisting of a Single nest and a Multiple nest is structurally invalid because the structural parameters of both nests are estimated to be below 1, meaning that the correlations of utilities within a nest is lower than that between nests. Due to similar reasons, the structure with a Regular nest and an Irregular nest is invalid as the structural parameter of the Regular nest is estimated to be below 1. The NL model with a Direct and a Peddling nest shows a structural advantage over NL model based on the likelihood ratio test result. The structural parameter of the direct nest is significantly different from 1 at 1% level, while that of the peddling nest is higher than 1 but not significant. All variables are significant at 20% level and the signs are consistent with a priori beliefs. The estimation results of this NL model are shown in Table 15.

Table 15: Estimation results of the NL model with a direct nest and a peddling nest

Variable	Coefficient	Std	Alternatives associated
<i># distinct P location - # distinct D location</i>	0.345**	0.151	RDID
	-0.0973*	0.0395	RP, mRP, RDRP
<i>pSpreadRatio</i>	0.644***	0.212	RD, RP, mRD, mRP, RDRP
<i>dSpreadRatio</i>	0.276^	0.186	RD, mRD, RDID
	-1.16***	0.353	RP, mRP
<i>carryConstructionMaterials_dummy</i>	1.74^	1.25	RD
<i>carryMineral_dummy</i>	-0.906*	0.5	RD, RP
<i>visitTransferterminal_dummy</i>	1.25***	0.353	RD, mRD
<i>startAfter10_dummy</i>	1.05*	0.556	RD, RP
<i>selfEmployed_dummy</i>	0.506*	0.287	MX
<i>driveForNonTruckingCompany_dummy</i>	0.961***	0.382	RD, mRD, RDID
<i>GarbageSanitaryWagon_dummy</i>	1.07^	0.799	RD, mRD, RDID
<i>LorryVehicle_dummy</i>	1.51^	1.2	RD
<i>MU_peddling</i>	1.13	0.47	
<i>MU_direct</i>	2.15***	0.561	
# Observations	501		
Initial log likelihood	-745.14		
Final log likelihood	-586.665		

Adjusted rho-square

0.193

^Significant at 20% \*Significant at 10% \*\* Significant at 5% \*\*\*Significant at 1%

*MU\_peddling* and *MU\_direct* are tested against 1 and others are tested against 0

CNL models are more appropriate to address the correlation within a nest. Based on likelihood ratio test, the CNL models show that the first nesting structure with Regular, Irregular, Direct, and Peddling nests is less preferred than the MNL model. The estimation results of the second CNL model with 1P1D, 1PmD, mP1D, and mPmD nests show that the scale parameter of the mPmD nest is smaller than 1. Since in the data, only 0.8% of the tours are of IP (mPmD) type and only the MX pattern contains IP type of tours, we removed the mPmD nest and thus MX belongs to all the 3 nests. The second nesting structure is preferred than MNL based on likelihood ratio test; but *MU\_1P1D* is unbounded from above so that the model is unidentifiable, thus the structure of the model is not well validated under this specification. This may be caused by the dominantly high ratio of RD type of tours. The estimation results of the CNL model with 1P1D, 1PmD and mP1D nests are shown in Table 16.

Table 16: Estimation results of Estimation results of the CNL model with 1P1D, 1PmD, mP1D nests

Variable	Coefficient	Std	Alternatives associated
# distinct P location - # distinct D location	0.218***	0.0647	RDID
	-0.193***	0.0542	RP, mRP, RDRP RD, RP, mRD, mRP,
<i>pSpreadRatio</i>	0.428**	0.22	RDRP
<i>dSpreadRatio</i>	0.523***	0.14	RD, mRD, RDID
	-0.0483***	0.008	RP, mRP
<i>carryConstructionMaterials_dummy</i>	0.308	0.327	RD
<i>carryMineral_dummy</i>	-0.0889^	0.0568	RD, RP
<i>visitTransferterminal_dummy</i>	1.06***	0.27	RD, mRD
<i>startAfter10_dummy</i>	0.088^	0.0611	RD, RP
<i>selfEmployed_dummy</i>	0.447^	0.36	MX
<i>driveForNonTruckingCompany_dummy</i>	0.0518***	0.0184	RD, mRD, RDID
<i>GarbageSanitaryWagon_dummy</i>	0.977^	0.71	RD, mRD, RDID
<i>LorryVehicle_dummy</i>	0.269	0.295	RD

<i>MU_IPID</i>	7.39	6.27
<i>MU_IPmD</i>	50	1.80e+308
<i>MU_mPID</i>	1.79***	0.598
<i>Alpha_RDID_IPID</i>	0.441***	0.107
<i>Alpha_RDRP_IPID</i>	1.00E-05	1.80e+308
<i>Alpha_MX_IPID</i>	0.273	1.80e+308
<i>Alpha_MX_IPmD</i>	0.000307	1.80e+308
# Observations	501	
Initial log likelihood	-736.672	
Final log likelihood	-563.785	
Adjusted rho-square	0.208	

^Significant at 20% \*Significant at 10% \*\* Significant at 5% \*\*\*Significant at 1%  
*MU\_IPID*, *MU\_IPmD*, and *MU\_mPID* are tested against 1 and others are tested against 0

## 6.2 Findings

Since the data size is limited and observations of some alternatives are very low, the results and analysis may not be completely compelling. A more robust analysis relies on the availability of a larger dataset. Major findings are summarized as below.

- 1) The difference between the number of distinct pickup and delivery locations is an important factor influencing the choice of tour-chaining patterns. When the number of distinct pickup locations is higher than that of delivery locations, RDID that involves multiple pickups in a tour is preferred; while RP, mRP and RDRP that involve multiple deliveries with one pickup in a tour are not preferred. This validates our a priori beliefs.
- 2) The spread ratio has significant impacts, indicating that it is a good measurement. If pickup locations are very dispersed, drivers prefer to choose patterns that have only 1 pickup in each tour, contributing to an increased preference for RD, RP, mRD, mRP and RDRP patterns. If distinct delivery locations are very dispersed, it positively impacts the choice of patterns having only 1 delivery in each tour (RD,

mRD, RDID), yet negatively impacts patterns with multiple deliveries in each tour (RP and mRP). If the variables are weighted by the distances by dwell time (a proxy of amount of pickup and delivery), then their impacts are not significant.

- 3) Types of commodities carried are important. Vehicles carrying construction materials are more likely to choose RD, yet there is no significant impact on mRD. This may be explained as there are correlations between other variables in the utility function and the dummy variable of carrying construction materials. Vehicles carrying minerals are more likely to choose patterns with multiple tours, but there are no strong negative impacts on patterns using irregular strategy. We do not observe any patterns involving only the irregular strategy, thus the impact of carrying minerals cannot be well reflected on patterns mixing the irregular with regular strategies.
- 4) If a transfer terminal is visited, the vehicle is more likely to choose RD and mRD, both consist of tours of 1 pickup plus 1 delivery. Visiting distribution center and retail stores do not have the expected impact on the choice.
- 5) Drivers who start working after 10 am are more likely to make only 1 tour in a day, thus showing preferences for patterns of RD or RP.
- 6) Self-employed drivers have higher chances to choose the MX pattern as expected; drivers driving for non-trucking companies prefer RD, mRD and RDID yet do not prefer those involved with peddling.
- 7) Drivers driving lorry trucks (lorry wooden or lorry metal) prefer the RD patterns, yet do not show strong preference for mRD, which again may be caused by correlations between this truck type and other variables in the utility function. Drivers driving garbage sanitary wagons tend to choose RD, mRD, RDID, which may reflect different work patterns – the RDID pattern is suitable for collection of garbage at residential locations while RD and mRD are likely involved with dumping a large amount of garbage from one site to a designated landfill site.
- 8) The variables in the dwell time category are excluded because the estimation results are count-intuitive, which may indicate that dwell time cannot be a good

proxy of pickup and delivery amount since it contains too much information of other types.

## 7 Conclusions and Future Work

### 7.1 Conclusions

This thesis presents a holistic method of identifying daily tour-chaining patterns and modeling the choice of tour-chaining pattern strategies of urban trucks.

We used data collected from *the Singapore Heavy Vehicle Survey Study* (Cheah et al., 2017), which is an advancement in freight data collection methodology and technology. The study collected processed GPS data which was later annotated by drivers as well as plentiful operations details and truck characteristics. A total of 504 vehicle days from 119 drivers are extracted after filtering, which is used for the daily tour-chaining pattern analysis. The ongoing tour and tour-chaining pattern identification studies by Alho et al. (2018b) gave us insights on how to use driver-annotated GPS data for the development of identification algorithms.

Taking into account important factors in logistics planning including stop purpose, stop time, and time of stop, we defined 4 types of tours: regular direct (RD), irregular direct (ID), regular peddling (RP), irregular peddling (IP). Specifically, RD tour consists of 1 pickup and 1 delivery; ID consists of multiple pickups and 1 delivery; RP consists of 1 pickup and multiple deliveries; and IP consists of multiple pickups and multiple deliveries. Only pickup stops, delivery stops, and long intermediate stops are considered in the pattern identification process. A day is defined as 4:00 am – 3:59 am to minimize splitting tours made across 2 calendar days. In the dataset, regular direct tours have the highest share while irregular peddling tours only have a very low proportion.

Based on the types of tours made in a day, 13 types of daily tour-chaining patterns are defined. If the day consist of exactly 1 tour, then the tour-chaining pattern is the same as the tour type, thus the pattern is RD, RP, ID, or IP. If the driver makes multiple

tours of the same type, then the pattern is mRD, mRP, mID, or mIP depending the type of tour made. If 3 out of the 4 types of strategies (regular, irregular, direct, peddling) are used, the day pattern is RDRP, RDID, IDIP, or RPIP. The pattern mixing regular/irregular and direct/peddling is mixed (MX). In the dataset, only RD, RP, mRD, mRP, RDRP, RDID, and MX has more than 1 observations and IP, mID, RPIP only has 1 observation, respectively. mRD has the highest share. The biasness in the share of patterns may be due to the monotony of truck and commodity types.

We developed discrete choice models to understand influencing factors in the choice of daily tour-chaining patterns. Such models are aimed for to be compatible with agent-based simulations. However, as a preliminary work there is a large room for improvement in the consideration of variables. The selection of candidate explanatory variables is based on a priori data analysis and a priori beliefs from previous literature and also intuition. Important factors identified include: number of distinct pickup locations and distinct delivery locations, geographical spread of pickup and delivery locations – the spread ratio, shipment type, time to start work, employment type, land use type of stops visited, and truck type. An MNL model shows good explanatory power whereas different structures of nested logit and cross nested logit models could not be validated based on statistical tests. This research is limited by the small dataset and the biased vehicle composition as all vehicles are heavy goods vehicles, yet it is an elementary step to a better understanding of the key strategy of tour-chaining behaviors in logistics planning.

## **7.2 Future work**

### *1) Integration with agent-based simulation:*

Tour-chaining pattern strategies is an important part in agent-based freight modeling. To feed into an agent-based simulator, this model should fit in the planning stage before the day starts. The introduction of discrete choice models will incorporate more behavioral considerations than optimization-based models.

Inputs required are explanatory variables specified in the model, mainly consisting of shipment sizes, shipment types, pickup and delivery tasks, land use types, truck types. These inputs are generated from the long-term planning.

Outputs of the model is the choice of a daily tour-chaining pattern strategy for each agent. For each pattern chosen, case-by-case analysis should be conducted. Some additional models will need to be estimated in order to realize the model outputs as traffic flows on the network. Such models may include: stop clustering, stop sequencing, and route choice.

*2) Model development and validation using other datasets:*

Other types of discrete choice models may be estimated and tested. Error components may be introduced to capture the correlation between alternatives, which has a similar function as the NL and CNL model. Some coefficients may be defined as randomly distributed. A candidate is the coefficient of the spread ratio so as to reflect the different tastes for travel distances. Another possible improvement is to take into account the panel effect, i.e. the intrinsic correlations between a sequence of choices one driver makes since each driver was tracked for 5 days.

Another direction for refinement is modeling the number of trips and tour-chaining patterns jointly. The MDCEV model Khan and Mechemehl (2017) specified is a good reference. The author wants to further analyze the rationality of applying the MDCEV model in this context, as well as better defining the discrete and continuous components in the choice set.

The validity and robustness of this model is limited due to the small and biased dataset. 6 tour-chaining patterns have 1 or 0 observation, making it impossible to estimate their utilities. In addition, light goods vehicles are expected to have very different tour-chaining patterns. For intercity trucks that travel for multiple days, their



strategies should not be studied on a daily basis. The identification of their “strategy planning unit” will bring insights to many prevalent models using a day as a unit of analysis. We will further develop the models after the large-scale integrated survey in the US projected in 2019.

*3) Improving data collection methodology:*

Data quality needs to be further improved by the advancement of the data collection methodology. For the US study in 2019, we suggest to make some changes in the survey design. One suggestion is that drivers should be tracked consecutively as in the Singapore study. Another suggestion is to impose logic checks on entries from users. In addition, the shippers should be surveyed in order to get more information about constraints in logistics planning.

*4) Incorporating logistics constraints in modeling:*

Constraints play very important roles in decision-making in logistics practices. Such constraints include time window, road restrictions on trucks, driver’s longest working hour, etc. We are interested in developing a holistic modeling system to address these constraints and incorporate them in the choice modeling.

*5) A more comprehensive consideration of explanatory variables and modeling structure:*

As a preliminary work, this thesis is limited in the considerations of explanatory variables. Specifically, variables describing the network properties, such as travel time dynamics, distance, time-varying toll rates, and parking supplies, are not considered. The challenges in using these variables come from the difficulties of deriving them from the daily tour-chaining patterns since decisions at lower levels also have impacts on the final realization of the travels. As a future interest, we would like to extend the current modeling structure to jointly model these choices, and thus test how other important variables have an influence on travel patterns of freight vehicles. In addition, we will consider what optimization objectives should be



incorporated in a way more valid theoretically. These improvements are meant to make the model more practical and more applicable for agent-based simulators.



## Reference

1. Alho, A., Sakai, T., Jeong, K., B. K., Bhavathrathan, and M. Ben-Akiva, 2018a. Revealing freight vehicle tours and tour patterns from GPS vehicle tracking and Driver Survey data. Accepted for presentation at the 7th Innovations in Transport Modelling Conference, Atlanta, U.S.A.
2. Alho, A., You, L., Lu, F., Zhao, F., Cheah, L. and M. Ben-Akiva, 2018b. Next-generation freight vehicle surveys by supplementing truck GPS with a driver activity survey: account and insights. Submitted to the 2018 IEEE Intelligent Transportation Systems Society (ITSS) Conference, Maui, Hawaii, United States.
3. Bhat, C.R., 2005. A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions. *Transport. Res. Part B* 39 (8), 679–707.
4. Bhat, C.R., 2008. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transport. Res. Part B* 42 (3), 274–303.
5. Burns, L., Hall, R., Blumenfeld, D., Daganzo, C., 1985. Distribution strategies that 7 minimize transportation and inventory costs. *Oper. Res.* 33 (3), 469–490.
6. Ben-Akiva, M.E., Bowman, J.L., 1998. Integration of an activity-based model system and a residential location model. *Urban Studies* 35 (7), 1231–1253.
7. Ben-Akiva, M.E., Lerman, S.R. and Lerman, S.R., 1985. *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.
8. Bowman, J.L., 1998. *The day activity schedule approach to travel demand analysis*. Ph.D. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
9. Bowman, J.L., Ben-Akiva, M.E., 2001. Activity-based disaggregate travel demand model system with daily activity schedules. *Transportation Research Part A: Policy and Practice* 35 (1), 1–28.

10. Battelle Memorial Institute, 1995. Comprehensive Truck Size and Weight (TS & W) Study Phase 1 – Synthesis: Logistics and Truck Size and Weight Regulations Working Paper 8. Federal Highway Administration.
11. Cheah, L. et al (2017). Next-generation commodity flow survey: a pilot in Singapore, 10th International Conference on City Logistics, 14-16 June 2017, Phuket, Thailand.
12. Chopra, S., 2003. Designing the distribution network in a supply chain. *Transportation Research Part E: Logistics and Transportation Review* 39 (2), 123–140.
13. Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M. and Zegras, P., 2013. Future mobility survey: Experience in developing a smartphone-based travel survey in singapore. *Transportation Research Record: Journal of the Transportation Research Board*, (2354), pp.59-67.
14. Ding-Mastera, J., Jing, P., Stinson, M., Manzi, E., Zhao, F., Marzano, V., Cheah, L. and Ben-Akiva, M., 2018. Combining GPS Data Collection with Assisted Machine Learning to Enhance Freight Vehicle and Driver Surveys: Methodology and Demonstration. In: *Transportation Research Board 97th Annual Meeting* (No. 18-00512).
15. Figliozzi, M.A., 2007. Analysis of the efficiency of urban commercial vehicle tours: data collection, methodology, and policy implications. *Transport. Res. Part B* 4, 1014–1032.
16. Gliebe, J., Cohen, O., Hunt, J.D., 2007. Dynamic choice model of urban commercial activity patterns of vehicles and people. *Transport. Res. Rec.* 2003, 17–26.
17. Hensher, D., Figliozzi, M.A., 2007. Behavioral insights into the modelling of freight transportation and distribution systems. *Transportation Research Part B* 41 (9), 921–923.
18. Holguin-Veras, J., Patil, G.R., 2005. Observed trip chain behavior of commercial vehicles. *Transport. Res. Rec.: J. Transport. Res. Board* 1906 (1), 74–80.

19. Hunt, J.D., Stefan, K.J., 2007. Tour-based microsimulation of urban commercial movements. *Transportation Research Part B: Methodological* 41 (9), 981– 1031.
20. Khan, M. and R. Machemehl., 2017. Analyzing tour chaining patterns of urban commercial vehicles, In *Transportation Research Part A: Policy and Practice*, Volume 102, Pages 84-97, ISSN 0965-8564
21. Kim, S., Park, D., Kim, S., Park, H., 2014. Modeling courier vehicles' travel behavior. *Transport. Res. Rec.: J. Transport. Res. Board* 2410 (1), 67–75.
22. Lin, J.J., Zhou, W., 2013. How urban commercial vehicle trips are chained: case studies in Texas and Idaho. In: *Transportation Research Board 92nd Annual Meeting* (No. 13-3121).
23. Liu, J., Li, C., Chan, C., 2003. Mixed truck delivery systems with both hub-and-spoke and direct shipment. *Transportation Research Part E: Logistics and Transportation Review* 39 (4), 325–339.
24. Nepal, S., Farnsworth, S., Pearson, D., 2007a. 2005 Amarillo Area Commercial Vehicle Survey Technical Summary. Texas Transportation Institute, Report prepared for Texas Department of Transportation Travel Survey Program.
25. Nepal, S., Farnsworth, S., Pearson, D., 2007b. 2006 Austin Area Commercial Vehicle Survey Technical Summary. Texas Transportation Institute, Report prepared for Texas Department of Transportation Travel Survey Program.
26. Nepal, S., Farnsworth, S., Pearson, D., 2007c. 2006 San Antonio Area Commercial Vehicle Survey Technical Summary. Texas Transportation Institute. Texas Transportation Institute, Report prepared for Texas Department of Transportation Travel Survey Program.
27. Miller, E., Roorda, M.J., Carrasco, J.A., 2005. A tour-based model of travel mode choice. *Transportation* 32 (4), 399–422.
28. Ogden, K.W., 1992. *Urban Goods Movement, a Guide to Policy and Planning*. Ashgate, England.
29. Ruan, M., Lin, J.J., Kawamura, K., 2012. Modeling urban commercial vehicle daily tour chaining. *Transport. Res. Part E: Logist. Transport. Rev.* 48 (6), 1169–1184.

30. Zhou, W., Chen, Q., Lin, J.J., 2014. Empirical study of urban commercial vehicle tour patterns in Texas. In: Transportation Research Board 93rd Annual Meeting (No. 14-5132).
31. Wang, Q., Holguín-Veras, J., 2008. Investigation of attributes determining trip chaining behavior in hybrid microsimulation urban freight models. *Transp. Res. Rec.* 2066, 1–8.
32. Wisetjindawat, W., Sano, K., Matsumoto, S., 2006. Commodity distribution model incorporating spatial interactions for urban freight movement. *Transportation Research Record: Journal of Transportation Research Board* 1966, 41–50.