BAA DTFH61-10-R-00013

Tour-based and Supply Chain Modeling for Freight in Chicago

Tour-based and Supply Chain Freight Forecasting Framework

Final Report

April 2012

Resource Systems Group
with
University of Illinois at Chicago
John Bowman

DATA ■ ANALYSIS ■ SOLUTIONS

Table of Contents

LIST OF FIGURES

List of Tables

1.0 INTRODUCTION

In recent years, freight forecasting has been identified as a way to understand the patterns of interstate and international trade, economic growth, and the impacts created by use of the nation's transportation system for the movement of freight. These impacts can include congestion and delay, potential exposure to hazardous materials and other safety concerns, as well as energy use and environmental consequences. The fact that today more and more freight is moved by heavy trucks on the nation's interstate system has become an area of particular concern to planners.

Despite recent advances in freight forecasting, the current methods are not adequate to address the increasingly complex issues related to freight demand. Most existing models are based on methods that were developed for personal passenger travel; this is true for Chicago where truck travel is loosely based on non-home-based passenger travel. Freight is clearly different than personal vehicle travel and requires a different technical approach. Given the transition that is currently underway to implement disaggregate modeling techniques, it is logical to also begin investigating disaggregate techniques for modeling the movement of freight.

Tour-based and econometric methods have become the state-of-the-art approaches in travel demand modeling. These new approaches offer a myriad of benefits that include the ability to model various aspects of choice behavior explicitly. These factors are relevant in personal travel, but are also important in freight modeling. The research described in this report implemented existing research on tour-based and supply chain models for urban commercial movements and provides a glimpse into the future of freight demand forecasting for practitioners. The research has demonstrated the potential of these new methods and how they might serve to address the limitations of current freight demand forecasting models.

1.1 Goals

The overall goal of this research was to identify a framework that can be adopted by Metropolitan Planning Organizations (MPOs) in the U.S. for use in evaluating transportation investments and their impacts on freight mobility. The details of the framework are specifically designed to address current weaknesses identified in standard practice freight forecasting, as follows:

- **The lack of detail at the traffic analysis zone level.** The framework is designed to synthesize firms and micro-simulate goods movements at the zone level rather than relying on available national data on commodity flow produced at a district level.
- **The lack of information about the local pickup and delivery trips.** The framework is designed to specifically model the delivery system at the end of the supply chain. This delivery system in many industries is based on a series of deliveries by a truck before returning to home base for additional goods. This delivery system is represented by tours that each truck makes to pick up and drop off goods.
- **The need to estimate shifts in long-haul and short-haul demand resulting from regional investments.** Current practice in freight forecasting uses commodity flow tables derived from national sources to represent long-haul goods movements and uses 4-step planning models to represent short-haul movements. Some freight models have estimated models that predict long-haul commodity flows. The framework is designed to represent the full supply chain for a specific commodity shipped from the supplier to the customer, including both the long-haul and short-haul components of the goods movement in a single framework rather than modeling these separately.

- **The ability to capture trip-chaining that occurs.** Trip-chaining is an important component of freight movement and commercial vehicle delivery of services. Trip-chaining can occur along the supply chain (goods travel from the supplier through a distribution center or warehouse to the retailer) or can occur during delivery (goods travel from the retailer to each customer along a tour). Trip-chaining is also expected for many service trips (service providers' travel from home base to each customer before returning to home base at the end of the day).
- **The need to represent commodities produced and consumed by different industries.** Commodities will travel differently based on their production and consumption characteristics. For example, the construction industry will consume many different commodities (forestry, mineral, metal, and chemical products). Forestry commodities are also consumed by material wholesalers. The supply chain for forestry products will be different if they are destined to a construction site as opposed to a material wholesaler. By identifying both the production and consumption industry for each commodity, we can more accurately predict the travel required to bring these products to market.

The framework described here is focused on two approaches: (1) using disaggregate representations of goods movement, and (2) representing the variety of supply chains that may apply to a specific commodity produced by a specific industry and consumed by a specific industry. Local pickup and delivery is explicitly represented by tour-based methods for goods movement and service-related commercial vehicle travel.

1.2 Typology

The following is a list of terms as they are defined for use in this study because many are defined differently in other sources or contexts:

- **Direct Tours.** These are tours with a single destination. A truck or commercial vehicle making a direct tour travels from its home base to the single destination and returns to the home base. A truck might make one or more direct tours each day.
- **Distribution Channels.** The depiction of different components of a supply chain that represent the movement of specific goods, such as from a supplier to a producer to a wholesaler to a retailer to a customer. Distribution channels are modeled as a choice for any specific commodity so that different distribution channels are used to represent the movement of goods rather than assuming that all goods of a certain commodity follow the same distribution channel.
- **Driver Surveys.** These are surveys collected at business establishments, firms who transport goods, or firms who provide services using commercial vehicles. Respondents are truck drivers and as a result, these surveys are limited to truck modes. Current driver surveys focus on tours of travel across one or more days and hence, these surveys are also referred to as **tour-based** or **truck diary** surveys. In this report, we have maintained the reference to tour-based or truck diary surveys to be consistent with the source of the surveys. For the purposes of modeling, these surveys have key similarities.
- **Establishment Surveys.** These are surveys collected at business establishments, firms who transport goods, or firms who provide services using commercial vehicles. Respondents are decision-makers such as operators of commercial vehicle fleets or shippers. Hence these surveys are also referred to as **operator** or **shipper** surveys. In this report, we have maintained the reference to operator or shipper surveys to be consistent with the source of the surveys. For the purposes of modeling, these surveys have key similarities.
- **Intermodal Transfer.** Shipments are sometimes transferred from one mode to another during the course of being shipped from supplier to receiver. For example, a container may be loaded

on to a truck at the supplier location, taken to an intermodal transfer location, and moved onto a rail car for the next part of the trip to the receiver.

- **Long-Haul.** Long-haul goods movements are used to depict those movements that begin or end outside the metropolitan area of interest and travel into, out of, or through the region or areas immediately adjacent to the metropolitan region of interest.
- **Non-establishment Trip End.** Establishment surveys often ask for details of specific shipments from an establishment or deliveries to an establishment. The non-establishment trip end is the other end of the trip from the establishment completing the survey. For inbound shipments, i.e. deliveries to the establishment being surveyed, the non-establishment trip end is the supplier or shipper of the shipment, i.e. the shipment's origin. For the outbound shipments, i.e. shipments from the establishment being surveyed, the non-establishment trip end is the customer or receiver of the shipment, i.e. the shipment's destination.
- **Peddling Tours.** These are tours with multiple destinations. A truck or commercial vehicle making a peddling tour travels from its home base to each destination in succession and then returns to the home base. A truck might make one or more peddling tours each day.
- **Production Mode.** This is the mode of transport on which the shipment is carried, e.g truck, rail, air, water.
- **Shipment Inbound/Outbound.** Establishment surveys often ask for details of specific shipments from an establishment or deliveries to an establishment. Inbound shipments are deliveries to the establishment being surveyed, while outbound shipments are shipments sent out from the establishment being surveyed.
- **Short-Haul.** Short-haul goods movements are used to depict those movements that begin and end inside the metropolitan area of interest, indicating that these goods are produced and consumed within the same region.
- **Stop Duration.** Stop duration is the length of time between a truck arriving at a stop location to a truck departing from a stop location.
- **Stop Location/ Stop Activity.** Truck tours are a sequence of stops with trips linking the stops. Stop location refers to the physical location of the stop; the stop location might also have characteristics such as the industry of the business establishment at the location. Stop activity refers to what takes place at the stop location: for example, a drop off, a pick up, or some ancillary activity such as fueling the truck or the driver getting a meal.
- **Supply Chain.** A series of trips that represents the movements from the supplier to the customer. These supply chains do not begin and end at the same place, because the supplier and the customer are different, even if they are close together.
- **Supply Chain Model.** A model that explicitly represents the transportation elements of the goods movements from a supplier to a customer. These goods may pass through a producer, a wholesaler, or a retailer on the way to the customer.
- **Tour-based Model.** A model that represents the local pickup and delivery of goods or services in a round trip with one or more stops to and from a home base of operation. The local pickup and delivery of goods can be the last piece of a larger supply chain.
- **Tours.** A series of trips that represent the pickup and delivery of goods or services from a home base to multiple destinations. A tour is completed with a return trip to the home base, because tours always begin and end at the same place.

1.3 Contents of this Report

This is the final report on the tour-based and supply chain freight forecasting framework ("framework") developed and demonstrated for the Chicago Metropolitan Agency of Planning (CMAP) region that includes Chicago and surrounding counties.

- Section 2 is a review of literature and data that exists in the field of freight modeling. This is a discussion on past models and data sources, some of which were used in the current work.
- Section 3 outlines the framework, which is used to demonstrate the tour-based and supply chain models for the Chicago metropolitan area. It also provides an overview of the demonstration data and application.
- Section 4 describes each model component of the framework in detail. It reports the data and the model development process along with estimation and demonstration results of each individual component. There is also a discussion on the performance of each model in the demonstration application and suggestions to improve the model.
- Section 5 outlines a data collection plan for obtaining data that would help improve the estimation and application of models in the framework.
- Section 6 summarizes the current effort and the final results.
- Section 7 includes a comprehensive bibliography.
- In the appendix, there is step-by-step documentation of the calculations contained in the model code written to demonstrate the framework for CMAP.

This report is a compilation and synthesis of several technical memoranda and presentations for project meetings that were developed during the course of the study.

1.4 Acknowledgments

The authors wish to thank Federal Highway Administration (Supin Yoder and Brian Gardner) for supporting the research throughout the project and for Chicago Metropolitan Agency for Planning (Kermit Wies) for supporting the application and demonstration project by providing data and feedback on all aspects of the work. Kouros Mohammadian, Kazuya Kawamura and Jane Lin provided invaluable technical assistance in bringing prior research into the framework. The authors of this report include Maren Outwater, Colin Smith, Bhargava Sana, and Jason Chen from Resource Systems Group and John Bowman. We also want to acknowledge those who provided data for our use on the project – Matt Roorda (University of Toronto), Vladimir Livshits (Maricopa Association of Governments), and Jane Lin and Kouros Mohammadian (University of Illinois Chicago). Kazuya Kawamura (University of Illinois Chicago) deserves a special commendation for translating aspects of the Tokyo survey from Japanese.

The authors wish to thank Dr. Rick Donnelly (Parsons-Brinckerhoff) and Professor Kenneth Boyer (Michigan State University) for freely sharing from their experience in the development of TRB Special Report 304, and exchanging ideas regarding prospects for and methods of obtaining the information needed to implement national scale models of the type considered in this project.

2.0 LITERATURE AND DATA REVIEW

2.1 Summary

While tour-based and logistics supply chain methods have been researched in the U.S. and applied in other countries (Canada, Japan, and Europe), there are currently no known examples of an application of these types of freight models in the U.S. for a metropolitan region. This literature review describes survey data that may be used to develop tour-based and supply chain models, and models that have been estimated and/or applied and that could be transferred. Several key sources, summarized in Table 1, were found to be the most applicable to the development of the framework. The first five sources discussed in the literature review are freight movement and truck surveys, the first four of which were obtained to support model estimation during this project. The fifth dataset, from Tokyo, was not available due to data confidentiality issues. The remaining key sources are tour-based and supply chain models that have been estimated and/or applied.

Table 1: Key Sources for Data and Model Elements

Source/Location	Shipper Survey	Driver Survey	Supply Chain Model	Tour Based Model
Phoenix: MAG Survey (2007)	✓	✓		
Texas: CV Survey (2005-2006)		✓		✓
FAME National Establishment Survey (2009-2011)	✓			
Toronto: Peel Survey (2007)	✓	✓		
Tokyo: Metropolitan Goods Movement Survey (1972, 1982, 1994, 2004)	✓		✓	
Denver: DRCOG Survey (1998)		✓		✓
Oregon: ODOT Model			✓	✓
Calgary: CV Model, Survey (2000)		✓		✓
Ohio: Statewide Model (2003-2004)		✓		✓
Netherlands: GoodTrip			✓	
Netherlands: SMILE			✓	
Norway/Sweden National Freight Model System			✓	

Table 2 summarizes the commodity and industry type variables found in the datasets from the four surveys, identifies inherent limitations in the datasets, and notes their sample sizes. A calculation of average trips per vehicle from the commercial vehicle surveys illuminates some of the differences in the coverage of the various surveys. For example, Toronto has the fewest trips per vehicle rate (1.5) and reflects an oversampling of manufacturing and midsize firms in Toronto, leading one to conclude that these types of firms have lower trips per vehicle rates. Toronto also has the smallest sample size, so this conclusion may be limited by this fact. Phoenix had an average trips per vehicle rate (5.5) that reflects only specific industries (agriculture, mining, construction, retail trade, local pickup and delivery, mail/parcel, and for-hire sectors). The Texas dataset had the highest average trips per vehicle rate (8.1) and included all industries and all firm sizes within small to medium size urban areas in Texas. Given the size of this sample and the fact that it was not oversampled or limited, this value is likely the most robust average trips per vehicle result.

Table 2: Survey Data Available for Model Estimation

Location/Type	Commodity/Industry Type	Limitations	Sample Size/Coverage
Texas Commercial Vehicle Survey	SIC codes for the truck's base location 22 Cargo commodity types 15 land use types for stops (at all stops including the base and customer)	Base location of a tour may not be the supplier Land use types for stops may not be coded consistently with other data sources that contain SIC/NAICS codes for industry types	1,711 vehicles 3,760 shipments 13,802 trips Covers 5 urban areas in Texas
Toronto Establishment Survey	Establishment survey has: commodity type, supplier address for inbound shipments, customer address for outbound shipments, establishment industry classification for surveyed establishment	Supplier/customer companies need to be matched with address database containing industry classification	597 firms 2,699 shipments Oversampled manufacturing and mid-sized firms in Toronto
Toronto Vehicle Survey	Truck survey has commodity type and land use type for stops	Land use types for stops may not be coded consistently with other data sources that contain SIC/NAICS codes for industry types	86 vehicles 42 GPS vehicles 131 trips
FAME Establishment Survey	10 Commodity types 6 Supplier or buyer industry types	For each shipment record, the dataset only has either the supplier or buyer of the commodity	241 firms 476 shipments Covers all U.S.
Phoenix Establishment Survey	Commodity type and supplier or customer industry for surveyed establishment in operator survey	Focuses on manufacturing, wholesale trade facilities and warehouse/distribution centers Does not include industry of the customer for outbound shipments/the supplier for inbound shipments.	562 firms 808 shipments
Phoenix Vehicle Survey	Has data on the industry of the truck operator and the addresses and land use type of the stops made but no commodity data	Focuses on agriculture, mining, construction, retail trade, local pickup and delivery, mail/parcel, and for-hire sectors	46 firms 236 vehicles 1,304 trips

The three shipper surveys (Phoenix, FAME, and Toronto) had a range of shipments per firm of 1.4 for Phoenix, 2.0 for FAME, and 4.5 for Toronto. The FAME dataset is unique in that it reflects both short-haul and long-haul freight movements in both urban and rural areas in the U.S. while the Phoenix and Toronto datasets represent select industries and oversamples for firm sizes, which may explain the variation in these results.

2.2 Tour-based Models

The research on tour-based models includes model components for tour generation, vehicle-type models, trip purpose models, time of day models, stop location models, and stop duration models. These are activity-based models that derive methods from the evolving world of activity-based passenger models. They focus on the tour characteristics of truck trips and are less concerned about what is being carried in the vehicle. Thus far, the focus of these models has been on truck trips, so they are not truly multimodal.

Holguín-Veras and Patil (2004) describes a comprehensive analysis of the observed trip chain behavior of commercial vehicles in the Denver region using data collected by the Denver Regional Council of Governments in 1998 and 1999. Wang and Holguín-Veras (2008) describes models developed using these data that estimate tour flows of commercial vehicles.

A well-documented example of a tour based model was developed in Calgary, Canada (Hunt et al. 2003). The model applies tour-based micro-simulation modeling concepts to urban goods movement modeling. The primary source of the data used in development is an extensive set of interviews about own-account commercial vehicle movements conducted at just over 3,100 business enterprises in the Calgary region.

Gliebe et al (2007) describes a disaggregate commercial vehicle model developed using establishment survey data collected by the Ohio Department of Transportation. The model generates entire daily patterns for workers who regularly travel as part of their jobs and creates tours through a dynamic choice process that incrementally builds tours.

There were three models, with data, that were particularly promising for this study: Phoenix, Texas and Toronto. These data and models are discussed in the following sections.

Phoenix Internal Truck Travel Survey

In 2007, Maricopa Association of Governments (MAG), the MPO for the metropolitan Phoenix area conducted a study to update their internal truck travel model, which included conducting internal truck travel surveys (Cambridge Systematics, 2007). The study included two survey components: truck diary surveys and operator surveys. The truck trip diaries were used for sectors that generated multi-stop tours that were short haul in nature. The surveys were designed to collect truck travel information that included origin and destination information, stop locations, land use types at stops, trip lengths, number of trips by truck type and sector, and time-of-day distributions of truck trips. Operator surveys or establishment surveys were used for sectors that generated truck traffic that were long haul in nature. These surveys were conducted by phone and were designed to collect information on the number of inbound and outbound truck trips at each facility or establishment, and the distribution of truck trips by trip distance and time of day.

Truck Diary Surveys

The truck diary surveys were used for the agriculture, mining, construction, retail trade, local pickup and delivery, mail/parcel, and for-hire sectors. The sample of businesses was stratified by employment size, type of industry, and area type. Truck drivers at the sampled businesses completed paper travel diaries that covered a 24-hour period. The diary data items collected included the number of truck trips carried out by each driver, the origin and destination of the trips, the time of day they were made, the type of trucks making the trips, information on the types of commodities being hauled, odometer readings, time, and stop type for each stop. There were a total of 236 completed truck diary surveys out of 3,276 sampled businesses in the dataset. In all, there were a total of 1,304 stops, which translates to an average of 5.5 stop per day among trucks surveyed traveling within the Phoenix metropolitan area. Four types of stops accounted for 75% of the stops: at a construction site (24%), at the respondents company/employer (23%), at a warehouse/wholesale store (17%), and at a house/other residential location (11%).

The truck trip diaries were used for sectors that generate multi-stop tours that are short-haul in nature. Information on stop locations and truck type was collected for a 24-hour period. Just over half of the sample (54%, 127/236) includes 6 digit NAICS codes, while the remainder has business names and addresses. The land use type at the stop location was recorded in the survey, but the commodity dropped-off or picked-up at each of the stop locations was not collected.

Operator Surveys

The sectors that the operator surveys focused on were manufacturing facilities, wholesale trade businesses, and warehouse/distribution centers. The survey was administered by telephone. The survey collected data on the location of the facility, the type of facility, the industry type, the number of trucks by size, the types of trucks, truck ownership, the number of employees, the number of drivers, the type of materials shipped from facility, the number of weekly outbound truck trips, the percentage of trips delivering to multiple locations, the percentage of trips delivering to single locations, the destinations for one day’s worth of trips from dispatch or GPS records, and the nature/land use of destinations. The survey also included data on inbound shipments received at the facility, allowing the shipments through warehouse/distribution centers to be at least partially understood. The survey received 562 completes out of 6,143 businesses contacted, of which 49% were manufacturers, 35% wholesale traders, and 16% warehousing and transportation businesses.

The operator survey focused on sectors that generate truck traffic that are long haul in nature such as manufacturing, wholesale trade facilities, and warehouse/distribution centers. Each establishment is coded to 6 digit NAICS codes (and so can be reclassified into more aggregate categories). The survey includes the questions “Which of the following best describes the primary commodity or good you ship?” and “Which of the following best describes the primary commodity or good that you receive?” The answer options for these questions are 10 commodity types which are groupings of SCTG codes. Applicability to this Project

The survey data collected by MAG is one of the only recent U.S. truck diary and establishment datasets available. The truck diary surveys contain a useful mix of trip tours, including detailed information about the number of stops, dwell times, and trip distances from a range of industries located in a large metropolitan area. The establishment survey contains a relatively large sample of records from three types of industries important in metropolitan areas that generate a large quantity of freight movement and truck trips. The survey data includes information about both inbound and outbound shipments as well as truck movements. Unfortunately, these data are not sufficiently detailed to track individual shipments in a fully disaggregate manner as they travel into, through and out of a facility.

Table 3 summarizes the operator survey shipments by primary commodity and industry type.

Applicability to this Project

The survey data collected by MAG is one of the only recent U.S. truck diary and establishment datasets available. The truck diary surveys contain a useful mix of trip tours, including detailed information about the number of stops, dwell times, and trip distances from a range of industries located in a large metropolitan area. The establishment survey contains a relatively large sample of records from three types of industries important in metropolitan areas that generate a large quantity of freight movement and truck trips. The survey data includes information about both inbound and outbound shipments as well as truck movements. Unfortunately, these data are not sufficiently detailed to track individual shipments in a fully disaggregate manner as they travel into, through and out of a facility.

Table 3: Phoenix Survey, Shipments by Industry and Commodity

	Aerospace and Aviation	Agriculture and Food Processing	Bio-industry	Development Industries	High-Tech	Mining and Primary Metals	Other Basic Industries	Plastics and Advanced Composites	Supplier Industries	Transportation and Distribution	Total
Raw agricultural and animal products	0	5	0	7	0	0	6	3	0	13	34
Food products, alcohol, and tobacco	0	2	0	4	1	0	14	2	4	36	63
Forestry, wood, and paper products	3	10	4	12	13	0	28	5	28	55	158
Chemicals and chemical products	2	2	0	2	2	0	6	1	8	24	47
Petroleum products	0	4	0	9	4	0	17	0	4	24	62
Mining materials	0	2	0	2	1	0	2	1	5	10	23
Manufactured metal and mineral products	1	3	0	27	8	0	38	5	30	60	172
Other manufactured products or equipment	2	6	1	17	9	2	34	5	21	97	194
Waste, refuse, and recycling	0	0	0	0	2	0	0	0	0	2	4
Miscellaneous	1	3	0	7	3	0	4	1	6	26	51
Total	9	37	5	87	43	2	149	23	106	347	808

Note: Highlighted rows are commodities that were used in model estimation.

Texas Commercial Vehicle Survey and Model

Regional freight surveys are good ways of capturing regional goods and services trips, which could provide information about goods pickup/delivery and service delivery in the region. Nationwide there are a number of public agencies that have conducted or plan to conduct regional freight surveys. Table 4 summarizes the existing survey methodologies, including vehicle traffic counts, establishment surveys, roadside intercept surveys, vehicle travel diaries, as well as Global Position System (GPS) surveys.

The Texas Department of Transportation (TxDOT) Travel Survey Program sponsored the external station surveys and the commercial vehicle surveys in a number of MPOs during 2001 and 2006 (Prozzi et al., 2006; Prozzi et al., 2004; Nepal et al., 2007a,b,c). External surveys aimed at capturing the inbound, outbound, and through traffic, while commercial vehicle surveys captured the internal activities in the region. As we will demonstrate later, the Texas data is considered especially suited for this project in understanding the tour structures by commodity and location. Therefore, we will focus on the Texas Commercial Vehicle Survey data in the rest of this section.

Table 4: Summary of Regional Freight Survey Methodology

Survey Approach	Data Collection Techniques	Collected Data
Vehicle Traffic Counts	Manual Counts Automated Counts	▪ Traffic flows by time, day, month and the proportion of total traffic flow ▪ Manual vehicle classification counts and some automated counts .
Establishment Survey (incl. Commodity flow survey)	Face-to-face Telephone Self-completion	▪ Data about total vehicle trips to/from particular establishments, and variation by time, day and month ▪ Data about type of goods delivered/collected ▪ Data about delivery/collection process: vehicle types, time taken to load/unload, where vehicle stopped, method of goods movement from vehicle, and origin of vehicle/goods
Roadside Intercept Survey	Face-to-face Mail-out/Mail-back	▪ Trip details and patterns of vehicles ▪ Origin/destination, trip purpose, goods carried, and vehicle type
Vehicle Travel Diary	Mail-out/mail-back Fax Self-completion Online submission	▪ Detailed information about the activities of a single vehicle (over a single day or few days) ▪ Data about exact locations served, route, arrival and departure times, time taken for delivery/collection/servicing, type of goods/service etc.
GPS Survey	Equipment/transmitter fitted in vehicle	▪ Record vehicle location and speed at frequent intervals ▪ Record stops for loading/unloading/parking

Survey Method and Data Summary

Table 5 provides a detailed illustration of both external station surveys and commercial vehicle surveys. We focused on the data from a subset of cities, San Antonio, Amarillo, Valley, Lubbock, and Austin, during 2005 and 2006 (for more details see Nepal et al. 2007a,b,c) as they are more recent. This particular subset consists of a total of 13,802 trips made by 1,711 commercial vehicles. Surveyed commercial vehicles were randomly sampled from a consolidated database of privately operated certified commercial vehicles in the study areas that included the vehicle registration database, the motor carriers database, and the employer database all maintained by Texas Department of Transportation. Drivers or operators of the sampled vehicles completed both a vehicle information form and a daily travel log on an assigned day. The vehicle information form contained basic vehicle data like vehicle type, vehicle model year, fuel type, gross weight, odometer reading, and vehicle base location information; the travel log recorded all trips the commercial vehicle made and all locations it visited during the study day. Key information recorded in the travel log is summarized below:

- **Longitude and latitude** - Stop coordinates
- **Departure/arrival time** - Departure/arrival time at stop
- **Cargo type (22) -** 1) Farm products, 2) Forest products, 3) Marine Products, 4) Metals and Minerals, 5) Food, Health, and Beauty Products, 6) Tobacco Products, 7) Textiles, 8) Wood Products, 9) Printed Matter, 10) Chemical Products, 11) Refined Petroleum or Coal Products, 12) Rubber, Plastic, and Styrofoam Products, 13) Clay, Concrete, Glass, or Stone, 14) Manufacturing Goods/Equip, 15) Wastes, 16) Miscellaneous Shipments, 17) Hazardous

Materials, 18) Transportation, 19) Unclassified Cargo, 20) Driver Refused to Answer, 21) Unknown to Driver, 22) Empty

- **Total cargo weight** - Total loaded or unloaded cargo weight
- **Activity type (9) -** 1) Base Location/Return to Base Location, 2) Delivery, 3) Pick-up, 4) Pick-up and Delivery, 5) Maintenance (fuel, oil, etc.), 6) Driver Needs (lunch, etc.), 7) To Home, 8) Others (specify), and 9) Refused/Unknown
- **Land use type (15)** - 1) Office Building, 2) Retail/Shopping, 3) Industrial/Manufacturing, 4) Medical/Hospital, 5) Educational (12th Grade or less), 6) Educational (College, Trade, etc.), 7) Government Office/Building, 8) Residential, 9) Airport, 10) Intermodal Facility, 11) Warehouse, 12) Distribution Center, 13) Construction Site, 14) Others (specify), and 15) Refused/Unknown.

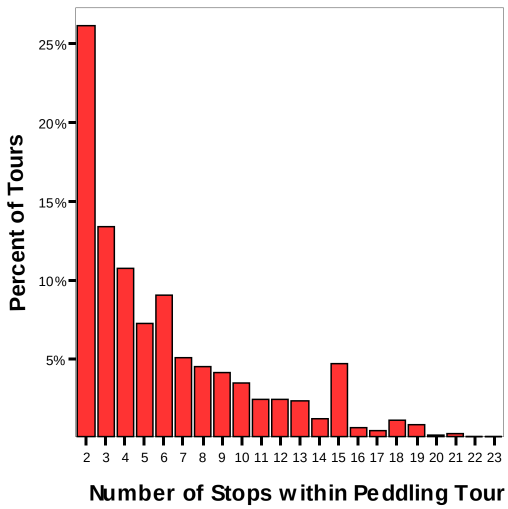
Table 5: Summary of the Texas Truck Data Collection

Topic	External Survey	Commercial Vehicle Survey
Survey Purpose	▪ Provide a clear picture of freight movements on a state's transportation system ▪ Determine the impact of freight on a state's road infrastructure (e.g., bridges and pavements) and the implications in terms of funding ▪ Evaluate strategies for improving freight mobility ▪ Forecast system performance ▪ Mitigate impacts of truck traffic on general mobility ▪ Determine the impacts on air quality ▪ Ensure effective land use planning ▪ Evaluate economic development impacts ▪ Improve the safety and security performance of the road network	
Year of Study	2001-2006	
Commodity Classification System	Self-defined classification	
Vehicle Classification Methods	Single Unit 2-axle (6 wheels), Single Unit 3-axle (10 wheels), Single Unit 4-axle (14 wheels), Semi (all Tractor-Trailer Combinations), Other (specify), and Unknown.	
Survey Method	Face-to-face interview/License plate match	Mail-back Survey
Survey Duration	During daylight hours for one day	One operation day of the vehicle
Sites Selection /Sampling	28 locations across the study area boundary and represent a site where travelers may enter and leave the study area.	Statistical formulas used to determine the number of trucking companies
Expansion Factors	24-hour vehicle classification counts	Highway Performance Monitoring System (HPMS) and Vehicle Classification Counts

Description of Individual Tours

There are a total of 1,493 direct and 1,314 peddling tours identified in the datasets. Figure 1 shows that the frequency distribution of peddling tours decreases as the number of stops in the tour increases. The average number of stops in a peddling tour is 6. More than 25% of the peddling tours have two stops while less than 5% have 15 stops or more. There is a noticeable jump for the number of tours with 15 stops for unknown reasons.

Figure 1. Distribution of Number of Stops within Peddling Tour



Some tour strategies are often chosen for one type of shipment more than other types, suggesting that these shipment strategies are not chosen at random. Instead, many tour strategies seem to be a result of specific shipment requirements. The key findings are highlighted as follows:

- The single direct tour has more than double the average trip distance (i.e., distance between two adjacent stops) and five times the dwelling time at stops than the others
- Within the single direct tour, frequent activities are retail delivery (22.2%) and shipment of manufacturing goods/equipment (26.2%) and infrequent activities are pick-up shipments (3.1%) shipment of construction materials (5.1%), and miscellaneous shipments (6.1%)
- Within the single peddling tour, frequent activities are retail delivery (50.2%), visit to/from distribution center (30.4%) and pick-up shipments (31.6%) and infrequent activities are shipment of construction materials (8.4%)

Applicability to this Project

The Texas commercial vehicle survey data is one of the few datasets in the country that contains rich tour-based information at the stop level, which makes it highly relevant and applicable to this project. This data set contributed to a number of model components in the project framework:

- **Shipment size and type.** The Texas data was used to extract information about shipment size and type in relation to, for example, cargo type, time of day, vehicle type, trip purpose, etc.
- **Tour patterns.** Texas data is essentially tour based. UIC has developed the urban truck tour based and tour-chain models using the Texas data: (1) Tour patterns include direct, peddle, and mixed runs; (2) Tour chain considers the chaining of those tour patterns (Ruan et al., 2011).
- **Stop duration and location.** The Texas data contains detailed stop level information such as location, dwell time, location type, arrival and departure time, loading/unloading commodities, shipment size, and vehicle size/type. This information was used for stop duration and location modeling.
- **Delivery Frequency/Time of Day.** The Texas data contains delivery frequency and time of day per vehicle. The UIC tour models contain delivery frequency and time of day variables.

The industry of the business at the commercial vehicle's base location is coded in 4-digit Standard Industrial Classification (SIC) codes. The base location is not necessarily the location of the supplier for the shipment. SIC codes can be converted to North American Industry Classification System (NAICS) codes.

The cargo commodity classifications in the dataset are shown in Table 6. The commodity classification used in the dataset is based on 2-digit Standard Transportation Commodity Codes[1] (STCC). It is possible to translate STCC codes to Standard Classification of Transported Goods (SCTG) codes (groupings of STCG codes are used in the FAME survey, for example).

Table 6: Texas Survey, Cargo Commodity Classification

Commodity type	STCC Code
Farm products	01
Forest products	08
Marine Products	09
Metals and Minerals	10
Food, Health, and Beauty Products	20
Tobacco Products	21
Textiles	22
Wood Products	24
Printed Matter	27
Chemical Products	28
Refined Petroleum or Coal Products	29
Rubber, Plastic, and Styrofoam Products	30
Clay, Concrete, Glass, and Stone	32
Manufacturing Goods/Equipment	33-36, 38-39
Wastes	40
Miscellaneous Shipments	41
Hazardous Materials	49
Transportation	37

Table 7 shows the total number of outbound and inbound shipments by the industry of the base location and commodity type. As noted above, the base location is not necessarily the location of the supplier for the shipment. For example, although no commercial vehicles captured in the survey were based at Agriculture locations (i.e. farms), shipments of farm products were captured in the survey. The highlighted rows in these, and subsequent tables, indicate the commodities that were used in the model estimation and application phases of this project.

[1] See for example, Commodity Classifications appendix of Quick Response Freight Manual II.

Table 7: Texas Survey, Number of Shipments by Industry and Commodity

	Agriculture	Construction	Finance and Real Estate	Manufacturing	Mining	Retail Trade	Services	Transportation	Wholesale Trade	Total
Farm products	0	85	0	10	17	14	0	134	0	260
Forest products	0	0	0	9	0	43	0	8	3	63
Marine Products	0	0	0	0	0	0	0	0	0	0
Metals and Minerals	0	215	0	27	0	0	4	147	25	418
Food, Health, and Beauty Products	0	0	0	160	0	0	225	15	0	400
Tobacco Products	0	0	0	0	0	0	2	0	0	2
Textiles	0	0	0	0	0	0	92	16	0	108
Wood Products	0	0	0	20	0	131	0	8	18	177
Printed Matter	0	0	0	0	0	0	0	1	0	1
Chemical Products	0	5	0	0	0	0	0	6	10	21
Refined Petroleum or Coal Products	0	8	0	0	0	0	0	14	0	22
Rubber, Plastic, and Styrofoam Products	0	0	0	0	0	0	0	9	0	9
Clay, Concrete, Glass, and Stone	0	12	0	597	3	8	0	64	3	687
Manufacturing Goods/Equip	0	53	0	14	0	191	88	292	47	685
Wastes	0	81	0	1	0	0	11	15	4	112
Miscellaneous Shipments	0	15	0	41	0	18	82	558	72	786
Hazardous Materials	0	0	0	0	0	0	0	0	0	0
Transportation	0	0	0	2	0	0	4	3	0	9
Total	0	474	0	881	20	405	508	1290	182	3760

Note: Highlighted rows are commodities which were used in model estimation.

Toronto: Region of Peel Commercial Vehicle Survey

The Region of Peel (Toronto, Canada) Commercial Vehicle Survey was a data collection effort that collected commodity, mode choice, and commercial vehicle movement data from a sample of approximately 600 shippers and their drivers. The survey was undertaken by the University of Toronto in the time period from October 2006 to May 2007. Peel is a suburban region, located to the west of the City of Toronto, which is a center of transportation, warehousing, and manufacturing activity in the Greater Toronto Area. The survey used a Mail-Out/Mail-Back (MOMB) survey questionnaire and part of the survey sample also completed a GPS-supplement. The survey was designed to understand truck trip generation rates, the characteristics of shipments, and the translation of shipments to a sequence of truck trips.

Shipper Survey

The MOMB shipper survey had an overall response rate of 25% resulting in a total of 597 completed shipper surveys. The MOMB shipper survey was a self-administered survey that contains a set of forms to be filled out by the firm's shipping manager. The MOMB shipper survey collected the following data:

- **Establishment information:** number of employees classified by occupation group; industry classification; square footage; age of establishment; estimated number and total value of annual shipments and services; yearly estimated values of shipments destinations that are local, interregional or international; number of commercial vehicles, classified by vehicle type, that entered and/or departed the establishment on the survey date; approximate annual frequency of shipments by mode; number of commercial vehicles that arrived and left the establishment on the survey date; and number of vehicles owned or leased by the establishment.
- **Outbound commodity information**: information about every good or service shipped by the business establishment on the survey day with name of destination, address of destination, shipment only or service only, mode(s) of transportation used, time sensitivity of the shipment or service, value of the shipment, commodity weight and commodity type;
- **Inbound commodity information**: the same information as the outbound commodity information for the case of inbound (received) shipments, except that the origin of the shipment or service is obtained, rather than the destination.

The sample included multiple industries including some wholesale trade and transportation and warehousing services.

The shipper survey data has been used to estimate models that are reported in Cavalcante and Roorda (2009). The paper presents a discrete/continuous model with shipment size as the continuous variable and vehicle-type choice as the discrete variable. The discrete model shows that small vehicles are more likely for shipments with higher value per unit weight, time sensitive shipments, and services (as opposed to good shipments) and larger vehicles are more likely for long distance shipments. The continuous model shows that shipment size (in kg) decreases with the increase of density values of the commodities ($/kg) and expected fuel operating cost of the vehicle-type chose ($/km) and with the decrease of distance (in km).

The sampling plan for the survey intentionally oversampled manufacturing and mid-sized firms. The industries have been classified into general categories that are groupings of NAICS codes. The non-establishment trip end of the shipment (i.e., the origin of inbound shipments and the destination of outbound shipments) is a business name and/or an address, and the addresses have been geocoded. The address provided by the survey respondent is not necessarily a full street address; it may just be the closest intersection. About 5% of the outbound and 10% of the inbound shipments have no information about the non-establishment trip end. Table 8 shows the number of outbound and inbound shipments by commodity and industry type from the Toronto establishment survey.

Table 8: Toronto Survey, Number of Outbound and Inbound Shipments by Industry and Commodity

	Construction	Manufacturing	Office and Service	Retail	Wholesale Trade and Transportation Handling	Agriculture	Heavy Industry	Transportation Handling	Waste Management	Utilities	Wholesale Trade	Government	Total
Food Products	0	29	24	92	4	27	0	31	0	0	181	1	389
Road Vehicles and Parts	0	20	4	49	11	0	0	12	0	0	19	0	115
Wood, Pulp, and Paper	21	62	36	14	4	0	1	1	0	0	19	2	160
Manufactured Products	12	273	57	94	70	6	27	83	0	0	137	0	759
Metals and Products	16	197	10	23	19	1	9	7	62	0	37	0	381
Non-Metallic Products	19	26	4	16	69	1	9	12	0	0	13	1	170
Oil and Gas	1	1	0	2	0	1	0	4	0	0	0	0	9
Chemical Related	0	26	7	8	0	3	0	8	0	0	450	0	502
Textiles	2	7	0	25	15	0	0	3	0	0	1	0	53
Mixed	0	5	4	14	1	0	5	11	0	0	8	1	49
Other	11	17	17	11	5	1	0	17	22	0	11	0	112
Total	82	663	163	348	198	40	51	189	84	0	876	5	2699

Note: Highlighted rows are commodities that were used in model estimation.

Driver Survey

The MOMB driver survey contained two forms: stop information and tour information. The stop information section collected information from the driver describing the vehicle and the stops made on the survey day including stop descriptions, locations, purposes, and arrival and departure times. The form provided enough space for a total of 14 stops. The tour information section collected behavioral information describing the characteristics of the tour as a whole, such as whether the tour is done routinely, and whether the timing of stops is typical compared to stops made in the past week. For the driver survey, 86 completed surveys were collected. Of the 86 driver surveys, 42 also completed the GPS portion of the survey where GPS units were installed on their trucks for 7 days.

In addition to the establishment survey data there is also a driver component to the survey which has a smaller sample size of 131 trip diaries. The drivers were required to record information about all the stops made on a particular day and also the commodities picked-up/dropped-off at each of the stop locations. The driver survey used the same commodity classification as the establishment survey.

Applicability to this Project

As with the Phoenix/MAG survey data, the sample is entirely from a metropolitan area (although it is a Canadian, and not a U.S., city), and as such the characteristics of the establishments are similar to those found in the Chicago area. The sample size of almost 600 establishments is also similar to that collected in Phoenix and provides data from a cross section of industries, but is smaller than the Texas survey and as a result was not used. The survey data includes information about both inbound and outbound shipments, including sufficient information about the origins of inbound shipments and the destinations of outbound shipments to characterize the industry class at both ends of the shipment of a commodity. This level of detail suggests that the shipment data can be grouped into the 3-way classification scheme of the industry, of the supplier, the industry of the customer, and the commodity being shipped, that is introduced later in this report. As demonstrated in Cavalcante and Roorda (2009), the data support the development of models such as the shipment size and vehicle-type choice model reported in that paper. The driver survey is less useful, given its smaller sample size, but potentially provides an additional source of truck tour information that could be grouped with other survey data.

2.3 Supply Chain Models

The literature review included papers discussing supply chain and logistics models, including some models developed for urban freight studies. Fischer et al. (2005) and Yang et al. (2009) provided summaries of recent developments in supply chain models.

A study by Tavasszy et al. (1998) is a prominent example of a supply chain and logistics modeling effort. They developed a series of disaggregate logistics models, called the Strategic Model for Integrated Logistics Evaluation (SMILE), together with an economic input-output model to provide a decision tool for policy evaluation in the Netherlands. Also, Boerkamps et al. (2000) developed an urban supply chain model, called GoodTrip, for the city of Groningen in the Netherlands. The GoodTrip is a disaggregate model that defines supply chain patterns and urban truck tours, and thereby provides insights into how the logistics decisions affect urban truck traffic.

De Long and Ben-Akiva (2007) also embarked upon the development of a logistics module to be included in the existing freight demand model for Norway and Sweden. This modeling approach is currently being used by Cambridge Systematics in a parallel freight forecasting study by CMAP and will be coordinated with efforts in this study. Holguin-Veras (2000) also discussed an urban freight modeling framework capable of incorporating logistic information and trip chaining behaviors.

In addition to these studies, three other models and data were of particular interest to this study's goals: FAME (a national model), Tokyo, and Portland. These are discussed in more detail in the following sections.

Freight Activity Microsimulation Estimator (FAME) National Establishment Surveys and Model

Survey Data

The University of Illinois at Chicago has developed and is currently continuing with the administration of an establishment survey. The goal of the survey is to collect information on Individual Shipments that are not available from other aggregate sources of data.

The survey questionnaire is broken into three main parts:

- **PART I, Establishment Information.** This includes establishment location, number of employees, suppliers, number of shipments, industry type, modes used, and access to intermodal facilities and warehousing.
- **PART II, Shipment Information.** This includes origin, destination, mode, commodity, value, weight, volume, cost, time, and other data about a set of individual shipments made by the establishment.
- **PART III, Contact Information.** This is for the establishment.

To date, two waves of the survey have been completed and a third is underway.

- Wave I was completed in November 2009. A total of 25,997 firms were initially targeted by emails, and a person with comprehensive knowledge of the firm's supply chain and transportation activities was contacted. A total of 316 establishments participated in the survey providing information on 881 shipments across the country.
- Wave II was completed in December 2010. A total of 12,000 firms were targeted with phone calls followed by emails. A total of 71 establishments participated in the survey.
- Wave III, which is currently underway, is targeting 100,000 firms by emails. In this wave, the goal is to collect data from 500 establishments.

Freight Activity Microsimulation Estimator (FAME) Model

Freight Activity Microsimulation Estimator (FAME) is a freight activity-based modeling framework with five basic modules. In the first module, all the firms in the study area are recognized and their basic characteristics are identified. Based on each firm's characteristics, the types and amounts of incoming and outgoing goods are determined, and the design of the supply chains is replicated in the second module. In the third module, the shipment sizes are defined based on the acquired information on the firms' characteristics and the way that they trade commodities between each other. The fourth module in which the shipping decisions are made is of great importance, because it predicts the decisions for shipping mode, haul time, shipping cost, warehousing, etc. Even though sophisticated firms make decisions on the physical infrastructure of the supply chain and logistics strategies simultaneously; those decisions are treated separately in order to make the modeling structure compatible with the available data. Finally, in the last module, the impact of the goods movements on the transportation network is investigated.

Applicability to this Project

The survey provides a disaggregate dataset of a significant number of shipments and as such is a valuable source of information about the inter-relationships between industry types, commodities, and the characteristics of the shipments that they make. The survey is a national sample, including both metropolitan and non-metropolitan establishments, and the establishment and shipment trip end locations are known so shipments can be further distinguished by area type.

The FAME model has some characteristics that are similar to the proposed framework, although it is designed specifically as a national-level model. These include the synthesis of firms, the supplier selection occurring in the supply chain module, and modeling shipment sizes directly.

The goal of the survey was to collect information on individual shipments that are not available from other aggregate sources of data. Industry type is based on NAICS codes. Commodity codes correspond to groups of SCTG codes.

The commodity classes with the highest number of shipments in the FAME sample are Machinery / Metal products and Wood / Paper / Textile / Leather products. These correspond with SCTG categories 32-34 and categories 25-30 respectively. Among the total 316 respondents, 92 were companies in the Manufacturing (NAICS code: 31-33) industry and 102 were companies in the Transportation and Warehousing (NAICS code: 48-49) industry. Table 9 shows the number of outbound and inbound shipments by commodity and industry type respectively.

Table 9: FAME Survey, Number of Outbound and Inbound Shipments by Industry and Commodity

	Agriculture, Forestry, Fishing, and Hunting (NAICS code: 11)	Manufacturing (NAICS code: 31-33)	Mining, Quarrying, and Oil and Gas Extraction (NAICS code: 21)	Transportation and Warehousing (NAICS code: 48-49)	Wholesale Trade (NAICS code: 42)	Information (NAICS code: 51)	Total
Agricultural products	11	11	0	13	1	0	36
Motorized and other vehicles (incl. parts)	0	8	5	13	10	0	36
Gravel/ Natural sands/ Cement	0	5	3	1	0	0	9
Chemical / Pharmaceutical products	0	28	0	13	9	0	50
Machinery / Metal products	1	49	0	32	18	2	102
Coal / Mineral / Ores	0	7	8	3	0	0	18
Wood / Paper / Textile / Leather products	0	45	0	41	23	2	111
Electronics	0	34	0	20	16	0	70
Mixed freight / Miscellaneous	0	12	0	48	11	0	71
Prepared foodstuffs	0	11	0	37	15	0	63
Total	12	210	16	221	103	4	566

Note: Highlighted rows are commodities that were used in model estimation.

Tokyo Metropolitan Goods Movement Survey and Model

Survey Data

The Tokyo Metropolitan Goods Movement Survey has been conducted four times, in 1972, 1982, 1994, and most recently in 2004. The 1982 survey obtained the largest sample, with information from approximately 46,000 firms. The data consists of records on commodity movement and truck movement of each firm. Each record provides information about firm characteristics, commodity movement, and truck movement. The information about firm characteristics includes industry type, number of employees, floor area, and other related information. In the same way, the information on commodity movement and vehicle movement includes commodity type, weight carried in and out, delivery frequency, truck type, carrier type, and other related information.

The more recent survey, from 2004, consisted of four sections:

- **Main survey**: a mail survey of establishments
- **Interview survey:** a business & logistics strategy of the 36 largest businesses, and a similar mail survey of 1,400 smaller businesses
- **Truck route survey:** where forwarders were asked to identify their main routes
- **Local delivery survey:** five case studies of "last mile" activities

The population sampled in the main survey was 120,000 business establishments in the Tokyo area. The FIRE, utility, mining, farming, and fishing sectors were not included, but all distribution facilities were included. Of the 30,000 business establishments contacted, 12,000 responded.

The survey included questions about the facility and business characteristics (size, site factors, functions, etc.), the origins of inbound shipments and the destinations of outbound shipments, in addition to the weight, mode, truck load, truck load factor, commodity, time window utilization, and vehicle type for each shipment.

Tokyo Metropolitan Area Freight Micro-simulation Model

Wisetjindawat and Sano (2003) developed a micro-simulation model for the modeling of urban freight transportation in the Tokyo Metropolitan Area using the 1982 survey data and validated it using the 1994 survey data. The proposed model considers the behavior of each freight decision maker individually. The model structure consists of three stages: commodity production and consumption, commodity distribution, and conversion of commodity flow to vehicle movement. The paper discusses the issues involved in model development and validation including the conceptual framework, mathematical formulation, and estimated results.

- **Commodity Production and Consumption Model -** Production and consumption amounts are estimated from the firm characteristics such as number of employees, floor area, and other indicators. This model utilizes regression techniques to estimate production and consumption amounts.
- **Commodity Distribution Model -** The model makes a connection from demand to supply according to the attractiveness of suppliers and distribution channels. The attractiveness of suppliers is derived from supplier location and the amount of commodity produced by that supplier. The model selects locations and shippers using multinomial logit models estimated using the survey data. Distribution channels were defined as the paths connecting the customers and shippers of a commodity. In a supply chain, there are a number of possible distribution channels. The probability of a given distribution channel being used for a given

commodity is calculated directly from the empirical data. The probability varies by the industry type of the customers, the industry type of the shippers, and the type of purchased commodity.

- **Conversion of Commodity Flows to Trip Chain -** Commodity flows are assigned to vehicle movements in three steps: delivery lot size and frequency, carrier and vehicle choice selection, and vehicle routing. The delivery lot size models are developed separately by the industry type of the customers and by purchased commodity type.

Applicability to this Project

The Tokyo surveys are particularly applicable to the framework discussed in this document and potentially useful for the development of models that will form elements of the framework. The surveys constitute a very large urban sample: 46,000 records, in 1982, with smaller samples (but still large by U.S. standards) in 1972, 1994, and 2004. The survey itself is also very comprehensive, and, as discussed above, the data have been used to develop models of commodity production and consumption, commodity distribution, and finally models that convert commodity flows to truck trip chains. The model system includes many of the elements discussed in the framework introduced later in this report. Unfortunately, we were not able to gain access to the survey data for model estimation purposes on this project and so these data were not used.

Portland, Oregon MOSAIC Model

Donnelly (2010) describes the development of the MOSAIC model system that is implemented in a metropolitan context in Portland, Oregon. The model is a microsimulation model that models (separately) each of the important decisions that characterize the demand for freight transport.

Table 10 shows the components of the model. Several data sources were used to develop various elements of the model. For example, the Canadian National Roadside Survey was used to estimate trans-shipments by commodity; the observed distributions of shipment size by commodity and mode were derived from surveys in Michigan and Oregon for trucks and from the Surface Transportation Board (STB) Carload Waybill Sample for rail; and data from the Commodity Flow Survey (CFS) and the Vehicle Inventory and Use Survey (VIUS) were used to determine carrier types.

Table 10: MOSAIC (Portland/Oregon) Model Components

Model Component	Description
Economic Drivers	Estimates overall value ($) of commodities produced in the region
Model Alternatives	Mode choice by commodity; tonnages by mode and commodity allocated to firm agents
Trans-Shipment	Probability of commodities being handled as a trans-shipment
Export and Import	Models to account for import and export flows
Shipment Generation	Flows disaggregated to daily flows and then individual shipments
Destination Choice	Allocates each shipment to a receiving firm
Carrier and Vehicle Choice	Private or for-hire carrier and then vehicle type
Tour Optimization	Shipments are allocated to vehicles and routed using the Traveling Salesman Problem (TSP) methodology.

3.0 Overview of the Tour-Based and Supply Chain Freight Forecasting Framework

3.1 Modeling System Overview

The tour-based and supply chain freight modeling framework is comprised of a series of model components which together identify the primary decisions for moving freight in a metropolitan region:

- Which firms are involved in moving freight?
- What is the demand of production and consumption of goods?
- How are these goods distributed in the region?
- What modes do the goods travel by?
- How many and what size trucks are needed to move the goods that are transported by truck?
- What time of day do these trucks travel?

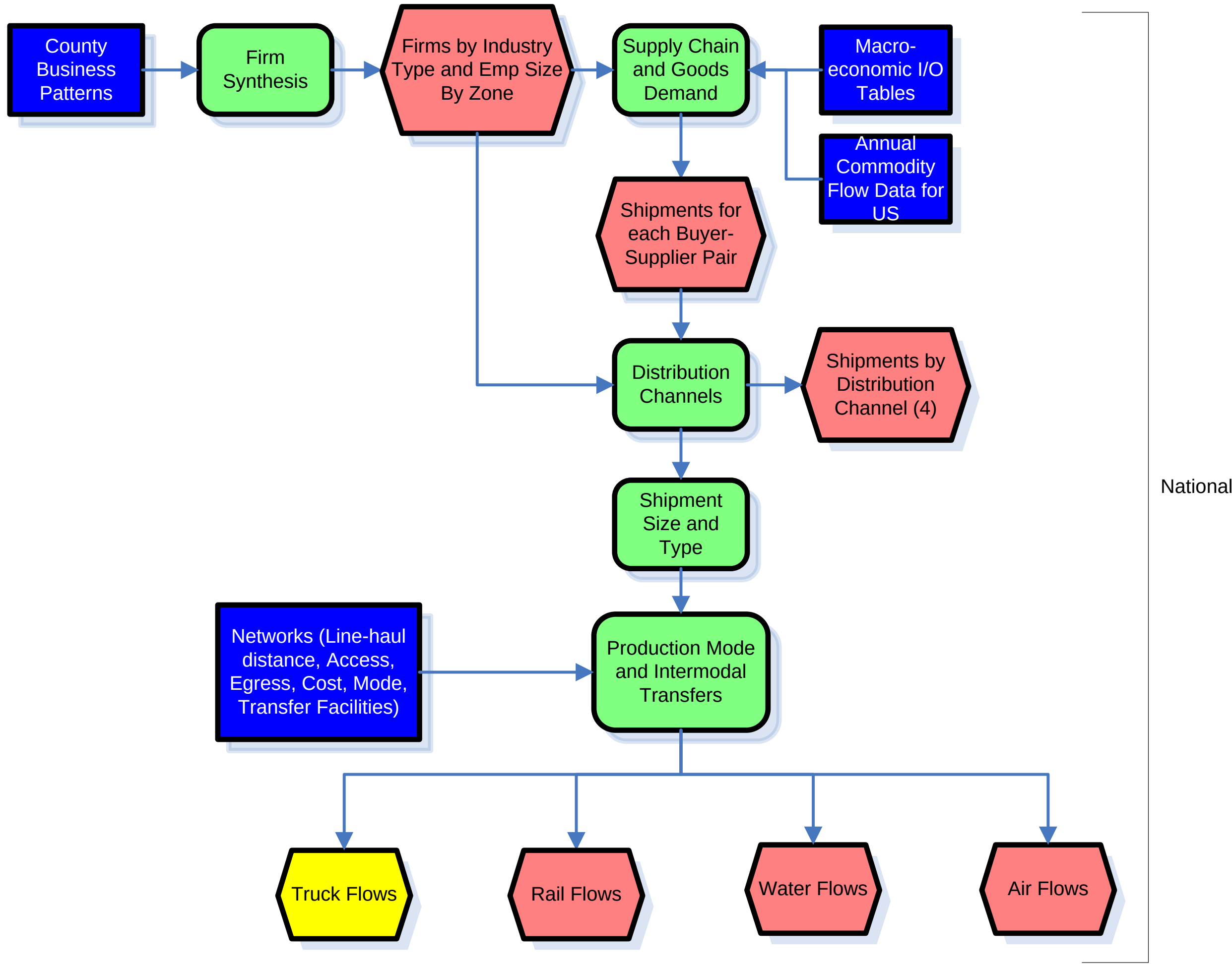
There is a final decision regarding the route choice for trucks that completes the framework, but, once the individual truck movements (trips) are simulated in the model system, this can be completed using existing aggregate assignment methods in use by MPOs around the country and so it is not a focus of this framework.

Each of the primary decisions is further subdivided into specific choices that are represented by individual model components. There are currently two primary types of freight or goods movement models that have included these model components: supply chain and tour-based models. Examples of these models were discussed in the literature review and demonstrated the different purposes and uses of these models to date. These modeling methods are brought together in the framework into a single modeling system.

The framework's sequence of model components is presented below in two figures, followed by a brief introduction to each model component. Figure 2 shows the national supply chain model and Figure 3 shows the regional tour-based truck model. The primary data inputs (economic, socioeconomic, and network data) are shown supporting specific model components. The flow of data between model components is also shown, although specific data products from each model component are not included for brevity (these details are elaborated in Section 4.0 which discusses each component of the model in turn).

After mode choice, the framework focuses only on trucks, since the capacity of the road system is a limitation for many metropolitan areas and a focus of regional planning and rail, although some rail systems also have capacity limitations, it is not feasible in most cases to accurately capture rail movements, which are often more than 1,000 miles long, in urban models. We expect that advancements in freight forecasting and modeling will subsequently include assignments of air and water modes, but we have not included these in the framework at this time in order to focus on the most important planning questions.

Figure 2: National Supply Chain Section of the Freight Forecasting Frameworks



National Supply Chain Model

- **Firm Synthesis -** The initial element of the framework synthesizes all firms in the study area, which is defined as the U.S., for purposes of capturing long-haul freight movements. The geographic detail within the region of interest is based on jurisdictional boundaries such as cities, and the geographic detail outside of study area is defined by Freight Analysis Framework zones. This model synthesizes firms by industry category and by size category to capture the primary drivers of commercial vehicle travel. Firm synthesis can be controlled by regional, county and state control totals obtained from State, MPO, or national sources.
- **Supply Chain and Goods Demand** - The next element of the framework predicts the demand in tonnage for shipments of each commodity type between each firm in the synthetic population. The demand represents the goods produced by each firm and the goods consumed by each firm in the U.S. The model is applied in two steps. In the first step, buyers who have a demand for goods are paired with suppliers who sell those goods using a probabilistic model. The connections between industry types for each commodity are based on input-output tables. Once the buyer-supplier relationships are established, the amount of commodity shipped on an annual basis between each pair of firms is apportioned based on the number of employees at the buyer and their industry so that observed commodity flows are matched nationwide.

- **Distribution Channels** - Using a multinomial logit model, each shipment between each buyer-supplier pair is assigned a probability of choosing a specific distribution channel to represent the supply chain it follows from the supplier to the consumer. The model predicts the level of complexity of the supply chain, for example whether it is shipped directly, or whether it passes through one or more warehouses, intermodal centers, distribution centers, or consolidation centers.
- **Shipment Size** - Shipment size is estimated using a discrete choice model based on a variety of firm, commodity and travel characteristics. It is as this point in the model that the units of analysis change from annual commodity flows between pairs of firms to discrete shipments that are individually accounted for and delivered from the supplier to the buyer.
- **Modes and Intermodal Transfers** - There are four primary modes (road, rail, air, and water) that are modeled. Detailed networks of road and rail for the U.S. are used, with simpler functions of distance and the value of goods being transported to represent the air and water modes. The modes and transfer locations on the shipment paths are determined based on the travel time, cost, characteristics of the shipment (perishable, expedited, containerized) and characteristics of the distribution channel (whether the shipment is routes via a warehouse, consolidation or distribution center, and whether the shipment includes an intermodal transfers, e.g. truck-rail-truck).
- **Daily Shipments and Warehouse Selection** - Once the modes and intermodal transfers have been assigned, the shipment list is converted from all annual shipments to a daily sample to represent the day being modeled. This component of the model can be calibrated to allow for seasonal variations in commodity flows where that information is available. This component of the model also assigns shipments to specific warehouse, distribution, and consolidation centers if the shipment passes through one of those locations.

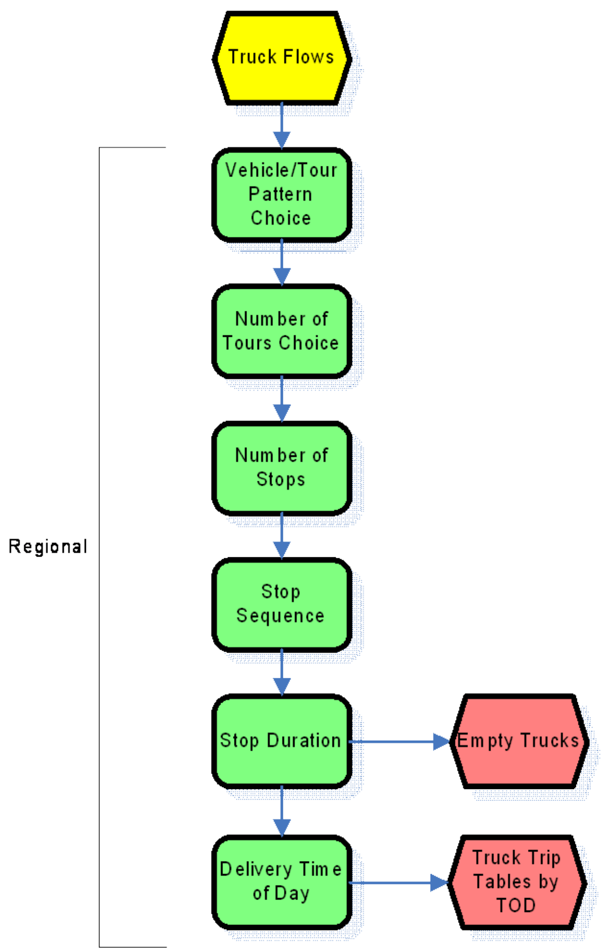
Regional Tour-Based Truck Model

The regional truck is a sequence of models that takes shipments from their final transfer point (i.e. the last point at which the shipment is handled before delivery, which is at the supplier for a direct shipment, or at a warehouse, distribution center, or consolidation center for shipments with a more complex supply chain) to their final delivery point. It does the same in reverse for shipments at the pick-up end, where shipments are taken from the supplier and taken as far the first transfer point. For shipments that include transfers, the tour-based truck model accounts for the arrangement of delivery and pick-up activity of shipments into truck tours. The models are sequenced as shown in Figure 3.

- **Vehicle and tour pattern choice** - This multinomial logit model predicts the joint choice of whether a shipment will be delivered on a direct tour from transfer to delivery (i.e. where a truck departs the transfer location, delivers the shipment, and returns to the transfer location) or a peddling tour where the truck makes multiple deliveries or pick-ups, and the size of the vehicle that will make the delivery.
- **Number of tours choice** - This multinomial logit model predicts the complexity of the peddling tour that a shipment is contained in; for example a truck might return to the transfer point after one large loop or might break their delivery schedule into two, three, or more tours.
- **Number of stops -** This model uses hierarchical clustering to divide the shipments into spatially collocated groups that can be reasonably delivered by the same truck.
- **Stop Sequence** - This model uses a greedy algorithm to sequence the stops in a reasonably efficient sequence but not necessarily a shortest path (our research shows that touring trucks only sometimes deliver in a sequence that is efficient in shortest path terms).

- **Stop Duration -** This multinomial logit model predicts the amount of time taken at each stop based on the size and commodity of the shipment.
- **Delivery time of day** - This multinomial logit model predicts the departure time of the truck at the beginning of the tours. Based on this, the travel time of each trip, and the stop duration of each delivery, all of the trips on the tour can be associated with a time period for assignment purposes.

Figure 3: Regional Tour-Based Truck Model



3.2 Classification Scheme

An important element of understanding the movement of commodities through supply chains is the shipment relationship between the supplying industry and the buying industry. A very detailed way of representing these relationships and the commodity movement between suppliers and buyers is with a 3-way classification scheme that identifies the industry of the supplier, the industry of the customer, and the commodity being shipped. This classification scheme is significantly more detailed than the 1-way classification scheme of commodity groups used in most freight forecasting models in practice today. The 3-way classification scheme identifies differences in shipping decisions for a particular commodity depending on the supplying and buying industries. Figure 4 is an example of a cross-classification of supplier and customer industries by commodity the group used in the Tokyo models.

Figure 4: Example of a 3-way Classification of Supplier and Customer Industries by Commodity Group

	Customer Industry Groups												
Supplier Industry Groups	1	2	3	4	5	6	7	8	9	10	11	12	13
1) Agriculture, Forestry, and Fishery													
2) Mining													
3) Construction													
4) Chemical Manufacturers													
5) Metal Manufacturers													
6) Machinery Manufacturers													
7) Other Manufacturers													
8) Material Wholesalers													
9) Product Wholesalers													
10) Retailers													
11) Warehouses													
12) Electricity, Gas and Water Suppliers													
13) Service and Government Work													

Commodity Groups:

1) Agricultural Products
2) Forestry Products
3) Mineral Products
4) Metal and Machinery Products
5) Chemical Products
6) Light Industry Products
7) Other Products
8) Wastes and Scraps

This type of classification scheme was found to be impossible to develop with the current U.S. freight survey data available during this project, since the surveys did not include sufficiently detailed questions about the customer industry. Some freight surveys in foreign cities, such as the large-scale effort carried out in Tokyo in 2004, do collect data in such a way that it is possible to identify the customer industry, but this dataset was not available for use in this study for confidentiality reasons. Instead, a simpler 2-way classification of supplier industry group by commodity group, which was supported by the survey data that were available, was explored and used in the framework.

3.3 Demonstration Project

The approach to this research recognized that the data to estimate a new freight forecasting model was not available in Chicago and the specific details of the preferred data collection program would not be known until the framework was more fully defined and tested. The details of the data collection program were completed as part of this research after the development of the framework and the testing of the framework on the demonstration project in the Chicago region; the data collection is described in Section 5.0. The demonstration project is, therefore, a proof of concept to test that the framework can be applied using local Chicago economic, socioeconomic and network data. The estimation of the model components for the framework was based on available freight survey dataset that contains the desired variables. The model estimations were based on several different datasets (one for each model component) and so the resulting models do not necessarily represent local Chicago behavior, but they do contain many of the desired sensitivities and have reasonable coefficients.

As this is a demonstration project, two classifications of commodity flows were identified that allowed the estimation and application of models for freight. Based on the cross-tabulations of commodity and industry types presented in Section 2.0, the following two commodity types were selected for project demonstration:

- Manufactured Goods (SCTG 32-35, 38-40)
- Food Products (SCTG 1-9)

The commodities were selected on the basis of two factors:

1. The sample sizes from available surveys. The two commodity groupings account for a large proportion of the data (in terms of the number of individual shipments or participating firms) in the surveys:
    - 48% of shipments from National (FAME) Survey
    - 57% of shipments from Toronto Survey
    - 37% of shipments from Texas Survey
    - 36% of firms from Phoenix Survey
2. Varied Industry and Supply Chain Characteristics: The commodity groups are different kinds of goods and have varied supplier/consumer industries and supply chain attributes. Two contrasting commodity types for the demonstration were focused on to produce interesting analytical outcomes from the framework and to open the possibility of finding that the framework is applicable across the full range of commodities.

Table 11 shows a summary and comparison of the commodity classifications in the different datasets and identifies the specific commodities that were categorized into the two proposed commodity groups. The sample sizes of the selected commodity types have been highlighted in the previous section's tables.

Table 11: Commodity Classification Comparison

Dataset	Industry Classification	Commodity Classification	Manufactured and Metal Products	Agricultural and Food Products
Texas	SIC (4 digit), can be converted to NAICS	22 categories (18 are STCC2 codes or groupings of STCC2 codes, plus "empty", "unknown to driver", etc.). STCC2 codes can generally be converted to similar groupings of SCTG codes	Manufacturing Goods/Equip	Farm products, Marine Products, Forest products, Food, and Health and Beauty Products
Toronto	NAICS, grouped into 14 high level categories	11 categories (groupings of SCTG codes, answers to survey question so cannot be disaggregated) used in both the establishment and driver survey	Manufactured Products and Metals and Products	Food Products
FAME	NAICS, grouped into 6 categories, but potentially could be disaggregated	10 categories (groupings of SCTG codes, answers to survey question so cannot be disaggregated)	Machinery / Metal products, Electronics	Agricultural products and Prepared foodstuffs
Phoenix	NAICS (6 digit) for both establishment and driver	Establishment survey: 10 categories (groupings of SCTG codes, answers to survey question so cannot be disaggregated) Driver survey: does not include question on commodity carried	Manufactured metal & mineral products, Other manufactured products or equipment	Raw agricultural and animal products, Food products, alcohol, and tobacco

3.4 Demonstration Application

The demonstration application, which is a full implementation of the framework for the Chicago region, is coded in the R programming language (using R version 2.13). This open source software is available from The R Project for Statistical Computing (http://www.r-project.org/). The input data for the demonstration application was obtained from the CMAP Mesoscale Model developed by Cambridge Systematics Inc.

The software development process was a three stage process. First, input data from the CMAP Mesoscale Model was converted from SAS (a statistical analysis application) datasets\databases to text files. In the second step, the national scale models in the CMAP Mesoscale Freight Model were converted from SAS to R. Finally, the additional national and regional scale models in the framework were developed in R.

The model components converted from SAS are:

- Firm synthesis
- Supplier selection
- FAF flow apportionment and business location

- Mode-Path selection

The model components directly coded in R

- Distribution channel
- Shipment size
- Vehicle choice and tour pattern
- Stop clustering and sequencing
- Stop duration
- Time of day

The demonstration application achieves a total runtime of between 80 and 90 minutes using an HP Z200 Workstation with a Intel Core i7 CPU 870 @ 2.93 GHz processor, 12.0 GB of RAM, and using the Windows 7 Professional (64-bit) operating system. Table 12 shows the runtimes for each model component in the demonstration application.

Table 12: Run times for model components in the demonstration application

Model Component	Run Time (minutes)	Notes
Firm synthesis	13	Synthesize 8 million firms and choose buyers (7.5 million) and suppliers firm types (1.4 million) for CMAP simulation
Supplier selection	24	Match supplier firm types for about 3 million firms
FAF flow apportionment and business location	19	Apportion FAF flows for 3 million buyer supplier pairs and locate 8 million firms to mesozones
Distribution channel	1.0	Predict distribution channels for 3 million buyer-supplier pairs using logit shares
Shipment size	1.5	
Mode-Path selection	20	Evaluation of annual logistics and transport costs for 54 mode-paths
Vehicle choice and tour pattern	1.5	Daily simulation for 300k deliveries\pick-ups from warehouses
Stop clustering and sequencing	1.5	
Stop duration	0.2	
Time of day	1.5	

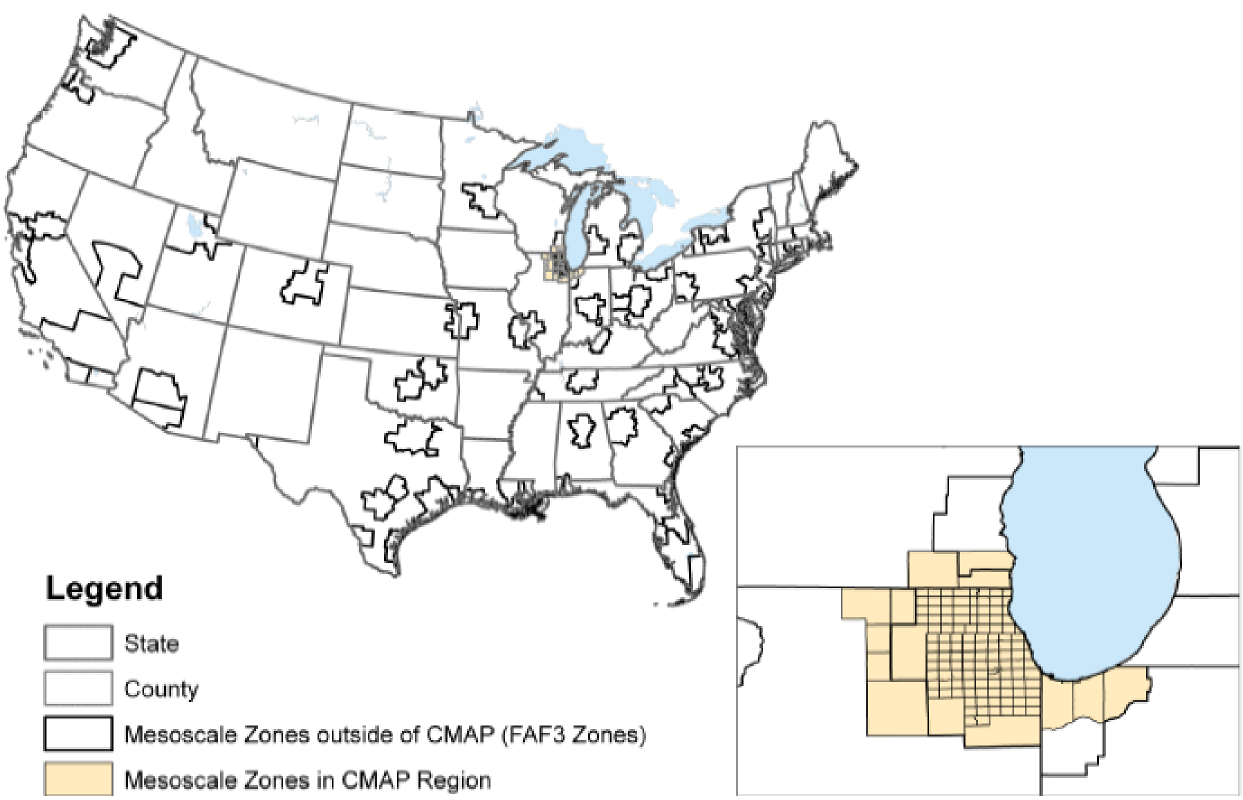
The demonstration application is a single R script implementing all of the model components. All the input data required is in the form of *.csv files located in the same directory as the script. The script reads and processes the input data when the particular input is required by a model component. All the information required for the subsequent simulation is stored in memory as R data frames or data.tables. At various intervals R workspace images (*.RData) are saved alongside the script and input data for debugging and output analysis.

4.0 Elements of the Tour-Based and Supply Chain Freight Forecasting Framework

4.1 Firm Synthesis

This initial element seeks to synthesize all firms in the study area, which is defined as the contiguous U.S., for purposes of capturing long-haul freight movements. The geographic details within the region of interest (in this case Chicago) are defined by jurisdictions and the geographic details outside the region are defined by Freight Analysis Framework version 3 (FAF3) zones. This geographic detail is consistent with parallel work conducted by CMAP for freight forecasting and will allow for compatibility among the data and models for the demonstration project (Figure 5). This model synthesizes firms by industry category and by size category to capture the primary drivers of commercial vehicle travel. Firm synthesis is controlled by regional, county and state control totals obtained from MPO or national sources.

Figure 5. Mesoscale Zone System for the CMAP Prototype Freight Model



Source: Federal Highway Administration for FAF zones and Cambridge Systematics for Mesoscale zones

Data Sources and Model Development

Firms by size and type are allocated to analysis zones using available observed data sources on employment by type, consistent with the data used in the passenger travel demand forecasting model. The observed data were obtained from prior work done at CMAP for fright forecasting. The data were primarily from County Business Patterns (CBP) collected by the U.S. Census Bureau but agriculture, construction and foreign employment numbers were enhanced using data from local sources (Cambridge Systematics, 2011). Tables 13 and 14 summarize the enhanced CBP data by NAICS and employee size category for the nation and CMAP region, respectively.

Based on the available data, it was decided that the model for firm synthesis would be a direct enumeration process of the firms based on the employment totals. Firms could be enumerated by two attributes – industry (NAICS) and the employee size category for each geographic unit.

Model Application and Results

Figure 6 shows a schematic of the firm synthesis process applied. The firm synthesis process enumerates lists of firms, then allocates these firms to zones and identifies them as either producers (suppliers) or consumers (buyers). Figure 7 depicts the firms enumerated in the CMAP region.

Figure 6. Firm Synthesis Process

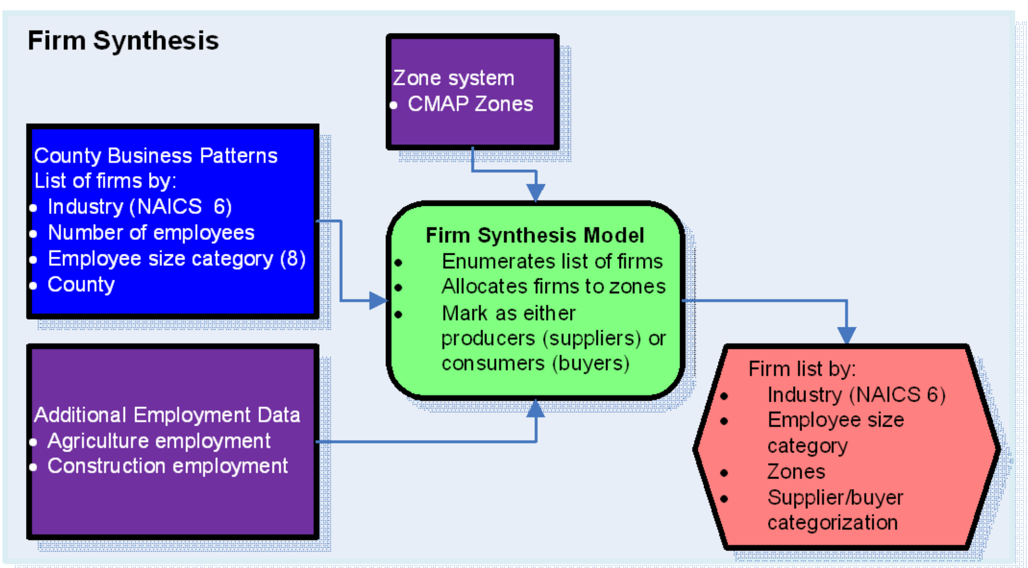


Table 13: National Firm Type Distribution by Industry and Employee Size Category

NAICS 2	Description	Employee Size Category								
		1-19	20-99	100-249	250-499	500-999	1000-2499	2500-4999	Over 5000	Total
11	Agriculture, Forestry, Fishing and Hunting	631,703	83,328	11,941	2,897	954	798	462	429	732,512
21	Mining, Quarrying, and Oil and Gas Extraction	19,713	4,399	664	215	98	185	166	160	25,600
22	Utilities	11,535	3,814	850	288	128	35	3	2	16,655
23	Construction	743,449	65,394	8,576	3,208	2,137	1,840	1,735	1,714	828,053
31	Manufacturing	228,381	74,451	18,942	6,170	2,384	3,260	2,575	2,481	338,644
42	Wholesale Trade	370,817	53,957	6,281	1,445	466	129	13	7	433,115
44	Retail Trade	963,962	132,041	22,687	4,491	399	37	7	0	1,123,624
48	Transportation and Warehousing	183,356	28,958	5,114	1,369	671	227	51	27	219,773
51	Information	116,504	20,803	3,961	1,265	584	237	61	33	143,448
52	Finance and Insurance	461,141	36,743	4,823	1,701	984	471	69	18	505,950
53	Real Estate and Rental and Leasing	363,511	14,644	1,480	324	113	46	3	0	380,121
54	Professional, Scientific, and Technical Services	800,852	55,971	7,083	1,973	726	274	73	26	866,978
55	Management of Companies and Enterprises	33,326	11,571	3,144	1,383	685	350	63	31	50,553
56	Administrative and Support and Waste Management and Remediation Services	319,534	46,408	11,940	3,495	1,244	383	50	13	383,067
61	Educational Services	66,653	16,252	2,466	679	415	300	47	44	86,856
62	Health Care and Social Assistance	664,691	94,252	17,654	3,844	1,753	1,448	388	113	784,143
71	Arts, Entertainment, and Recreation	105,503	16,454	2,348	576	199	109	15	3	125,207
72	Accommodation and Food Services	452,549	169,686	8,721	946	377	160	28	12	632,479
81	Other Services (except Public Administration)	694,640	45,377	3,409	548	163	40	12	3	744,192
	Total	7,231,820	974,503	142,084	36,817	14,480	10,329	5,821	5,116	8,420,970

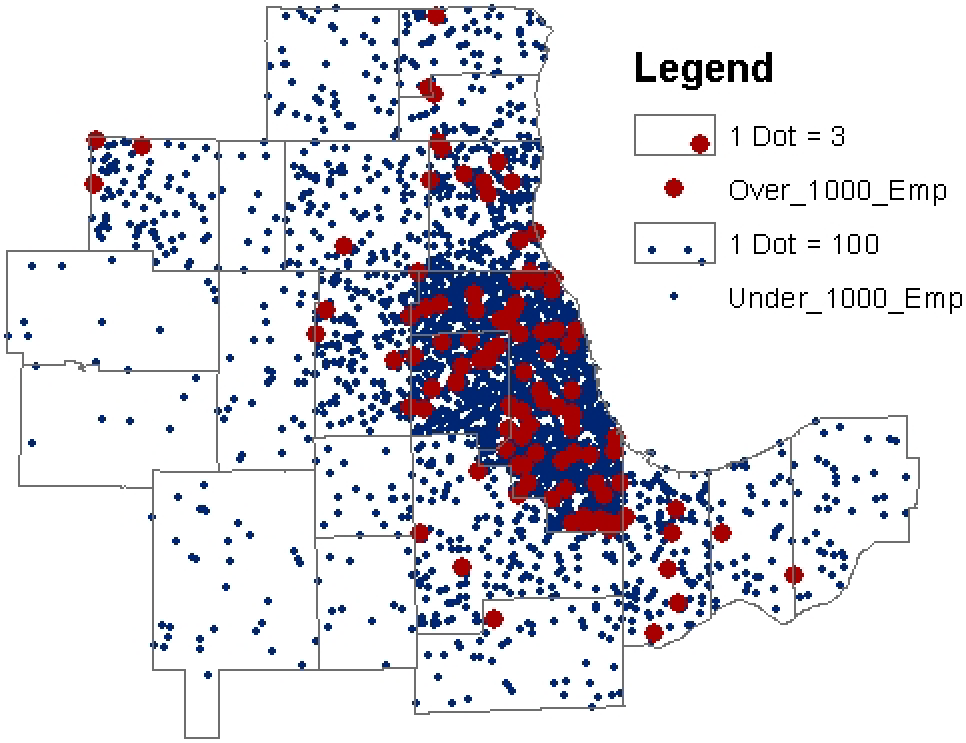
Source: Cambridge Systematics, 2011, Table 4.3, page 4-5

Table 14: CMAP Region Firm Type Distribution by Industry and Employee Size Category

NAICS2	Description	Employee Size Category								
		1-19 Emp	20-99	100-249	250-499	500-999	1000-2499	2500-4999	Over 5000	Total
11	Agriculture, Forestry, Fishing and Hunting	12,950	1,718	240	41	10	1	0	0	14,960
21	Mining, Quarrying, and Oil and Gas Extraction	107	31	6	2	1	0	0	0	147
22	Utilities	165	43	25	9	12	2	0	0	256
23	Construction	27,542	2,236	342	126	42	10	7	2	30,307
31	Manufacturing	8,918	3,576	936	245	81	22	4	2	13,784
42	Wholesale Trade	14,536	2,254	297	69	30	10	2	0	17,198
44	Retail Trade	27,976	4,552	1,054	128	15	3	0	0	33,728
48	Transportation and Warehousing	7,854	1,051	266	55	22	11	3	3	9,265
51	Information	3,648	688	140	46	20	6	0	0	4,548
52	Finance and Insurance	17,539	1,715	277	71	43	19	5	0	19,669
53	Real Estate and Rental and Leasing	11,345	541	72	16	2	2	0	0	11,978
54	Professional, Scientific, and Technical Services	33,190	2,147	284	89	43	17	3	1	35,774
55	Management of Companies and Enterprises	1,116	474	167	72	47	17	4	2	1,899
56	Administrative and Support and Waste Management and Remediation Services	11,494	1,736	485	177	77	26	1	2	13,998
61	Educational Services	2,354	695	90	29	14	20	1	2	3,205
62	Health Care and Social Assistance	22,183	2,816	568	98	54	66	14	5	25,804
71	Arts, Entertainment, and Recreation	2,992	577	77	17	8	4	0	0	3,675
72	Accommodation and Food Services	15,397	5,553	303	46	15	8	0	0	21,322
81	Other Services (except Public Administration)	22,844	1,873	164	34	13	1	1	0	24,930
	Total	244,150	34,276	5,793	1,370	549	245	45	19	286,447

Source: Cambridge Systematics, 2011, Table 4.3, Page 4-5

Figure 7. Firms in the Chicago Region by Size



A correspondence between NAICS and SCTG is used to attach a unique SCTG commodity to each of the enumerated firms (Samimi et al, 2010). The firms are then split into producers and consumers based on whether or not they produce any commodities. Some industries produce more than one commodity (such as wholesale). To account for this, commodity for each firm is simulated based on probabilities of the multiple commodities that it could produce. A producer firm types database is also created which is used later in the supplier selection model. A firm type is defined by a unique combination of industry NAICS, commodity SCTG, and the geographic id of a firm.

A consumers/users database is created next. This consists of all the firms in the firm database merged with processed I/O data (user5top_array.csv) based on NAICS of the consumer. The processed I/O data (user5top_array.csv) identifies the top 5 SCTG commodities (in terms of value) consumed and the corresponding supplier NAICS for each consumer industry NAICS. It was thought that top 5 would include most of the commodities consumed a particular firm and hence this limit was set to decrease computational burden. Similar to what is done during firm synthesis; SCTG commodity for suppliers who could produce more than one SCTG commodities is simulated using probability thresholds. This is done for all the top 5 suppliers that are being considered. It is being assumed that a certain percentage (in this case, 30%) of consumers would work with a wholesaler instead of directly with a producer. Therefore, some suppliers to consumers who themselves are not wholesalers, are probabilistically mutated to an appropriate wholesale supplier (NAICS) based on the SCTG commodity being consumed.

Finally, two datasets are created in this step. A makers/suppliers dataset is created from the producers firm type database and consists of firms that are located in a Chicago zone, all firms that have at least 1,000 employees and one firm by each unique industry-commodity type for each geography. A users/buyers dataset is created from the consumers database and consists of firms that are located in a Chicago zone, firms that have more than 500 employees, one firm by each unique industry-employee size category for each geography, and a set of 5% of randomly selected consumers. All these firms are summarized by location in Table 16.

Table 15. Supplier and Consumer Firms by Location Type

Location	All firms	Suppliers	Supplier Firm Types	Consumers	Modeled suppliers	Modeled consumers
Domestic - CMAP region	286,447	40,537	4,621	258,189	4,621	258,189
Domestic - not in CMAP region	8,125,427	1,339,460	59,900	7,310,902	28,233	484,544
Foreign	9,096	9,096	6,861	9,096	6,861	9,096
Total	8,420,970	1,389,093	71,382	7,578,187	39,715	751,829

Source: Cambridge Systematics, 2011 and Samimi et al, 2010 adapted for use in this project

4.2 Supplier Selection and Goods Demand

The next element pairs up buyers and suppliers among the firms that have been synthesized in the previous step and predicts the demand in tonnage for shipments of each commodity type by each firm in each industry. The demand is developed to represent the goods produced by each firm and the goods consumed by each firm (and household) in the U.S.

Data Sources and Model Development

The models for supplier selection and goods demand are based on earlier freight modeling work at CMAP (Cambridge Systematics, 2011) and Samimi et al, 2010). For each buyer/consumer firm a supplier is selected from the suppliers/makers dataset. The selection of supplier does not mean the selection of an exact business but that of a firm type (a combination of industry NAICS, commodity SCTG, and the geographic id of a firm). The exact firm is determined after the next step of business location (of each firm to a mesozone) is done.

The probability of a supplier being paired with a buyer firm (type) depends on the employment sizes of both the firms and the geographic distance between them. The coefficients used are presented in Table 17. These coefficients are attested and not estimated due to unavailability of data of this nature. In general, the probability of a supplier firm being chosen increases with its employee size and its proximity to the buyer firm. The distances between the firms are Great Circle Distance (GCD) values obtained from Oak Ridge National Laboratory (ORNL) county-to-county skims.

Table 16. Supplier Selection Parameters

Consumer Business Size (Number of Employees)	Coefficient							
	Producer Business Size (Number of Employees)			Great Circle Distance between Consumer and Producer (Miles)				
	1 to 99	100 to 499	500+	Over 1,509	596 to 1,509	150 to 595	1 to 149	0 (Intracounty)
1 to 99	0.2	0.2	0.4	-0.4	-0.3	-0.2	0	0.1
100 to 499	0.2	0.6	0.6	-0.2	-0.1	-0.05	0	0.1
500+	0.4	0.6	0.6	-0.1	-0.05	0	0	0.1

Source: Cambridge Systematics,2011, Table 2.8, page 2-23

Once buyer and supplier pairs have been established, the annual flow between each of the pairs is estimated. The FAF3 flows dataset is used to apportion goods demand to each buyer supplier pair based on the size of the buyer firm. An estimate of consumption (of the commodity being consumed) by a buyer firm is calculated based on the value (in dollars) consumed per employee which is obtained using Input-Output (I/O or make-use) economic tables. The values consumed per employee are calculated for each combination of supplier-buyer industry NAICS from the I/O tables.

Model Application and Results

Figure 8 shows a schematic of the supplier selection and goods demand model. A choice set of suppliers is created for each buyer firm based on the top 5 commodities it requires and the corresponding NAICS of the suppliers. Also, a supplier firm is excluded from the choice set, if no flows for the commodity being traded are observed in the FAF3 dataset between the relevant FAF zones. The GCD values are merged based on the buyer and supplier CBP zones (which are counties). A score for each buyer and potential supplier pair is then calculated using the attested coefficients and adding a random value for stochasticity. For each buyer firm, the supplier firm with the best (highest) score is selected. Table 18 shows the number of buyer-supplier firm pairs by geographic region.

Figure 8. Supplier Selection and Goods Demand Model Process

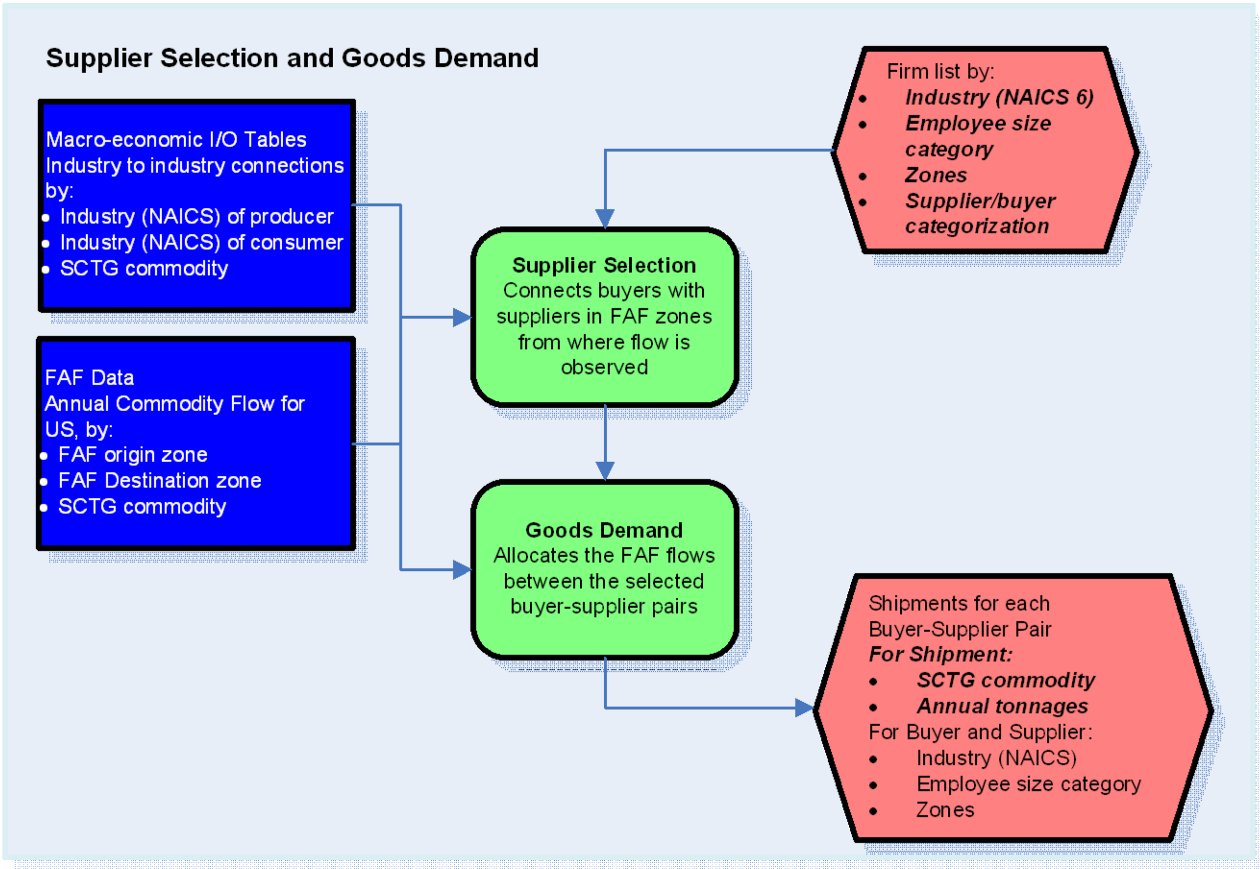
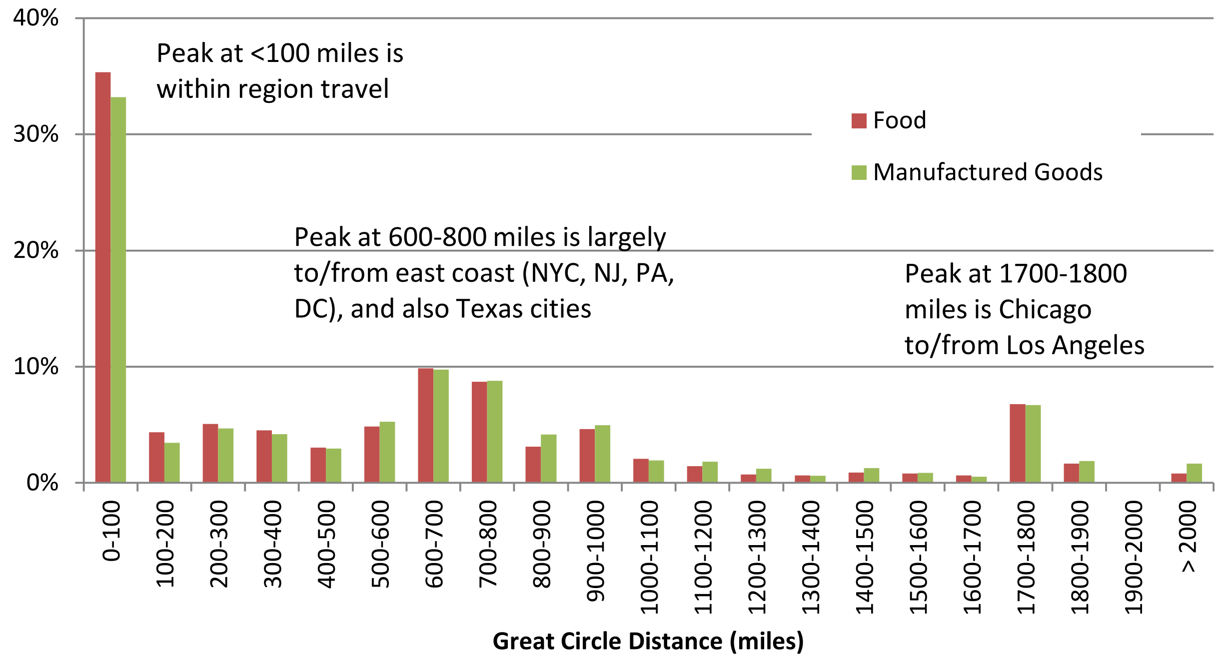


Table 17. Shipments by Buyer and Supplier Firm Locations

Supplier Firm Location	Buyer Firm Location	Number of Firm Pairs
CMAP Region	CMAP Region	1,270,372
CMAP Region	External	2,271,175
External	CMAP Region	7,452

A frequency distribution of the GCDs between buyer-supplier pairs was plotted to check of reasonableness of the supplier selection model (**Error! Reference source not found.**). Each of the three peaks in the plot are for destinations that have higher trading with the Chicago region.

Figure 9. Distance Distribution of Buyer-Supplier Pairs



The values consumed per employee are used to calculate a consumption estimate (in dollars) for each buyer firm. A share of consumption for each firm in a particular FAF3 zone is then calculated based on the consumption estimate. These shares are used to apportion FAF3 flows for each commodity into a FAF3 zone to individual buyer firms. This results in an estimate of annual goods demand between each of the buyer-supplier pairs. Table 19 shows the aggregate flows (values and tons) between CMAP and external regions.

Table 18. Commodity Flows by Location (Tonnage and Value)

Flow direction	Tons	Value ($)	SCTG
CMAP-CMAP	365,094,523	209,935,463,752	All
CMAP-External	279,106,389	369,765,253,227	All
External-CMAP	111,754,530	171,306,433,160	All
CMAP-CMAP	41,468,521	29,694,462,726	Food
CMAP-External	63,124,523	37,068,773,659	Food
External-CMAP	18,254,015	13,272,772,700	Food
CMAP-CMAP	30,807,818	83,647,403,614	Mfg
CMAP-External	53,246,106	166,094,491,885	Mfg
External-CMAP	14,121,701	97,658,015,279	Mfg

Figures 10 and 11 show the flows from the CMAP region for food and manufactured commodities, respectively. Figures 12 and 13 show the flows into the CMAP region for manufactured and food commodities, respectively.

Figure 10. Commodity Flows for Food Products from the CMAP Region

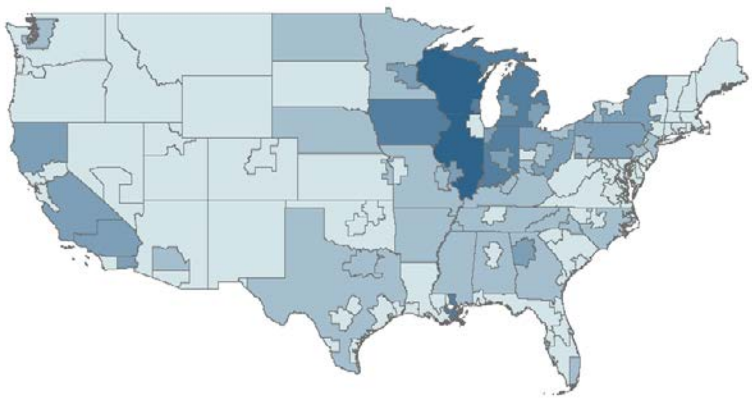


Figure 11. Commodity Flows for Manufactured Products from the CMAP Region

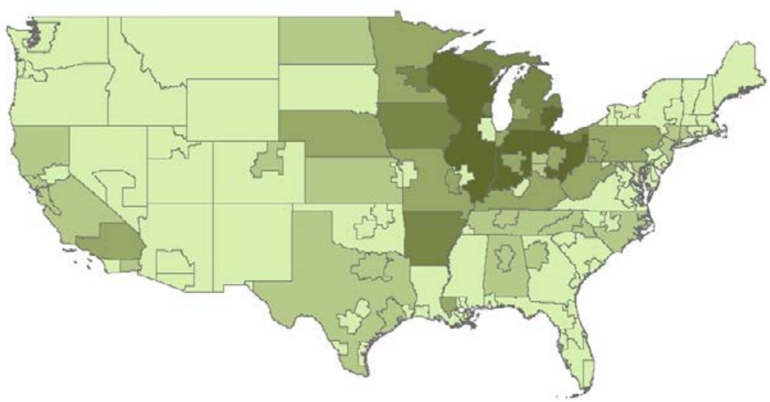


Figure 12. Commodity Flows for Food Products into the CMAP Region

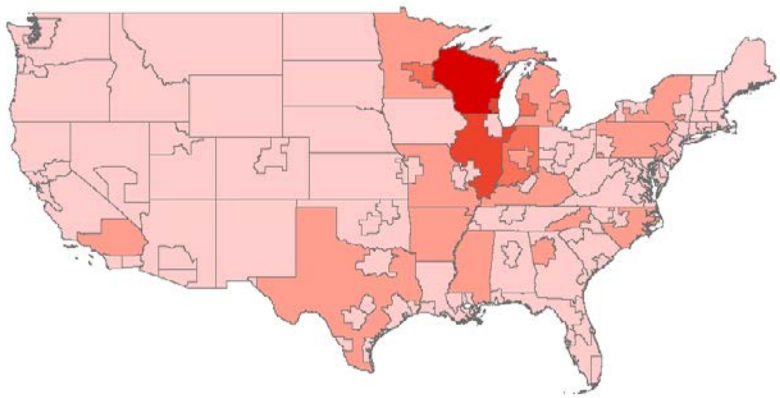
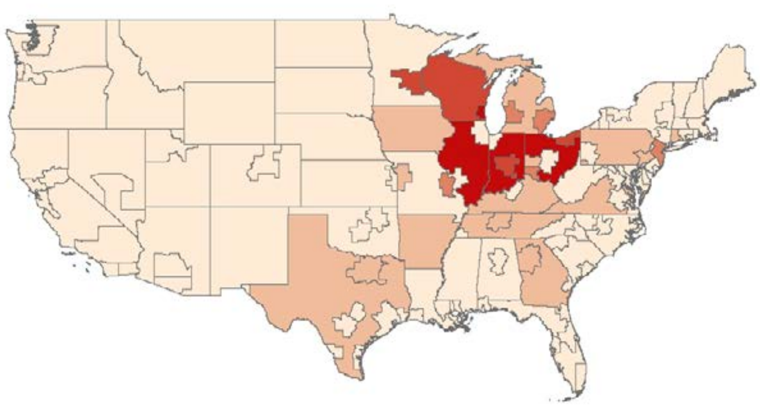


Figure13. Commodity Flows for Manufactured Products into the CMAP Region



Business/Firm Location

Until this step, the geographic identifier for all zones has been the CBP zone (represented by counties for zones within CMAP and FAFs zones for those outside). For the purpose of mode choice and simulation of freight traffic the firms in CMAP region are now assigned to a higher resolution geographic id called the mesozone. There are a few counties/CBP zones which correspond to only one mesozone. The firms in these counties are directly assigned a mesozone. The other counties correspond to more than one mesozone. Firms in these counties are assigned to mesozones based on employment ranking by industry.

A dataset is prepared from employment data which contains the percentile ranking of each of 21 NAICS categories by mesozone based on employment numbers in each of those industries. Higher employment numbers implies a higher percentile rank. For industries in each of the 21 NAICS categories considered, candidate mesozones are identified based on firm size and the ranking of a particular NAICS in a mesozone. The probability of a mesozone getting assigned to a particular firm increases with the rank of the firm's NAICS in the mesozone and the number of employees in the firm. For example, if a firm belongs to the manufacturing industry and has a firm size greater than 5,000 then all mesozones which have manufacturing ranked 9^{th} or 10^{th} are candidates for the particular firm. Once candidate mesozones are assigned to each firm, one of the candidates is randomly selected as the firm's mesozone. Firms not in the Chicago region are assigned a mesozone number based on their CBPZone number and combined with the Chicago firms dataset to create a full firms database with mesozones attached.

4.3 Distribution Channels

A distribution channel refers to the supply chain a particular shipment follows from the supplier to the consumer/buyer. The supplier firms may use their own transportation resources or send shipments to the buyer using third-party logistics firms. The distribution channel might affect the cost, shipment size and frequency of shipments between a buyer-supplier firm pair.

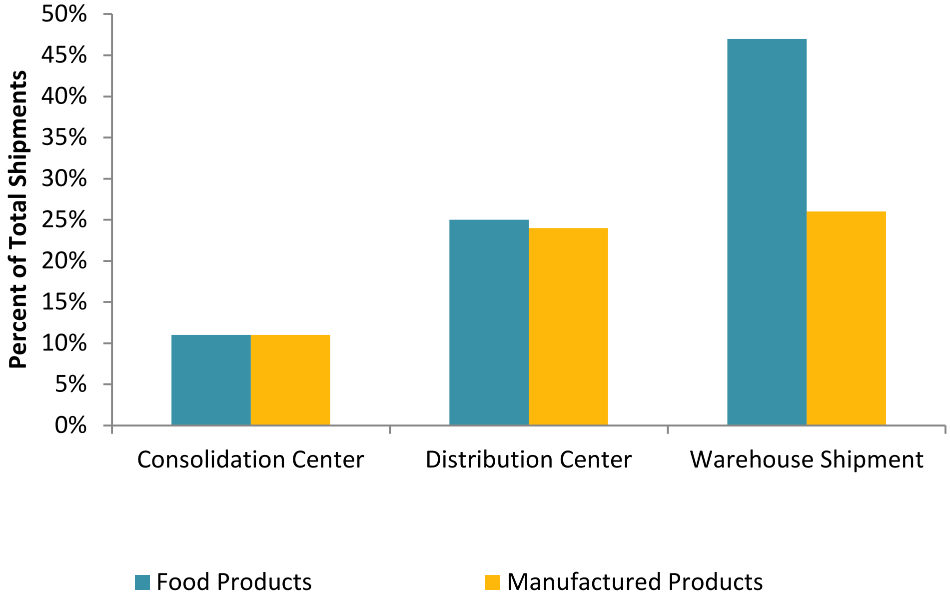
Another version of a supply chain model (FAME) was developed by the University of Illinois at Chicago survey by Samimi et al (2008). This model identifies a supplier and buyer pair of firms using a fuzzy, rule-based model based on their distance from each other and their employee sizes. In this framework, we prefer to identify the full scale of the supply chain to include all establishments that goods move through as they travel from the producer to the consumer. Unfortunately, data for these detailed supply chains is limited. The supply chain model uses discrete choice methods to identify the unique aspects of the supply chain. The initial model developed for this framework was simplified due to data limitations, but the conceptual approach will also work when more detailed data sources are available.

Data Sources and Model Development

The idea was to use the Tokyo dataset to estimate a logit model for the choice of distribution channel. Since Tokyo dataset could not be obtained, other (available) data sources were explored that could potentially be used to develop a distribution channel model.

The FAME establishment survey was used to represent data on the supply chain, which contained data on whether the goods went through a consolidation center, a distribution center, and/or a warehouse. Other aspects of the supply chain were not possible with this dataset and other datasets did not have the national coverage or details about the supply chain for this purpose. The Tokyo dataset was considered for this purpose and did contain more information about the supply chain, but was not available for use in this project. Due to the lack of other data sources, it was determined that the FAME data be used to identify a portion of the supply chain. Figure 13 shows the proportions of shipments that use particular elements of a distribution channel for the two proposed commodity types from the FAME survey.

Figure 13. Shipments with Intermediate Transfer Locations



The concept of distribution channel was further simplified to obtain a reasonable sample for model estimation (Figure 14). Four alternatives for distribution channel were identified – Direct, one stop type, two stop types, and three stop types where stop type is a warehouse, distribution center, or consolidation center. So the distribution channels that involved only one warehouse stop or only on distribution center stop were considered the same.

Figure 14. Distribution Channels

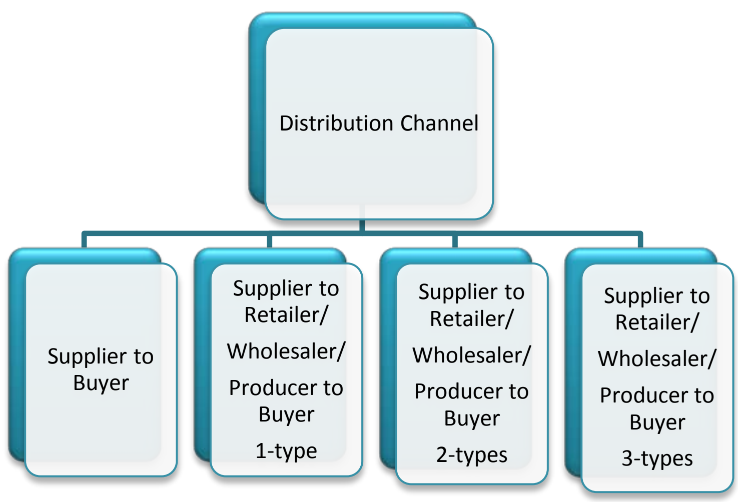


Table 20 shows the Multinomial Logit (MNL) model estimated for food products from the FAME survey. As it can be seen from the constants, the direct distribution channel is the mode preferred everything else remaining constant. The other variables that affect the choice of distribution channel in a significant manner are firm size and the industry type of the firms involved. It is acknowledged that the estimated model does not have a very rich specification but it is reasonable given the data constraints.

Table 19. Distribution Channel Model Specification for Food Products

Choices	Variable Description	Variable Name	Coefficient	t-stat
Direct	Alternative Specific Constant	ASC_V1	0 (fixed)	-5.26
1-Type Used	Alternative Specific Constant	ASC_V2	-0.841	-2.39
2-Types Used	Alternative Specific Constant	ASC_V3	-3.44	-3.33
3-Types Used	Alternative Specific Constant	ASC_V4	-3.44	-3.33
Direct	49 or less employees firm involved	EMP49_1	0.881	2.15
1-Type Used	Manufacturing industry firm involved	MFGIND2	1.85	3.43
2-Types Used	Transportation\warehousing or wholesale trade firm involved	TRWIND3	3.71	3.47
3-Types Used	Transportation\warehousing or wholesale trade firm involved	TRWIND4	3.37	3.12
Number of Observations	**Final Log Likelihood**	**Rho-squared**		
182	-176.182	0.302		

Table 21 shows the distribution channel MNL model estimated for manufactured goods. The kind of explanatory variables are the same as those in the food products model.

Model Application and Results

Figure 15 shows a schematic of the distribution channel model. The distribution channel choice was simulated for shipments between all the buyer-supplier pairs based on the type of commodity. For commodities other than food or manufactured, the manufactured good model was applied. Figure 16 show the distribution of various distribution channels based on geographic flow.

Table 20. Distribution Channel Model Specification for Manufactured Products

Choices	Utility Equations			
	Variable Description	Variable Name	Coefficient	t-stat
Direct	Alternative Specific Constant	ASC_V1	0 (fixed)	-5.26
1-Type Used	Alternative Specific Constant	ASC_V2	-1.96	-5.25
2-Types Used	Alternative Specific Constant	ASC_V3	-2.68	-6.53
3-Types Used	Alternative Specific Constant	ASC_V4	-3.58	-6.04
3-Types Used	50 to 199 employees firm involved	EMP3_4	1.32	2.06
1-Type Used	200 or more employees firm involved	EMP4_2	0.698	1.91
1-Type Used	Transportation\warehousing or wholesale trade firm involved	WHIND2	1.88	4.8
2-Types Used	Transportation\warehousing or wholesale trade firm involved	WHIND4	1.5	3.16
Number of Observations	**Final Log Likelihood**	**Rho-squared**		
182	-176.182	0.302		

Figure 15. Distribution Channel Model Process

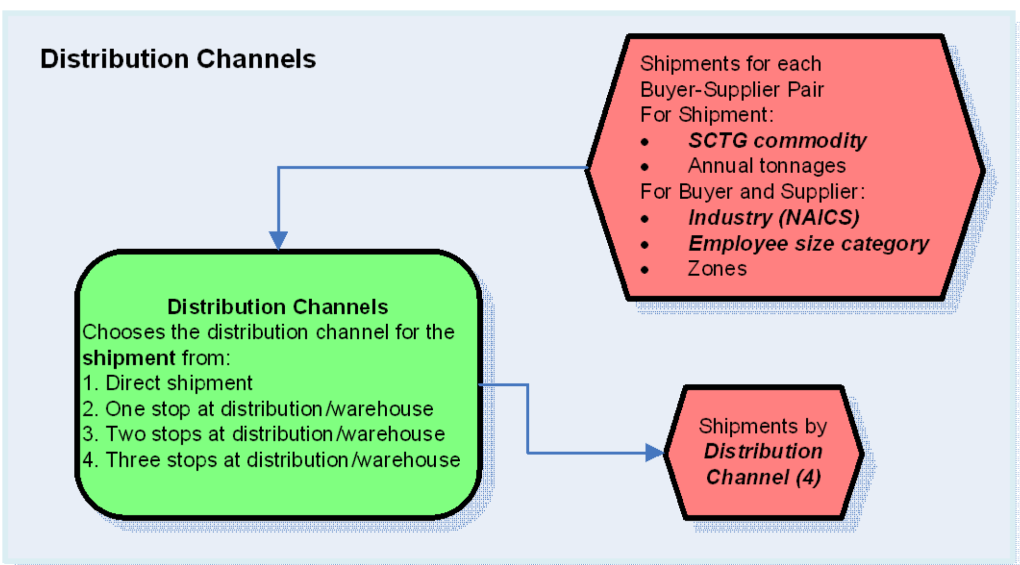
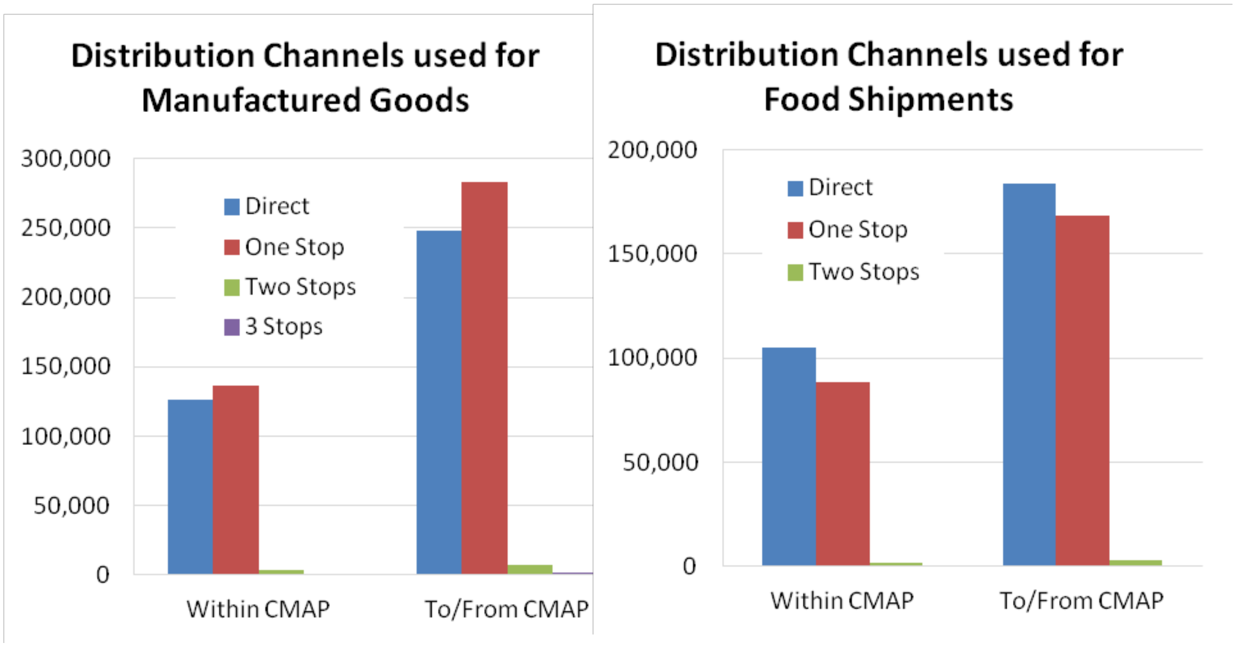


Figure 16. Distribution Channels for Manufactured and Food Products



In case of food products, it appears that a higher proportion of shipments are direct than in case of manufactured goods. It would have been really useful if the distance between the firm pairs was also found to be significant in this model.

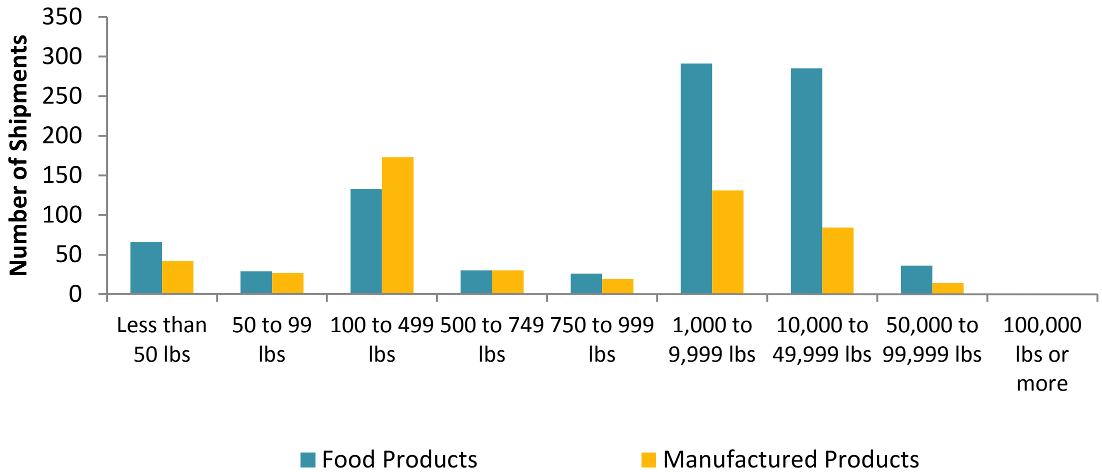
4.4 Shipment Size and Frequency

In this step, the annual goods flow between buyer-supplier firms pairs are broken down into individual shipments. The shipment size (weight) and the corresponding number of shipments per year are determined. Shipment size affects the mode used to transport the shipment. This framework is not designed to optimize the shipments or identify the logistics of how shipments may be combined to make a truckload or rail delivery.

Data Sources and Model Development

An MNL model is estimated for choice of shipment size. The Texas commercial vehicle survey dataset was used for estimating the discrete choice model due to its relatively high sample size. This dataset is not ideal for the shipment size model because the shipments represented in the dataset are likely to within an urban area. However it was thought to be the most appropriate considering the sample sizes in other datasets. Figure 17 shows the distribution for shipment sizes for food and manufactured products from the Texas dataset.

Figure 17. Number of Shipments per Year by Size for Food and Manufactured Products



Based on the distribution, three alternatives selected to form the choice set – less than or equal to 999 lbs, 1000-9999 lbs, and greater than 9999 lbs. It was hypothesized that the distribution channel would influence the choice of shipment size. The distribution channel was not directly available in the Texas dataset. The stop level data were transformed into tour level data and the distribution channel was assigned based on the stops made by the truck at ports, intermodal facilities, warehouses, and distribution centers.

Separate models are estimated for food and manufactured products. Table 22 shows the shipment size choice model results for food products. It appears that the shipment size between 1000 and 9999 lbs is the most preferred for food products everything else being equal. An indirect distribution channel in which the shipment stops at three types of facilities seems to positively influence the highest shipment size category (>= 10,000 lbs). The other explanatory variables in the model specification are trip length until current shipment stop from the base location and industry types at the stop location. Longer trip lengths seem to be associated with shipments greater than 10,000 lbs.

Table 21. Shipment Size Model Specifications for Food Products

Choices	Utility Equations			
<= 999 lbs	ASC_V1 * one + SIC11 * SIC1 + SIC21 * SIC2			
1000-9999 lbs	ASC_V2 * one + DISTCHAN12 * DISTCHAN			
>=10000 lbs	ASC_V3 * one + SIC23 * SIC2 + DISTCHAN32 * DISTCHAN_2 + Cost2 * cost			
Choices	**Variable Description**	**Variable Name**	**Coefficient**	**t-stat**
<= 999 lbs	Alternative Specific Constant	ASC_V1	0	
1000-9999 lbs	Alternative Specific Constant	ASC_V2	0.546	3.85
>=10000 lbs	Alternative Specific Constant	ASC_V3	-1.71	-5.98
>=10000 lbs	Trip Length	Cost2	0.245	2.48
1000-9999 lbs	Distribution Channel with 1-Type Used	DISTCHAN12	-0.788	-3.58
>=10000 lbs	Distribution Channel with 3-Types Used	DISTCHAN32	0.759	3.05
<= 999 lbs	Service Industry	SIC11	5.84	5.77
<= 999 lbs	Transportation/Construction Industry	SIC21	0.975	3.57
>=10000 lbs	Transportation/Construction Industry	SIC23	2.88	9.9
Number of Observations	**Final Log Likelihood**	**Rho-squared**		
738	-554.922	0.316		

Table 23 shows the shipment size choice model results for manufactured products. The explanatory variables in this model are similar to those in the food products model. Shipments less than or equal to 999 lbs seem to be the most preferable everything else being equal. Here, longer trip lengths seem to be associated with shipment sizes less than or equal to 999 lbs.

Table 22. Shipment Size Model Specification for Manufactured Products

Choices	Utility Equations			
<= 999 lbs	ASC_V1 * one + cost1 * cost + SIC11 * SIC1 + SIC31 * SIC3			
1000-9999 lbs	ASC_V2 * one + SIC32 * SIC3 + DISTCHAN12 * DISTCHAN			
>=10000 lbs	ASC_V3 * one + DISTCHAN33 * DISTCHAN_3			
Choices	**Variable Description**	**Variable Name**	**Coefficient**	**t-stat**
<= 999 lbs	Alternative Specific Constant	ASC_V1	0	
1000-9999 lbs	Alternative Specific Constant	ASC_V2	-0.107	-0.5
>=10000 lbs	Alternative Specific Constant	ASC_V3	-0.349	-1.63
<= 999 lbs	Trip Length	cost1	0.15	1.69
1000-9999 lbs	Distribution Channel with 1-Type Used	DISTCHAN12	-0.911	-3.65
>=10000 lbs	Distribution Channel with 3-Types Used	DISTCHAN33	-1.35	-2.77
<= 999 lbs	Service Industry	SIC11	2.27	4.16
<= 999 lbs	Manufacture/Retail/Wholesale/Mining Industry	SIC31	1.98	6.3
1000-9999 lbs	Manufacture/Retail/Wholesale/Mining Industry	SIC32	1.03	2.86
Number of Observations	**Final Log Likelihood**	**Rho-squared**		
552	-431.443	0.289		

Model Application and Results

Figure 18 shows a schematic of the shipment size and frequency model. The shipment size choice is simulated for all the buyer-supplier firm pairs using the estimated models. For commodities other than food and manufactured products, the manufactured goods model was applied. Figure 19 shows the model simulation results of shipment size categories compared to the distributions observed in the Texas dataset.

Figure 18. Shipment Size and Frequency Model Process

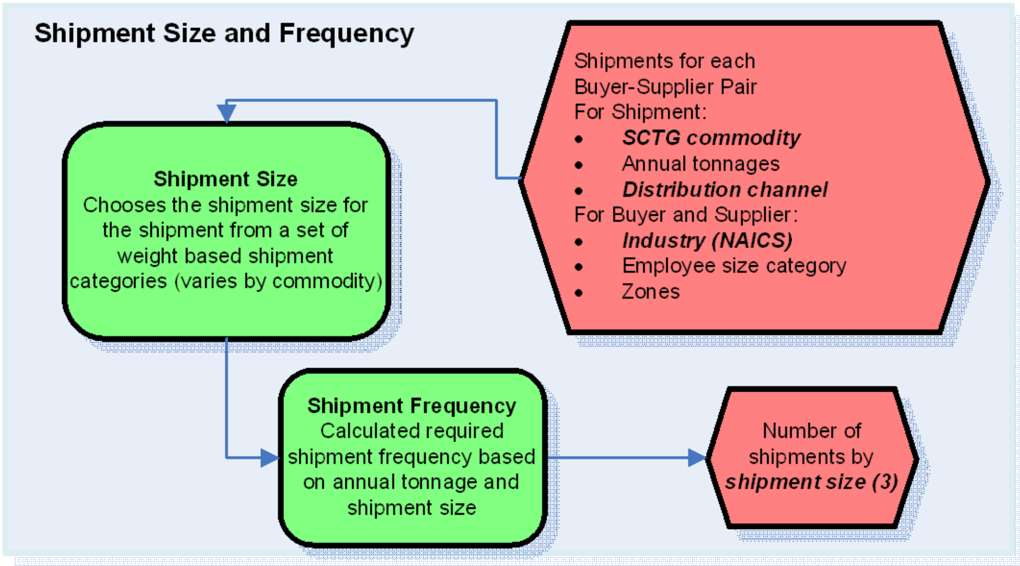


Figure 19. Percentage of Shipments by Size and Commodity Group

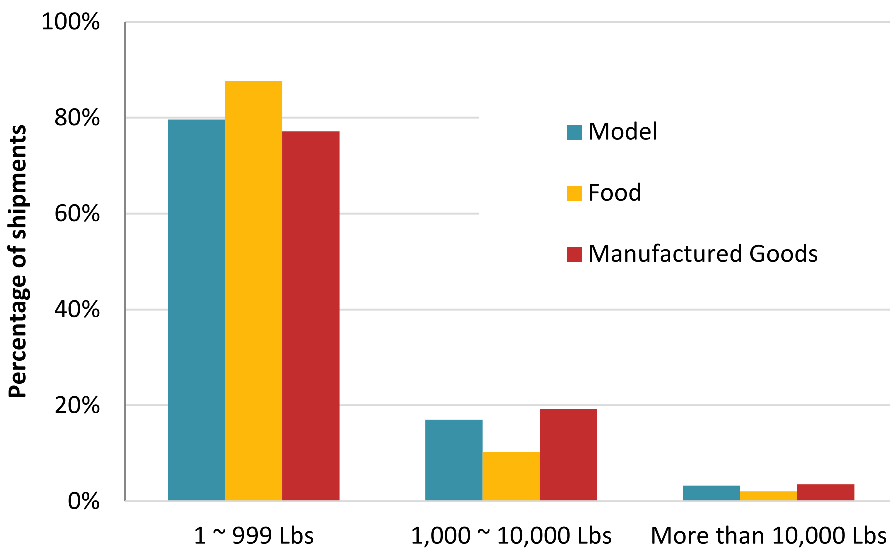
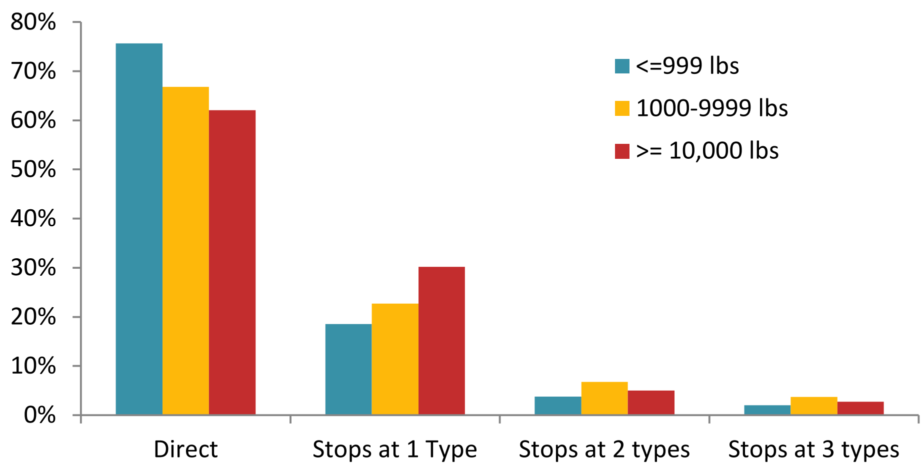


Table 24 shows the shipment size distribution by distribution channel. The percentage of these shipment sizes vary by size category, as shown in Figure 20, where direct shipments are much more likely to be smaller shipments.

Table 23. Shipments by Size and Distribution Channel

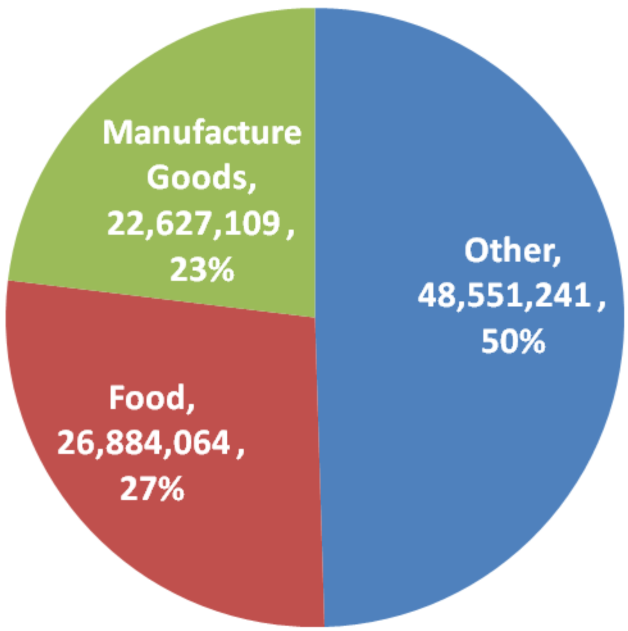
	<=999 lbs	1000-9999 lbs	>= 10,000 lbs
Direct	1,104,328	584,214	288,486
Stops at 1 Type	270,198	198,708	140,182
Stops at 2 types	55,269	59,263	23,376
Stops at 3 types	29,357	32,259	12,667

Figure 20. Percentage of Shipments by Size and Distribution Channel



Simulation of shipment size results in the assignment of a shipment to one of the three broad shipment categories. To obtain a more accurate shipment size each size category was split into 10 bins and probability thresholds for a shipment being in one of those 10 bins were calculated from Texas survey data. For example, the first shipment category was split into < 50, 50-150, 150-250, 250-350, 350-450, 450-550, 550-650, 650-750, 750-850, 850-950, 950-999 lbs bins and the probability of a shipment falling into each of these categories was computed from the Texas survey. All the shipments in the first modeled category (0-999 lbs) were then assigned to one of the finer shipment size categories using monte carlo simulation. The annual delivery frequency is then calculated using the annual commodity flow (in tons) and the individual shipment size for all the buyer-supplier firms. Figure 21 shows the distribution of the annual delivery frequencies by commodity.

Figure 21. Annual Delivery Frequency by Commodity Type



4.5 Production Mode and Intermodal Transfers

This step assigns a mode for shipments transported between each buyer-supplier pair. There are four primary modes (road, rail, air, and water) modeled. Air and water goods movements are typically a small portion of the overall goods movement in a region and so these approximations seem reasonable for the size of the market.

Data Sources and Model Development

A parallel study by CMAP (Cambridge Systematics, 2011) has adopted methods from de Jong and Ben-Akiva (2007) to predict the path and mode of long-haul movements of freight into, out of, within, and through the Chicago region. The path includes identifying the location of intermodal transfer facilities, distribution centers, or warehouses where shipments are consolidated or de-consolidated. Detailed networks of road and rail for the U.S. were used, but detailed networks of air and water were estimated using simple functions of distance and the value of goods being transported.

These models for mode choice and intermodal transfers have been adopted for the current demonstration project, based on the formulation developed by de Jong and Ben-Akiva (2007:

$$G_{mnql} = \beta_{0ql} + \beta_1 * \left(\frac{Q}{q}\right) + T_{mnql} + \beta_2 * j * v * Q + \beta_3 * t_{mnl} * j * v * \frac{Q}{365} + (\beta_4 + \beta_5 * v)\left(\frac{q}{2}\right) + a * \sqrt{LT * \sigma_Q^2 + Q^2 * \sigma_{LT}^2}$$

The descriptions of variables and parameter notations are provided in Table 25. A low (0.01), medium (0.05), or high (0.25) discount rates is used based on the type of commodity being transported. Bulk natural resources have a low discount rate. Animals and intermediate processed goods have a medium discount rate. A high discount rate is applied for finished goods.

Table 24. Mode Choice and Intermodal Transfer Model Parameters

Variable or Parameter	Description or Interpretation (of Parameters)	Value
G_{mnql}	Logistics cost between shipper m and receiver n with shipment size q and logistics chain l	Calculated in the mode choice model
Q	Annual flow in tons	From goods demand model
q	Shipment size in tons	From shipment size model
β_{0ql}	Alternative-specific constant	Asserted values based on commodity category
β_1	Constant unit per order	50
T	Transport and intermediate handling costs	From network skims
β_2	Discount rate	0.01/0.05/0.25 based on commodity
j	Fraction of shipment that is lost or damaged	0.01
v	Value of goods (per ton)	From FAF flow apportionment
β_3	Discount rate of goods in transit	0.01/0.05/0.25 based on commodity
t	Average transport time (days)	From network skims
β_4	Storage costs per unit per year	5,000
β_5	Discount rate of goods in storage	0.01/0.05/0.25 based on commodity
a	Constant used to set the safety stock in a way that generates a fixed probability of not running out of stock	0.50
LT	Expected lead time (time between ordering and replenishment)	10
σ_Q	Standard deviation in annual flow (i.e., anticipated variability in demand)	1.00
σ_{LT}	Standard deviation of lead time	1.00

Mode-path skims (times and costs) were developed for 54 alternatives in the meso-scale freight model study at CMAP (Table 26). These alternatives included direct modes (such as truck – Full truck load, truck – less than truck load, rail etc), indirect mode (such as rail-truck, water-truck etc), and also intermodal facilities (such as airports, truck terminals, rail terminals, and ports).

Table 25.Path Cost Parameters

Parameter	Description	Value
B_1	Constant unit per order	50.00
B_4	Coefficient for discount rate	5,000.00
j	Fraction of shipment that is lost or damaged	0.01
a	Safety stock constant	0.50

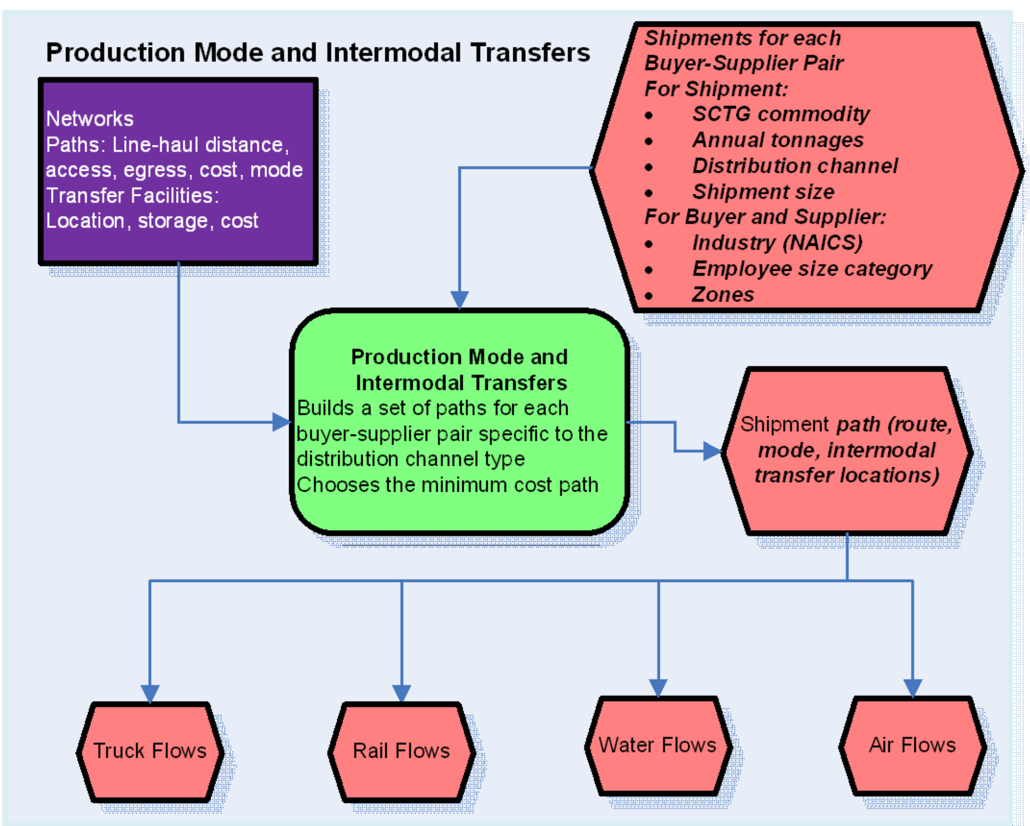
LT_OrderTime	Expected lead time (time between ordering and replenishment)	10.00
sdQ	Standard deviation in annual flow	1.00
sdLT	Standard deviation in lead time	1.00
LowDiscRate	Low-discount rate	0.01
MedDiscRate	Medium-discount rate	0.05
HighDiscRate	High-discount rate	0.25
CAP1FTL	Truckload capacity (tons)	30.00
CAP1Carload	Carload capacity (tons)	32.00
CAP1Airplane	Air cargo hold capacity (tons)	1.00

Source: Cambridge Systematics, 2011, Table 2.9, page 2-24.

Model Application and Results

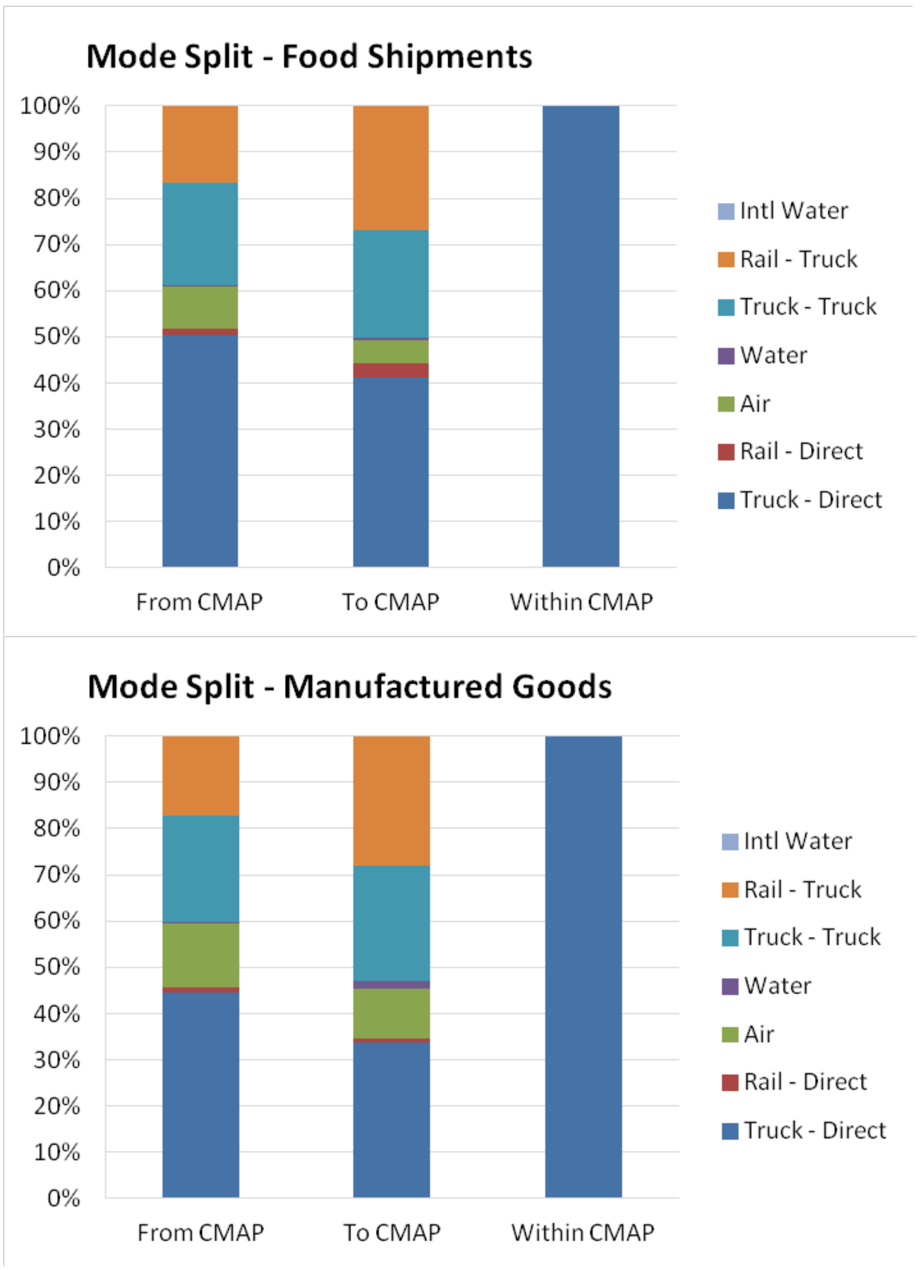
Figure 22 shows a schematic of the mode-path choice model. The buyer-supplier pairs dataset now has information on buyer firm ID, supplier firm type ID, commodity type (SCTG), annual flow in both tons and dollars, distribution channel,, and the shipment size. At this stage an actually business identified by randomly assigning a business from the pool of businesses that belong to the same firm type. Also, both the buyer and supplier mesozones are merged in from the output of the business/firm location model. The mode-path skims developed for all 54 mode-paths are merged next into the buyer-supplier pairs dataset.

Figure 22. Mode Choice and Intermodal Transfer Model Process



Based on the distribution channel the choice set of mode-path evaluated is restricted. For firms pairs using a direct distribution channel only direct mode paths are evaluated – Full Truck Load (FTL) Direct, Less than Truck Load (LTL) Direct, Carload Direct (rail), and IMX Direct (rail). The remaining 50 mode-path were evaluated for firms using indirect distribution channels. Once the generalized cost was evaluated for all the alternatives for each buyer-supplier pair, the least cost alternative was chosen as the mode-path. Mode splits of major modes by commodity are shown in Figure 23.

Figure 23. Mode Split for Food and Manufactured Products

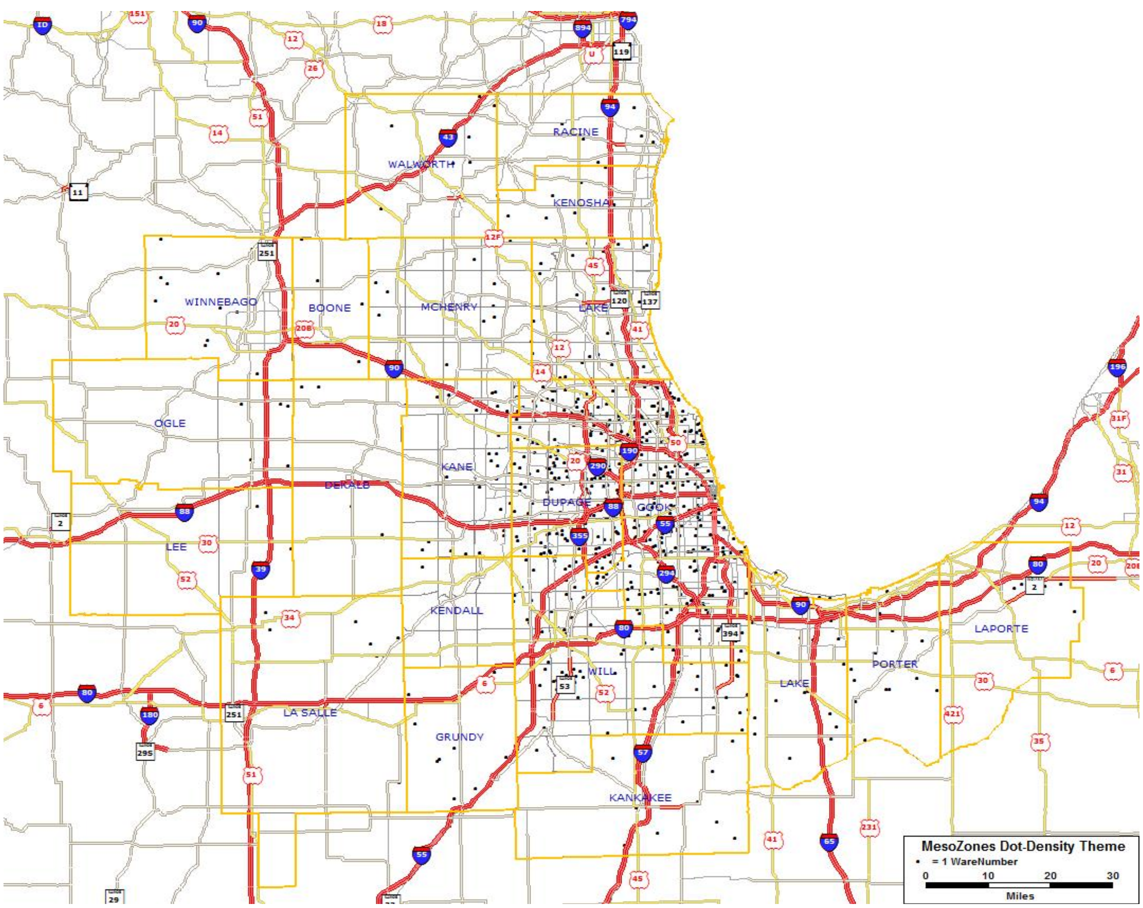


At this point, there is a transition from national level models to regional level models in the framework. The demand of goods and the mode-path is known for all the shipments. The next steps consist of loading the shipments on trucks and simulating truck trips in the region.

After the mode-path for each firm is selected, the daily frequency of delivery is determined. The annual tonnage between each firm pair is divided by the shipment weight to get the number of annual shipments and the number of annual shipments is divided by 310 to get daily frequency. Not all firm pairs have daily deliveries or pickups. A random number between 0 and 1 is generated and firm pairs for which the random number is less than the daily frequency are assumed to have deliveries on the day being simulated. This also ensures that firm pairs with a daily frequency greater than one have daily deliveries or pickups. From this point onwards, only the selected firm pairs and corresponding shipments are simulated to represent a typical day scenario.

After the mode-path for all shipments between buyer-supplier pairs are chosen, regional warehouses/intermodal facilities are assigned to the shipments which are using indirect distribution channels. These warehouses represent the first stop if a shipment is originating from CMAP region and the last stop (before final delivery) if a shipment is terminating in the CMAP region. Firms that could be potential warehouses were filtered out based on their NAICS industry code. The industries that were in air, rail, water transportation, warehousing and storage (NAICS 481-483, 493) were chosen as potential regional warehouse locations. Figure 24 shows the locations of the firms chosen as warehouses in the CMAP region.

Figure 24. Warehouses and Intermodal Facilities in the CMAP Region



A local warehouse (within CMAP region) is randomly selected for an indirect distribution channel shipment. Based on the mesozone of the warehouse and whether or not a shipment is incoming or outgoing, the origin and destination mesozones of all the shipments that are to be simulated in a day are determined. Based on the mesozone of the warehouse and whether or not a shipment is incoming or outgoing, the origin and destination mesozones of all the shipments that are to be simulated in a day are determined (Table 27).

Table 26. Daily Shipment Origin and Destination Mesozones

	Origin	Destination
Buyer in Chicago, Direct	Seller mesozone	Buyer mesozone
Buyer in Chicago, Indirect	Warehouse mesozone	Buyer mesozone
Buyer outside Chicago, Direct	Seller mesozone	Buyer mesozone
Buyer outside Chicago, Indirect	Warehouse mesozone	Seller mesozone

If distribution channel is direct, the origin mesozone is the mesozone of the seller, if it is not direct, it is the warehouse mesozone (in case of a drop-off a truck will start from the warehouse with a shipment and in case of a pick-up, truck will start from the warehouse without a shipment). If the distribution channel is indirect and the buyer is outside Chicago, the destination mesozone is the seller mesozone (to simulate a pickup), if not, it is the buyer mesozone.

4.6 Vehicle Choice and Tour Pattern

As mentioned earlier, this step marks the beginning of the regional component of the freight model system. At this stage, both the number of shipments from warehouse facilities to various buyer firms and the number of shipments from various suppliers to warehouse facilities are known. In this step, the choices of vehicle and tour pattern are simulated for each of the warehouse facilities. This model applies only to trucks.

Data Sources and Model Development

Due to the inter-relationship that exists between the choices of vehicle or truck types and tour patterns, we jointly simulated the type of truck and the tour pattern for each of the shipments. This also circumvents the complexity of having to estimate two separate models – one each for tour pattern and vehicle choice. Again, the Texas truck survey dataset was the most useful for model estimation, due to its larger sample size. The Texas Commercial Vehicle Survey Datahave been used earlier for developing tour-based truck models (Ruan et al, 2011). Daily patterns of trucks were mined from this data. These daily patterns included single or multiple direct tours (with a single destination), single or multiple peddle tours (with multiple destinations), or a mixture of these options. An MNL model was estimated with six alternatives which are combinations of 2 tour patterns (direct and peddling) and three vehicle types (2 axle, 3-4 axle, and semi/trailer).

The results are shown in Table 28. Just from the constants, it appears that two-axle trucks are more likely to be chosen when compared to the heavier trucks and peddling tours are more likely than direct tour. The type of commodity (food/manufactured), pick-up/drop-off weights (shipment sizes), and the type of industry at the stop location (delivery/buyer location) of the shipment along with county total employment (at the drop-off or pick-up location) are some of the other explanatory variables in the specification. Food shipments tend to be on 2-axle trucks and peddling tours. Manufactured goods are less likely to use peddling tours and larger trucks (3-4 axle and

Table 27. Vehicle Choice and Tour Pattern Model Specifications

Choices	Variable Description	Variable Name	Coefficient	t-stat
Direct, 4 Tires	Alternative Specific Constant	ASC_V01	0	
Direct, 6-8 Tires	Alternative Specific Constant	ASC_V02	-4.5	-7.25
Direct, Semi/Trailer	Alternative Specific Constant	ASC_V03	-4.41	-7.55
Paddling, 4 Tires	Alternative Specific Constant	ASC_V04	3.89	11.32
Paddling, 6-8 Tires	Alternative Specific Constant	ASC_V05	-1.29	-3.07
Paddling, Semi/Trlr	Alternative Specific Constant	ASC_V06	-2.94	-6.91
Paddling, 4 Tires	Cargo is Food Products	CARGO_FOOD_V04	1.21	9.95
Paddling, 6-8 Tires	Cargo is Manufactured Products	CARGO_MANU_V05	-1.21	-8.86
Paddling, Semi/Trlr	Cargo is Manufactured Products	CARGO_MANU_V06	-0.294	-2.71
Direct, 6-8 Tires	Cargo Weight at Dropoff (lbs)	Cargo_Weight_DO_V02	0.412	6.16
Direct, Semi/Trailer	Cargo Weight at Dropoff (lbs)	Cargo_Weight_DO_V03	0.371	5.91
Paddling, 4 Tires	Cargo Weight at Dropoff (lbs)	Cargo_Weight_DO_V04	-0.209	-4.89
Paddling, 6-8 Tires	Cargo Weight at Dropoff (lbs)	Cargo_Weight_DO_V05	0.283	6.45
Paddling, Semi/Trlr	Cargo Weight at Dropoff (lbs)	Cargo_Weight_DO_V06	0.263	6.03
Direct, 6-8 Tires	Cargo Weight at Pickup (lbs)	Cargo_Weight_PU_V02	0.355	5.59
Direct, Semi/Trailer	Cargo Weight at Pickup (lbs)	Cargo_Weight_PU_V03	0.401	6.64
Paddling, 4 Tires	Cargo Weight at Pickup (lbs)	Cargo_Weight_PU_V04	-0.156	-3.7
Paddling, 6-8 Tires	Cargo Weight at Pickup (lbs)	Cargo_Weight_PU_V05	0.282	6.53
Paddling, Semi/Trlr	Cargo Weight at Pickup (lbs)	Cargo_Weight_PU_V06	0.263	6.1
Direct, Semi/Trailer	Destination Industry is Manufact.	MANUFAC_V03	1.19	3.53
Paddling, 6-8 Tires	Destination Industry is Manufact.	MANUFAC_V05	1.41	11.97
Paddling, Semi/Trlr	Destination Industry is Manufact.	MANUFAC_V06	1.34	11.24
Paddling, 6-8 Tires	Destination Industry is Office	OFFICE_V05	-0.834	-2.66
Paddling, Semi/Trlr	Destination Industry is Retail	RETAIL_V06	0.928	9.1
Paddling, 6-8 Tires	County Total Employment	cbp98EMP_V05	0.137	7.47
Paddling, Semi/Trlr	County Total Employment	cbp98EMP_V06	0.309	16.55

Number of Obs.	Final Log Likelihood	Rho-squared
5,314	-5,765.731	0.392

semi/trailer). An intuitive result is that the trucks get larger with heavier shipments (look at the coefficient of Pickup and Drop-off weights). Also, heavier shipments are more likely to be put on direct tours than peddling tours. Delivery to an office location is less likely to be on peddling tours and 3-4 axle trucks. Delivery to a retail location is more likely to be on peddling tours and semi/trailer trucks.

Model Application and Results

Figure 25 shows a schematic of the vehicle choice and tour pattern model. Vehicle choice and tour patterns are simulated for all the daily shipments using the estimated model. Shipments using a direct distribution channel are forced to have a direct tour pattern keeping the vehicle type choice same as that output by the simulation\prediction. Figure 26 shows the distribution of vehicle type and tour patterns for shipments by commodity type.

Figure 25. Vehicle Choice and Tour Pattern Model Process

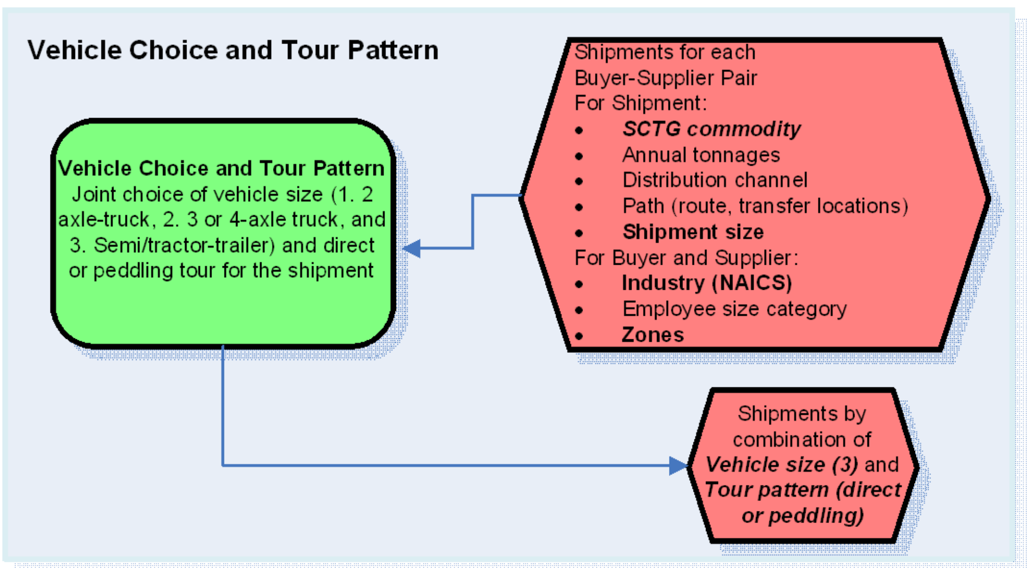
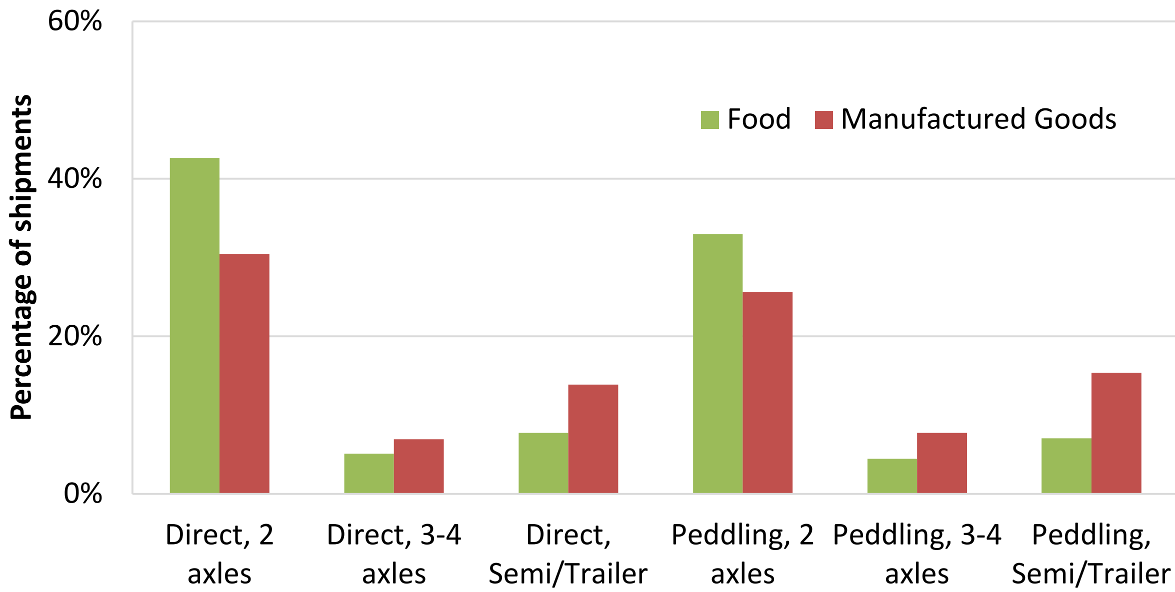


Figure 26. Percentage of Shipments by Vehicle Type and Tour Pattern



4.7 Number of Tours Choice

The purpose of this step is to group the various delivery/pick-up locations of shipments from/to a warehouse into truck tours. This is applicable only to indirect distribution channel shipments and is achieved in two steps. First, the category of number of truck tours for each shipment is determined. Each shipment record now represents a stop to be made on a truck tour. The stops for the indirect shipments are then clustered according to the number of truck tours category.

Data Sources and Model Development

The Texas dataset was used in the estimation of the number of truck tours model. Instead of modeling the number of stops in a particular tour, the object is to predict for each stop a number of tours category. There are four alternatives for the category of number of truck tours – all (stops) in one tour, two tours, three tours, and four tours. All in one tour means the shipment belongs to a tour category in which a truck covers all the stops assigned to it in a single tour. The purpose of this is to be able to determine the number of tours for all the stops serviced from a particular warehouse.

Table 29 shows the result of the MNL model estimated for predicting the number tours category for each indirect shipment. All being equal, the most cost effective method - all stops in one tour - is most preferred. The explanatory variables in the model specification are primarily stop types (or the industry type at the stop / buyer location) and commodity types. Delivery/pick-up of food is more likely occur by two or three separate trucks tours whereas delivery to a construction firm is less likely to use multiple truck tours. Deliveries to distribution centers are more likely to be routed through one or two truck tours. Also, as the weight of the shipment increases, it is more likely that it is a part of a multi-tour truck.

Table 28. Number of Tours Model Specifications

Choices	Variable Description	Variable Name	Coefficient	t-stat
All stops in one route	Alternative Specific Constant	ASC_V1	4.52	16.29
All stops in two routes	Alternative Specific Constant	ASC_V2	3.31	11.83
All stops in three routes	Alternative Specific Constant	ASC_V3	2.14	6.65
All stops in four routes	Alternative Specific Constant	ASC_V4	0	
All stops in two routes	Cargo is Food Products	CARGO_5_V02	0.672	4.65
All stops in three routes	Cargo is Food Products	CARGO_5_V03	1.06	6.13
All stops in two routes	Destination Industry is Construction	CONST_V02	-0.787	-6.04
All stops in three routes	Destination Industry is Construction	CONST_V03	-0.473	-3.66
All stops in one route	Destination Industry is Distribution Center	DISTCN_V01	2.09	7.33
All stops in two routes	Destination Industry is Distribution Center	DISTCN_V02	1.49	4.47
All stops in one route	Cargo Weight at Pickup or Dropoff (lbs)	Duty_V01	-0.375	-13.17
All stops in two routes	Cargo Weight at Pickup or Dropoff (lbs)	Duty_V02	-0.375	-13.17
All stops in three routes	Cargo Weight at Pickup or Dropoff (lbs)	Duty_V03	-0.254	-7.35
All stops in one route	Destination Industry is Manufacturing	MAC_V01	1.22	8.96
All stops in two routes	Destination Industry is Manufacturing	MAC_V02	0.713	3.99
All stops in one route	Destination Industry is Office	OFFICE_V01	1.71	3.28
All stops in two routes	Destination Industry is Office	OFFICE_V02	1.78	3.2
All stops in one route	Destination Industry is Retail	RETAIL_V01	1.57	10.34
All stops in three routes	Destination Industry is Retail	RETAIL_V03	-0.749	-2.5
All stops in one route	Destination Industry is Warehouse	WARE_V01	0.958	5.6

Number of Obs.	Final Log Likelihood	Rho-squared
4,326	-3,973.536	0.334

Model Application and Results

Figure 27 shows a schematic of the number of tours and stops model. The number of truck tours category is simulated for each indirect shipment using the estimated model. Figure 28 shows the distribution of the simulated number of truck tours categories by shipment size.

Figure 27. Number of Tours and Stops Model Process

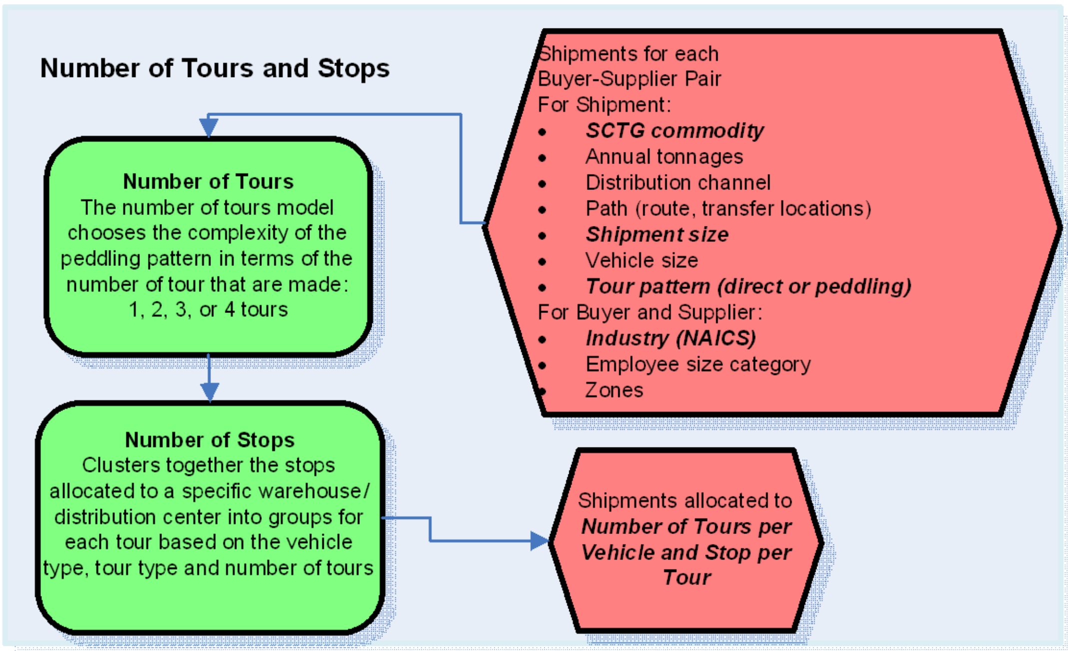
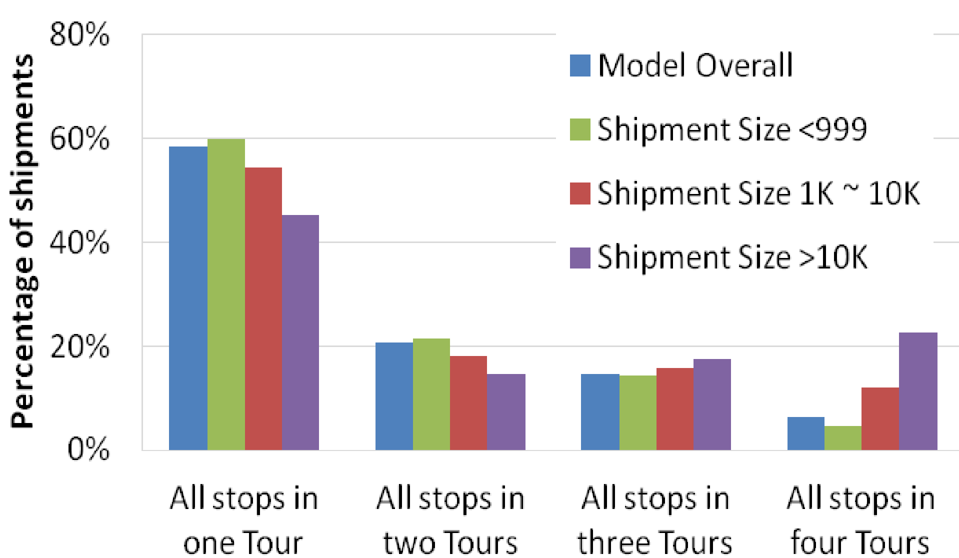


Figure 28. Shipments by Number of Tours and Shipment Size



The shipments stops are now clustered based on the number of truck tours category predicted. This is done only for indirect shipments since direct shipments are serviced by a direct tour with a single stop. The count of number of stops by warehouse, vehicle type, and number of tours category is obtained. It is ensured that the count of shipment stops is at least as much as the number of tours in the corresponding category. For example, if a particular shipment falls in a 4 tour category, there should be at least 3 more shipments that are in the 4 tour category assigned to the same warehouse for consistency. If that is not the case, the number of tours category is modified to be consistent. Next, shipment stops are clustered based on the number of tours category and vehicle type. For example, if a warehouse has 5 indirect shipment stops assigned to it and all of them fall in the two tours category using a 2 axle truck, the clustering function would cluster the five stops in to two clusters using Euclidean method. Figure 29 shows the distribution of the number of stops clustered into all the tours from the warehouses.

Figure 29. Number of Stops per Tour from Warehouses

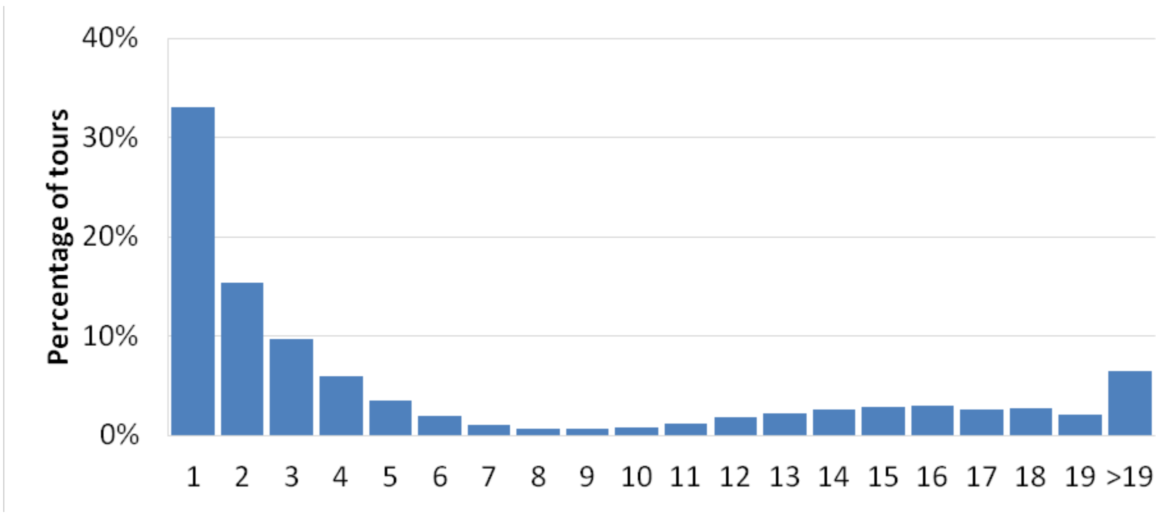
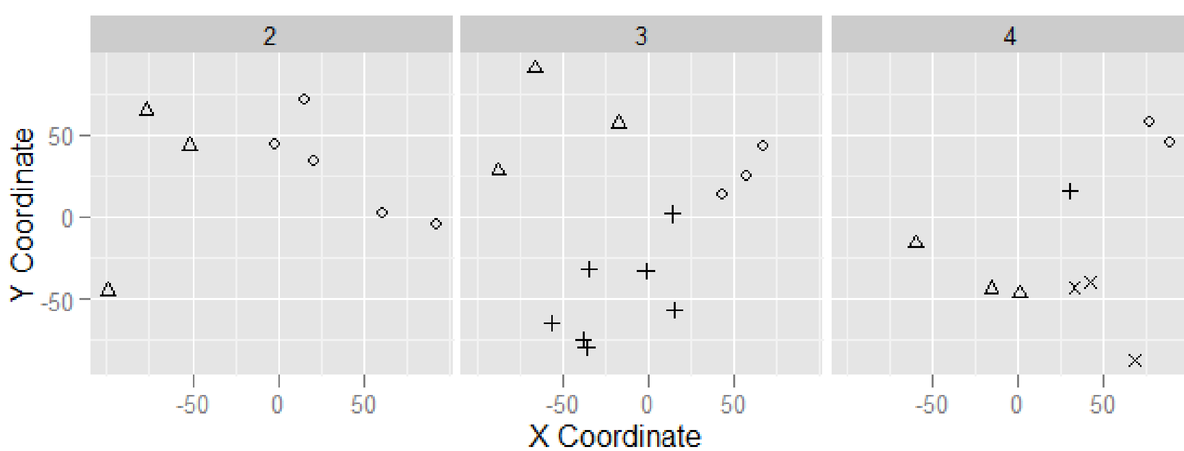


Figure 30 shows an example of how shipment stops from a warehouse are clustered. The stop which are in 2 tours category are clustered into two groups or tours, those in 3 tours category are clustered into 3 tours, and those in 4 tours category are clustered into 4 tours.

Figure 30. Clustered Stops for 2, 3 and 4 Tour Trucks



4.8 Stop Sequence and Duration

In this step, the shipment stops that had been clustered into individual tours in the previous step are sequenced and the duration of the drop-off/pick-up at each of those stops is simulated.

Data Sources and Model Development

For sequencing all the stops within a tour, a greedy algorithm is applied for lack of a better dataset to model (subsequent) stop location. The process assigns the first stop as the one that is closest to the base (warehouse). It then keeps adding the next closest stop from the previous stop to the trip sequence until all stops in a particular tour are assigned a sequence.

A discrete choice model was estimated for stop duration using the Texas survey data (Table 30). There are 6 alternatives for stop duration – 15 minutes or less, 15-30 minutes, 30-45 minutes, 45-60 minutes, 60-75 minutes, and over 75 minutes. The results from this model are fairly intuitive. If the shipment weight is higher it is likely to increase the duration of the stop. It there are more stops involved in the tour, it is less likely that the trucks will stop for a long time. If the vehicle is larger, it will take longer to unload and the stop will be longer. Food products have longer stops than manufactured products. Direct tour patterns have longer stops and retail stops tend to have stops that are 60-75 minutes long.

Model Application and Results

Figure 31 shows a schematic of the stop sequence and duration models. The greedy algorithm for sequencing shipment stops is only applied for indirect shipments whereas the stop duration model is used to simulate the stop durations of both direct and indirect shipments. Figure 32 shows an example of how stop sequencing algorithm sequences the stops in two separate tours from a warehouse.

Figure 31. Stop Sequence and Duration Model Process

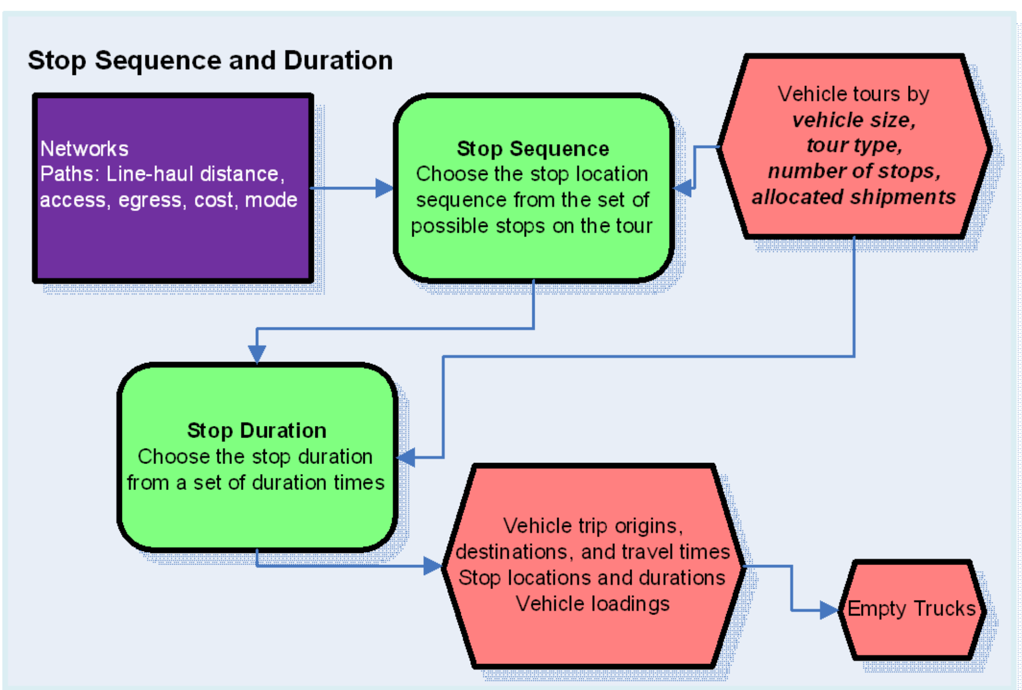
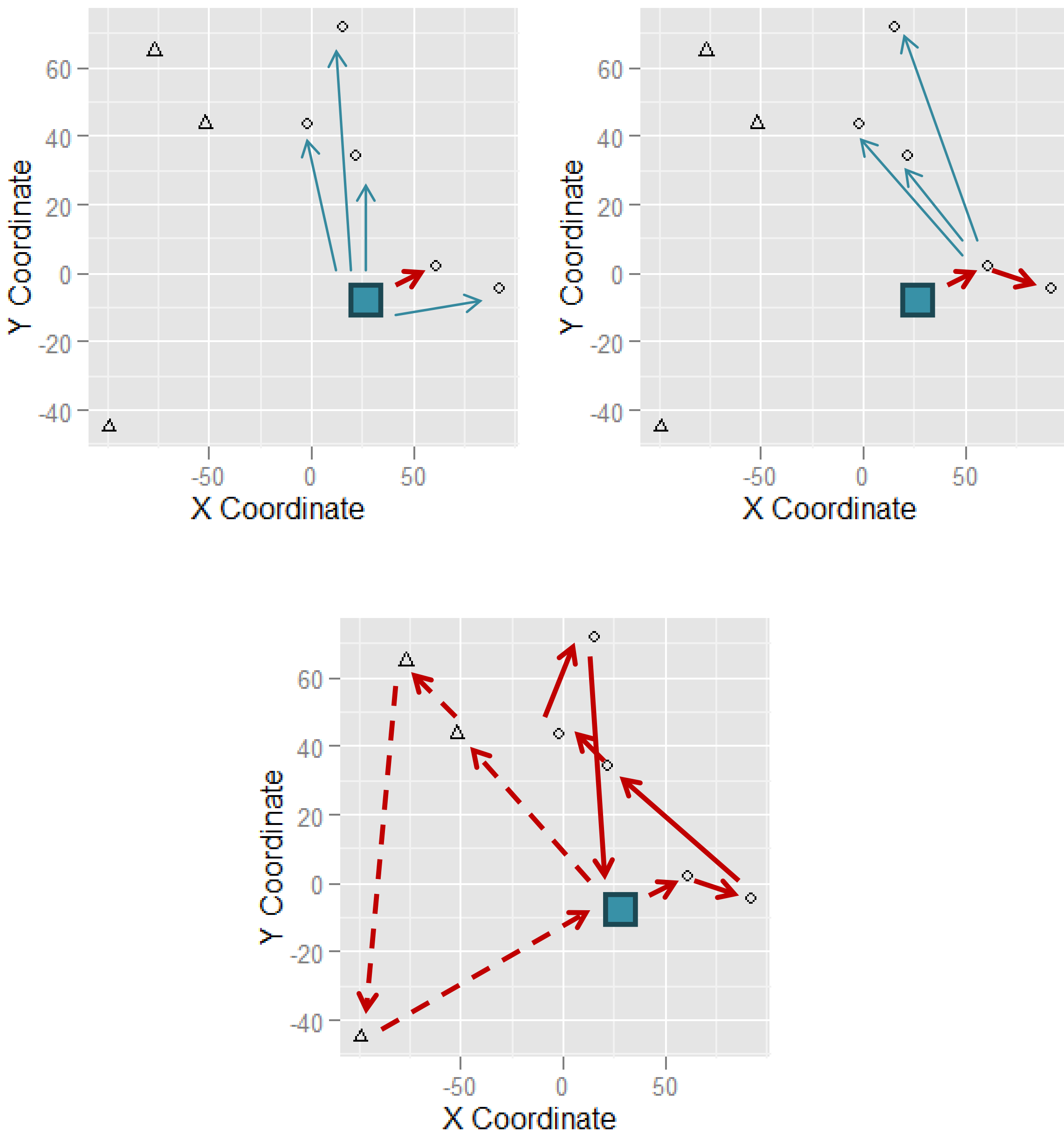


Table 29. Number of Tours Model Specifications

Choices	Variable Description	Variable Name	Coefficient	t-stat
<=15 mins	Alternative Specific Constant	ASC_V01	0	
15 ~ 30 mins	Alternative Specific Constant	ASC_V02	1.19	8.05
30 ~ 45 mins	Alternative Specific Constant	ASC_V03	1.9	9.99
45 ~ 60 mins	Alternative Specific Constant	ASC_V04	2.69	13.11
60 ~ 75 mins	Alternative Specific Constant	ASC_V05	2.43	10.48
>75 mins	Alternative Specific Constant	ASC_V06	4.4	22.55
30 ~ 45 mins	Cargo is Manufact. Products	Cargo14_v03	-0.191	-1.69
>75 mins	Cargo is Manufact. Products	Cargo14_v06	-0.219	-1.89
>75 mins	Cargo is Food Products	Cargo5_v06	-0.336	-2.32
15 ~ 30 mins	Pickup/Dropoff Weight (Log Lbs)	Duty_V02	0.0505	3.93
30 ~ 45 mins	Pickup or Dropoff Weight (Log Lbs)	Duty_V03	0.0951	6.58
45 ~ 60 mins	Pickup or Dropoff Weight (Log Lbs)	Duty_V04	0.1	6.4
60 ~ 75 mins	Pickup or Dropoff Weight (Log Lbs)	Duty_V05	0.111	6.24
>75 mins	Pickup or Dropoff Weight (Log Lbs)	Duty_V06	0.0804	5.49
60 ~ 75 mins	Stop is Retail	RETAIL_V05	0.269	2.07
>75 mins	Tour Pattern is Direct	RP_Direct_V06	1.18	5.37
15 ~ 30 mins	Number of Stops on Route	Tot_stops_V02	-0.0463	-4.75
30 ~ 45 mins	Number of Stops on Route	Tot_stops_V03	-0.14	-13.04
45 ~ 60 mins	Number of Stops on Route	Tot_stops_V04	-0.241	-19.31
60 ~ 75 mins	Number of Stops on Route	Tot_stops_V05	-0.281	-19.32
>75 mins	Number of Stops on Route	Tot_stops_V06	-0.409	-29.17
30 ~ 45 mins	Vehicle is 2-Axle Truck	veh_class1_v03	-0.346	-3.24
45 ~ 60 mins	Vehicle is 2-Axle Truck	veh_class1_v04	-0.758	-6.29
60 ~ 75 mins	Vehicle is 2-Axle Truck	veh_class1_v05	-0.695	-4.98
>75 mins	Vehicle is 2-Axle Truck	veh_class1_v06	-0.699	-6.1
15 ~ 30 mins	Vehicle is 3-4 Axle Truck	veh_class2_v02	-0.388	-3.33
30 ~ 45 mins	Vehicle is 3-4 Axle Truck	veh_class2_v03	-0.514	-3.86
45 ~ 60 mins	Vehicle is 3-4 Axle Truck	veh_class2_v04	-0.903	-6.24
60 ~ 75 mins	Vehicle is 3-4 Axle Truck	veh_class2_v05	-0.857	-5.37
>75 mins	Vehicle is 3-4 Axle Truck	veh_class2_v06	-0.871	-6.18

Number of Obs.	Final Log Likelihood	Rho-squared
5,723	-8,776.958	0.141

Figure 32. Stop Sequencing Algorithm



Figures 33 and 34 show the distribution of tour lengths as a result of the stop sequencing algorithm compared to the actual distribution of tour lengths observed from the Texas survey data for all tours and tours with at least 3 stops, respectively. The greedy algorithm does a reasonable job of matching observed tour length distributions.

Figure 33. Observed and Estimated Tour Length Distribution for All Tours

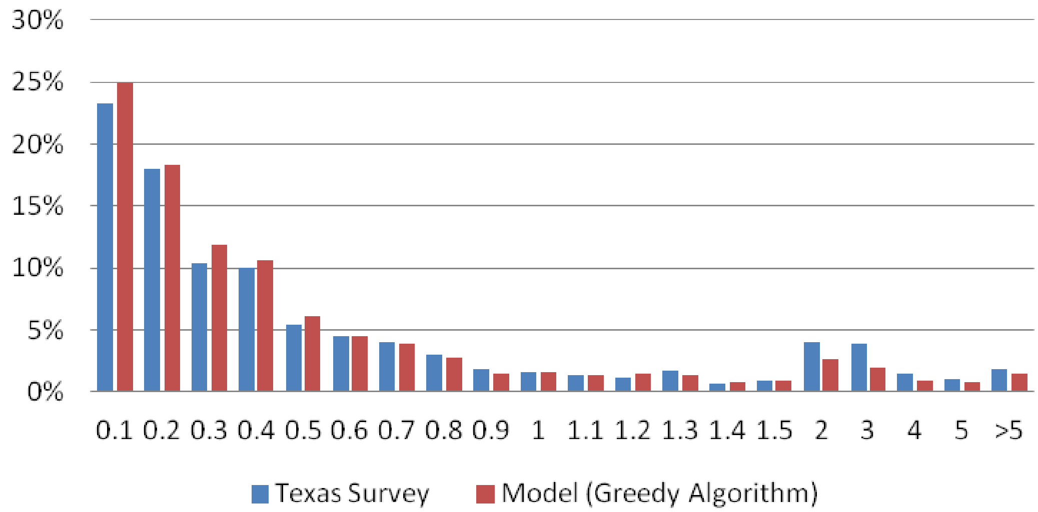


Figure 34. Observed and Estimated Tour Length Distribution for Tours with at least 3 Stops

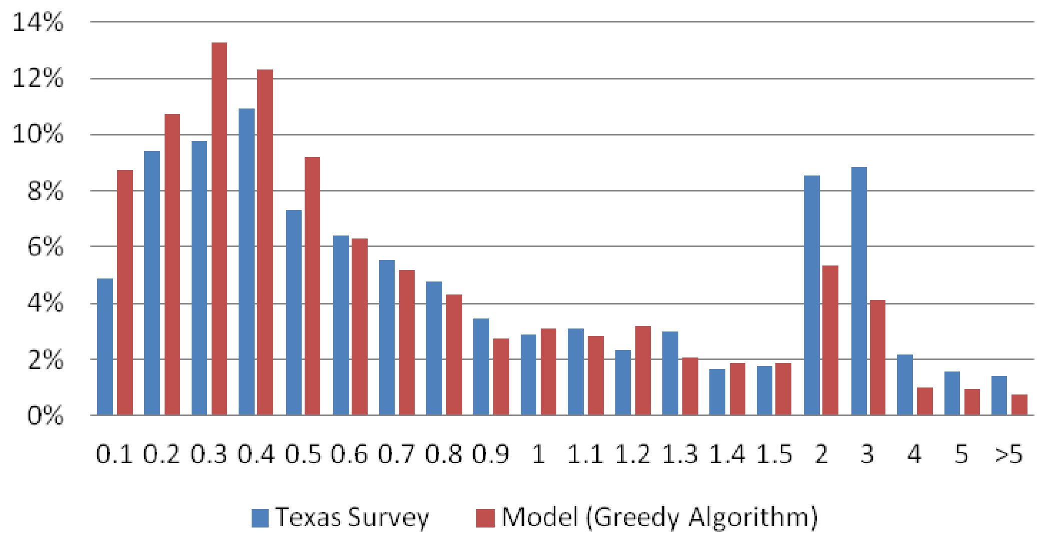
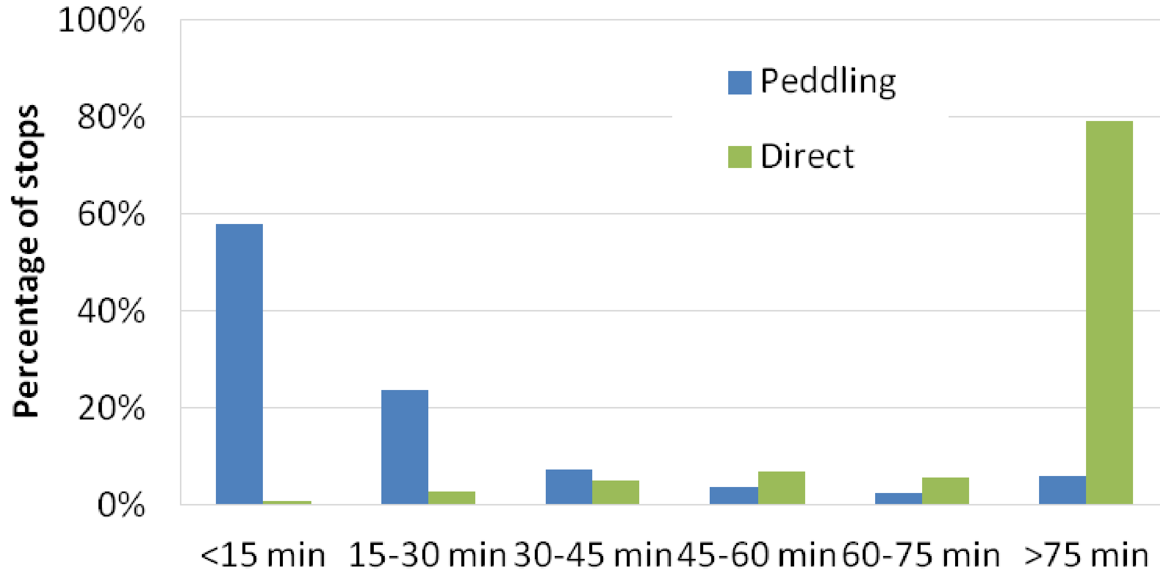


Figure 35 shows the distribution of stop durations by tour pattern (direct or peddling). As expected from the estimated model, stop durations on peddling tours tend to be shorter than those on direct tours.

Figure 35. Distribution of Stop Durations by Tour Pattern



4.9 Tour Time of Day

This step simulates the tour start times of all tours from the warehouses. The tour start times and origin and destination mesozones could be used to develop truck trip tables that can then be assigned with passenger trip tables in a multiclass assignment in a travel model.

Data Sources and Model Development

The Texas survey data are used for estimating a tour time of day model. There are 5 alternatives time periods modeled - before 6AM, 6-8 AM, 8-9 AM, 9-10 AM, and after 10 AM. Table 31 shows the results of the estimated model. The time of departure is influenced by the total load on particular tour, the total tour length and the dwell time. As the total load for a tour increases, the probability of a truck tour starting earlier in the day increases. The tour start time from the base location is more likely to be earlier in the day as the dwell time and tour length increase.

Table 30. Tour TIme of Day Model Specifications

Choices	Variable Description	Variable Name	Coefficient	t-stat
12:00AM -6:00AM	Alternative Specific Constant	ASC_V1	-5.97	-9.13
6:00AM - 8:00AM	Alternative Specific Constant	ASC_V2	-3.13	-6.04
8:00AM - 9:00AM	Alternative Specific Constant	ASC_V3	-2.05	-4.47
9:00AM -10:00AM	Alternative Specific Constant	ASC_V4	-1.6	-2.62
10:00AM - 11:59PM	Alternative Specific Constant	ASC_V5	0	
12:00AM -6:00AM	Total load (log lbs)	Duty_V1	0.0454	1.9
6:00AM - 8:00AM	Total load (log lbs)	Duty_V2	0.0484	3.05
12:00AM -6:00AM	Total Dwell Time	Dw1	0.429	5.28
6:00AM - 8:00AM	Total Dwell Time	Dw2	0.429	5.28
8:00AM - 9:00AM	Total Dwell Time	Dw3	0.429	5.28
9:00AM -10:00AM	Total Dwell Time	Dw4	0.226	2.08
12:00AM -6:00AM	Total Tour length (log miles)	TL_V1	0.706	6.41
6:00AM - 8:00AM	Total Tour length (log miles)	TL_V2	0.282	3.99
Number of Observations	**Final Log Likelihood**	**Rho-squared**		
1,046	-1,529.711	0.084		

Model Application and Results

Figure 36 shows a schematic of the tour time of day model. The tour start time period is simulated using the estimated model for both direct and indirect shipments. The actual start time is assigned as 5:00 AM, 7:00 AM, 8:30 AM, 9:30 AM, or 10:30 AM based on the alternative predicted for a tour. The tour start time represents the start time of the first trip in the tour. Since the skims are available and the stop durations are known, trip start times of subsequent trips are calculated in a straight forward manner. The start times for each of the individual trips can be calculated by adding the travel duration to current stop and stop duration at the current stop to the start time of the previous trip. Figure 37 shows the distribution of tour start times by tour pattern. As expected, a higher percentage of direct tours start later in the day whereas a higher percentage of peddling tours start earlier in the day.

Figure 36. Delivery TIme of Day Model Process

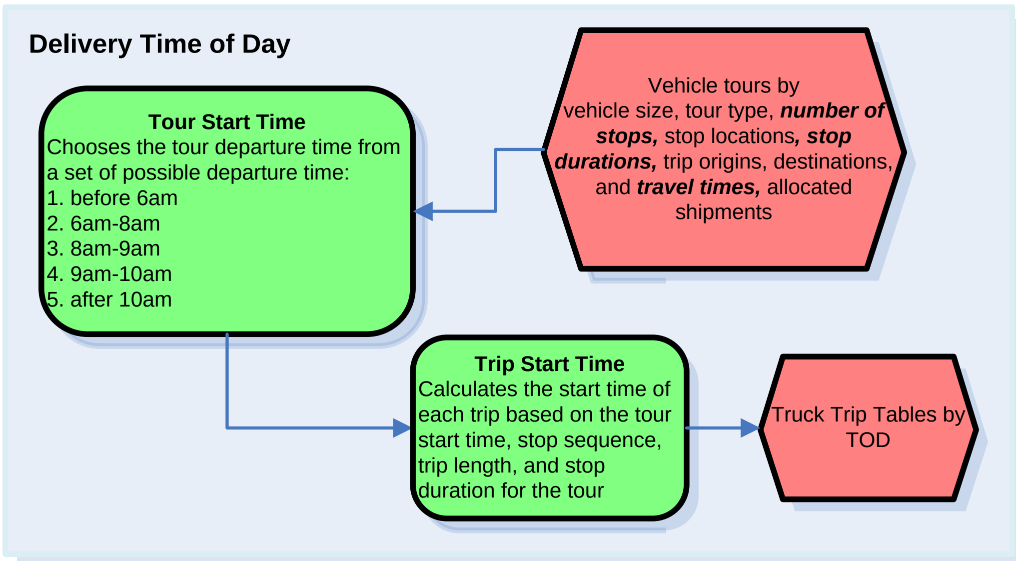
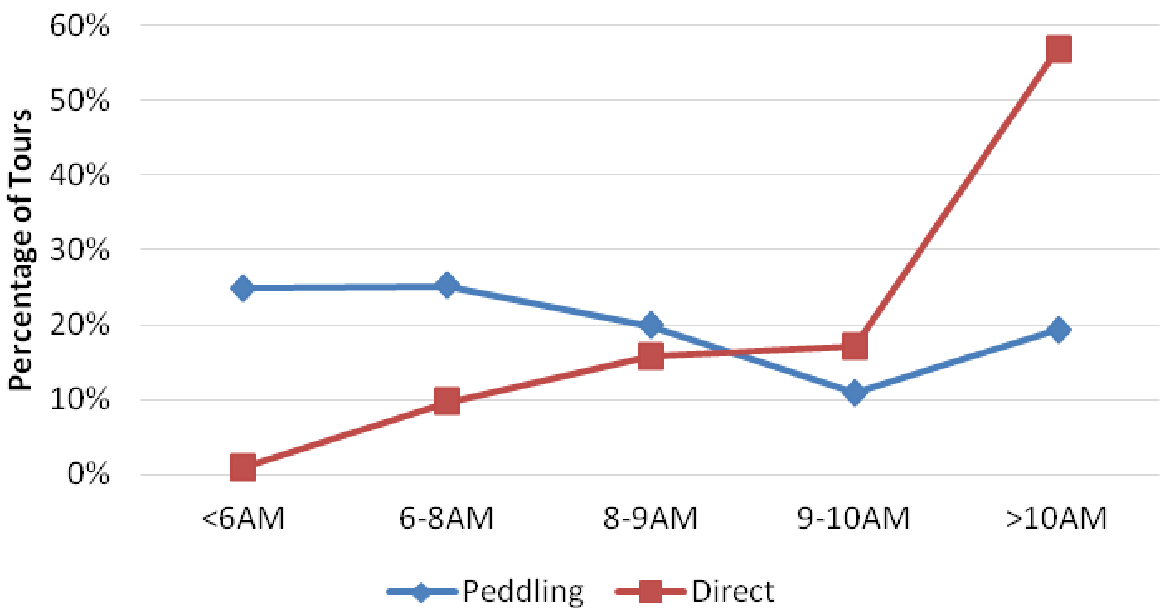


Figure 37. Tour Start Time by Tour Pattern Type



5.0 Data Collection Plan

5.1 Introduction

This data development plan addresses the data requirements for implementing and using the advanced freight forecasting models described in this report. In developing this plan we take the view that, in light of the challenges in collecting freight and commercial transport data, and the national and international scale of some freight transport and the data needed to model it, substantial responsibility for collecting the needed data should fall to the federal government. Furthermore, data are expensive and slow to develop, and should also be systematically maintained, enhanced and archived.

The scope focuses on models that predict surface transport required for the delivery of goods. It includes the models of demand for transport and models of traffic flows on the transport networks, to the extent that they must be different than regional traffic models used for personal transport. It focuses on enabling estimates of traffic flows within a region, but includes flows at a national and international scale because flows at that level affect flows within any given region.

Data are required for model development, calibration and validation, and prediction, as shown in Table 32. The next section of this report identifies the data needed for model development, first for the nation-level model components and then for the regional components. It is followed by sections on calibration data and data needed for forecasting. The final section provides recommendations for further research and development.

Table 31. Major types of data and their use for modeling

Types of data	Modeling Uses		
	Model Development	Calibration and validation	Prediction
Spatial data	✓ (survey years)	✓ (calibration and validation years)	✓ (projections)
Network data	✓ (survey years)	✓ (calibration and validation years)	✓ (projections)
Economic data	✓ (survey years)	✓ (calibration and validation years)	✓ (projections)
Surveys	✓	✓	
Traffic data		✓	

5.2 Data Needed for Model Development

National Model Components

The national model components include firm synthesis, supplier selection, distribution channels, production mode & intermodal transfers, and shipment size & frequency. Table 33 provides an overview of the types of data needed by each of these models. It is followed by descriptions of the data types.

Table 32. Usage of the Data Types by the National Model Components

Model component	Spatial data	Network data	Economic data	Survey data
Firm synthesis	✓			
Supplier selection	✓	✓	✓	✓*
Distribution channels	✓	✓*		✓
Production mode and intermodal transfers	✓	✓		✓
Shipment size and frequency	✓	✓		✓

* Needed but not in current model

Spatial data

Zone system. The national model components predict flows among zones partitioning the entire nation and representing international ports. The zone system maintained by the US government for the Freight Analysis Framework (FAF)[2] serves as the basis for the national model components, but the FAF zones need to be subdivided to the level of county or smaller within the study region, and perhaps also within certain other FAF zones that may be of competitive interest to the study region because of their special impact on flows into, out of and through the study region.

Zonal attributes. For each zone in the national zone system, it is necessary to estimate employment and payroll by industry category, using the current standard NAICS classification system[3]. These data are used for firm synthesis, and also to identify the presence and size of distribution centers, consolidation centers, warehouses and intermodal transfer facilities for each commodity group within the zone (data on spatial scale (i.e. acreage or square footage) of these centers could be used to advantage in the national models if it were available.) The primary source of the employment and payroll data is the national **County Business Patterns (CBP)** data[4]. CBP data are produced once per year by the U.S. Census Bureau from administrative records and survey data. Known weaknesses of CBP data are the exclusion of agriculture, construction and most government employment, as well as inaccurate industry classification and location of leased workers, which CBP associates with the leasing company rather than the client for whom the workers actually work. To help overcome the weaknesses of CBP, it can be augmented with data from public sources (state Employment Security Data[5] (ESD)) and private sources (e.g., InfoUSA[6]).

[2] see http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/

[3] see http://www.census.gov/eos/www/naics/

[4] see http://www.census.gov/econ/cbp/

[5] see, for example, http://www.ides.illinois.gov/

[6] see http://www.infousa.com/

Network data

National freight network representation. The national model components use a composite network representation of freight paths. It includes links for all major modes (air, road, rail and water); nodes for all major transfer facilities (intermodal, distribution centers, warehouses and consolidation centers) within the study region; zonal centroids for access and egress; impedance for traversal of links and nodes; impedance for access and egress at centroid connectors; capacity and impedance for handling at transfer facilities. The national network is augmented with additional detail in the study region. It is used to build paths for the national distribution channel, mode and intermodal transfer models.

Mode-specific zone-to-zone transport times, costs and other attributes. As with regional personal travel models, these data are needed to develop the national model components. They are either skimmed from the national freight network, or developed exogenously.

Economic data

Macro-economic Input/Output tables. The national model components use tables available from the U.S. Department of Commerce Bureau of Economic Analysis (BEA) that identify the commodity groups (SCTG codes[7]) made by each industry category (NAICS code) and the commodity groups used by each industry category. It is used in the supplier selection model to link buyers and suppliers.

Commodity flows. A very important input to the national model components is a detailed estimate of flows of the various commodity groups among all the national zones. This is provided by the Commodity Flow Survey[8] (CFS), conducted every five years by the Census Bureau and the Bureau of Transportation Statistics (BTS). It identifies the annual flows, in tonnage, of commodities (SCTG codes) between origin zone and destination zone, as defined in the federal Freight Analysis Framework (FAF). It is used in the supplier selection model to connect synthetic buyers and suppliers. It is also used to provide control totals in the Goods Demand model that apportions the CFS flows among the buyer-supplier pairs.

Survey data

Integrated shipment survey of sellers/shippers, carriers and buyers/receivers. This commodity-class-specific survey provides detailed information about shipments, with each shipment characterized by commodity class, shipper and receiver. The survey collects basic information about each step in the distribution of the shipment from seller/shipper to buyer/receiver. There may be many steps, including one or more occurrences of collection, consolidation, warehousing, line haul, de-consolidation, and distribution. However, transformations such as refinement, manufacture, value-added changes, and resale by a wholesaler or other middleman constitute the end of a shipment and, potentially, the beginning of one or more other shipments of the same or a different commodity class. Thus, the focus of the survey is on the shipment of goods, rather than on their production.

Unlike the spatial, network and economic data, there is no existing shipment survey that provides data of this type that is adequate for fully developing the models in the proposed model system.[9]

---

[7] http://bhs.econ.census.gov/bhs/cfs/Commodity%20Code%20Manual%20%28CFS-1200%29.pdf

[8] http://www.bts.gov/publications/commodity_flow_survey/

[9] The closest examples of this kind of survey from other countries include the French ECHO survey (see Canal et al, 2006, and Arunotayanun, 2009) and the Tokyo Metropolitan Goods Movement Survey (MLIT, 1982, 1994, 1999, 2004).

This survey needs a national/international component, sponsored by the US government. It can also include a compatible regional component for goods and services that are predominantly transported within the study region, or are of special importance to the study region.

The sample frame includes all business establishments that buy, sell, ship, receive, carry or manage the shipment of commodities for business purposes. Carriers and managers include transportation firms and third party logistics companies, as well as logistics units within companies that have their own distribution systems. Each respondent completes those parts of the survey for which it has information, and identifies other business establishments involved in their shipments that could complete those parts for which they lack information.

The objective is to get complete information for a shipment. However, in some cases, partial information could be useful in developing model components of the framework. In particular:

- Shipment information that connects buyer and seller without identifying all steps of the distribution chain could be used to develop the national supplier selection and goods demand models, and the regional models of shipment size and type.
- Shipment information that provides a complete description of the logistical steps for shipments, without identifying the buyer and seller, could be used to develop the national models of distribution channel, transport mode, and intermodal transfer.
- Shipment information that provides a complete description of collection and distribution within a region could be used to develop the regional models of vehicle choice, tour patterns, and stop location and timing for delivery from regional shippers, warehouses and distribution centers.

This survey can be developed so as to enable the use of electronic information systems that may or may not be in use by the surveyed firms. It may be possible to implement the survey in such a way that major firms and carriers routinely supply information to the survey database in the context of their existing shipment information systems.

Regional Model Components

The regional model components include tour patterns & vehicle choice, number of stops in a tour, stop sequencing, stop duration and time of day. Table 34 provides an overview of the types of data needed by each of these models. It is followed by descriptions of the data types.

Table 33. Usage of The Data Types by the Regional Model Components

Model component	Spatial data	Network data	Survey data
Tour patterns and vehicle choice	✓	✓*	✓
Number of stops in a tour		✓*	✓
Stop sequencing		✓	✓
Stop duration			✓
Time of day		✓	✓

* Needed but not in current model

Spatial data

Zone system. The zone system used for the regional freight model components is the same zone system that is used in the study region for models of personal transport and for highway traffic assignment.

Point, parcel, microzone or zonal attributes of employment, residence and network connectivity As with the regional personal travel models, these data are used for model estimation, calibration and application of the regional components of the model system. In the prototypes developed for this project, the attributes are at the zonal level. However, as with regional personal travel models, the quality of the model results might be improved if models of spatial choices were able to use more detailed spatial attributes, such as census blocks or parcels.

Network data

Regional road network representation. This is the region's standard highway network used for regional travel demand modeling, augmented with information about truck restrictions if these are important in the study region. It is used for the development and application of regional models of vehicle choice, tour patterns, and stop location and timing for delivery from regional shippers, warehouses and distribution centers. It is also used for the assignment of these trips, as well as the assignment of long-haul truck trips within the region identified by the national shipment path choice models (truck shipments into, out of and through the region, as well as truck drayage.

Mode-specific zone-to-zone transport times, costs and other attributes. These are used in the regional model components, and are skimmed from the regional road network.

Survey data

Integrated shipment survey of sellers/shippers, carriers and buyers/receivers. This is the same survey described in detail above for the national level components. It would be used for the regional models of shipment size and type, vehicle choice, tour pattern, and stop location and timing.

Regional survey of intra-regional distribution centers and business establishments that conduct last-leg distribution of goods and services within the region[10]. This survey captures the information needed to develop detailed region-level models of goods deliveries within the region, including those that are contained within the region as well as those that serve as the last leg of long-haul shipments from the national model for distribution by truck within the region. It provides detailed truck tour information, including vehicle type, day and tour pattern, tour timing, and location and duration of stops.

The survey interviews a stratified random sample of several thousand business establishments doing business in the region, including establishments located in the region and establishments located outside the region but with substantial transport operations within the region. The interview is analogous to a household travel diary. Sampled establishments provide information on the movements of their entire fleet over a 24 hour time period, including origin, destination, purpose, fleet and commodity information. The data are used to estimate the models and are also expanded by industry, size and location to represent the total population of commercial enterprises, for purposes of model calibration.

---

[10] Hunt, et al (2006) describe a survey of this type conducted in the Edmonton and Calgary regions of Alberta, Canada, and its use to develop models similar to the regional components of the integrated model framework.

An additional component of this survey is a stated preference exercise in which the establishment provides choices in response to hypothetical scenarios. The resulting data are used to estimate values of time for the various categories of establishment, for use in the model.

In order to conduct the survey of business enterprises it is necessary to develop a list of the establishments. This list is used as a sample frame (master list) to randomly select establishments for the interview. It is also used to determine how many firms are in each of the model's categories (by type and number of employees), so that the resulting sample can be expanded to represent the entire population. There isn't necessarily a single good data source for this, although the US Census Bureau's employment data products (such as OnTheMap) may be useful.

Some of the data collected in this survey might be gathered through the use of GPS devices on the delivery trucks, reducing the respondent burden and enabling the collection of large enough samples for good model development, larger than have historically been collected in surveys of this type (see Bassok, Outwater and McCormack, 2011).

5.3 Data Needed for Model Calibration

Once the models have been developed, it is important to calibrate them so that they match observed behavior in the base year used for model development. Although it would be desirable to use data other than was used in model development, it is common to use the same data for calibration as for model development; unlike passenger travel, where census data that is not used for model development can provide good calibration totals for some dimensions of the model system, little such data is available for calibrating freight models. But in addition to the data sources used for model development, it is important to use traffic count data of several types for calibration of the model components to match observed counts.

Traffic data

Counts of medium and large trucks entering and leaving the region. These counts are used for purposes of calibrating the national models.

Counts of trucks at port, intermodal, terminal and warehouse facilities. These counts enable expansion of survey data to represent all trips to and from the regional facilities.

Counts of trucks by type on screenlines and a sample of all arterial and larger roadways within the study region and on major national corridors and ports of entry/exit. These counts are used for purposes of calibrating the regional and national model components.

Intercept surveys. Intercept surveys are needed at the same places as count data in order to gather information about details such as commodity being carried that a count cannot determine.

5.4 Data Needed for Model Forecasting

National Model Components

Spatial data

Zone system. Typically, the same zone system would be used for forecasting as was used for model development.

Zonal attributes. For each zone in the national zone system, the same zonal attributes are needed for forecasting as for model development. The simplest approach to this would be to rescale the data that was used for model development, based on national and regional forecasts of employment overall and by sector.

Network data

National freight network representation. The network used for forecasting would typically be the same network that was used for model development, with changes made that reflect likely major changes in the national network, as well as likely or scenario-based changes within the study region.

Mode-specific zone-to-zone transport times, costs and other attributes. The zone-to-zone attributes would be skimmed from the national freight forecasting network, or developed exogenously.

Economic data

Macro-economic Input/Output tables. The same Input/Output tables can be used for forecasting as were used for model development, although in some cases it might be desirable to adjust the tables for known or anticipated changes in production technologies.

Commodity flows. The commodity flows, provided by the CFS for the base year, must also be developed for forecast years. These are available from the FAF, and can also be purchased, in more disaggregate form, from Global Insight as part of the Transearch database[11]. Limitations in the FAF and Global Insight forecasting procedures make it desirable to develop procedures or models that modify their commodity flow forecasts in response to regional conditions, or else to develop complete alternatives to their forecasts. In particular, their forecasts assume that base year mode splits are maintained in the forecast year for each commodity type.

Regional Model Components

Spatial data

Zone system. Typically, the same zone system would be used for forecasting as was used for model development.

Point, parcel, microzone or zonal attributes of employment, residence and network connectivity For each zone, the same zonal attributes are needed for forecasting as for model development. The simplest approach to this would be to rescale the data that was used for model development, based on regional forecasts of employment overall and by sector.

Network data

The network used for forecasting would typically be the same network that was used for model development, with changes made that reflect likely or scenario-based changes within the study region.

Regional road network representation. The network used for forecasting would typically be the same network that was used for model development, with changes made that reflect likely or scenario-based changes within the study region.

[11] See http://www.ihs.com/products/global-insight/industry-analysis/commerce-transport/database.aspx

Mode-specific zone-to-zone transport times, costs and other attributes. These are skimmed from the forecasting version of the regional road network.

5.5 Recommendation for Next Steps

The Short Term View

The key missing data for the proposed freight forecasting framework is disaggregate shipment data, described above as coming from a national level integrated shipment survey. However, major barriers stand in the way of the development of such data, as described in TRB Special Reports 276 and 304 (TRB 2003, TRB 2011), and it is unlikely that these barriers will be overcome in the short term. Nevertheless, further steps can be taken in the short term toward a full implementation of the regional components, and enhancing the national components.

Recommendation 1 (Regional Freight Survey): The first recommendation is to conduct a regional freight survey for a selected region, and implement regional components of the model framework for that region. The objective would be to demonstrate a full implementation of the regional components of the model framework. The survey is identified and described in detail above in the section on data needs for regional model components called "Regional survey of intra-regional distribution centers and business establishments that conduct last-leg distribution of goods and services within the region."

Recommendation 2 (Pilot National Data Collection): At the national scale, the key missing data for the proposed freight forecasting framework is disaggregate shipment data, described above as coming from a national level integrated shipment survey. As a first step toward closing this gap, a project could be conducted to find practical ways of obtaining the needed national data by carrying out pilot data collection efforts. The data collection methods would include interviews, Delphi panels, stated preference surveys, and shipment surveys. The in-depth interviews and Delphi exercises would target supply chain architects and shipment decision-makers representing shippers, carriers and third-party logistics firms. Prior to developing and implementing these exercises, a review of prior similar work would be conducted, so as to leverage those results[12]. The aim of these exercises would be to learn more about their decision processes and outcomes, and to inform the design and implementation of surveys and stated preference exercises. Small scale surveys and stated preference exercises would be implemented in order to test the adequacy of their design, and to fill data gaps. The final step in the project would be to refine the survey designs and model framework. This pilot would achieve an important incremental step toward bridging the gap between data need and data availability for implementing the national scale components of the model framework. One of the major challenges in a project of this type, and a key to its success, would be identifying and gaining the cooperative participation of the right experts in the right organizations.

The Long Range View

As the above sections of this report make clear, U.S. government agencies are the primary providers of the existing data used in the proposed freight modeling framework. The Bureau of Transportation Statistics and Census Bureau provide the Commodity Flow Survey and Freight Analysis Framework, the Census Bureau provides County Business Pattern data, and the Department of Commerce Bureau of Economic Analysis provides economic input/output tables.

[12] See, for example, the appendices of Cambridge Systematics (2010)

The key missing data for the proposed freight forecasting framework is disaggregate shipment data, described above as coming from a national level integrated shipment survey, although national intercept survey and count data for model calibration would also be beneficial. Here again, the U.S. government is uniquely positioned to lead the development of this data, for several reasons:

- **Geographic Scope.** The required shipment data must be national in scope, and also include information about flows through international ports, because many shipments are national and international in scope.
- **Shipment Complexity.** Often a shipment is handled by many different entities, none of which knows the complete shipment information. For this reason, traditional surveys may not work, and innovative data programs that collect and fuse disaggregate shipment data from multiple sources, perhaps via administrative procedures, may be required.
- **Confidentiality.** Freight data is tightly held by private entities that generate it, for commercial and privacy reasons. Probably no entity other than the federal government is in a position to unlock for legitimate purposes the information contained in confidential data, while at the same time assuring that the confidentiality requirements are satisfied.
- **National Security.** The U.S. government may have national security concerns related to freight shipments that lead it to develop freight data, some of which could serve other important purposes, including the shipment information required by the proposed freight model framework.

The above reasons for US government involvement were identified in TRB Special Report 276, A Concept for a National Freight Data Program (TRB, 2003), which, as the title suggests, recommended the development of an ongoing national freight data program. Similarly, TRB Special Report 304, How We Travel: A Sustainable National Program for Travel Data (TRB 2011) also called for the implementation of a National program for freight data collection. Both of these studies identified the needs, as well as issues preventing the implementation of such a program, but neither of them carried out the work needed to resolve the issues. This report adds its voice to those of SR 276 and SR 304. A federally sponsored effort is needed that involves the federal agencies and private entities with a vested interest in freight data. The effort would define a data program that meets the needs determined in this report as well as SR 276 and 304, and would resolve the issues preventing the implementation of the program.

6.0 SUMMARY

6.1 Freight Forecasting Framework

This freight forecasting framework represents a unique mixture of methods proposed for other urban areas. The principles of the freight forecasting framework are as follows:

- Use discrete choice methods wherever possible to determine probabilities of an outcome.
- Combine supply chain and tour-based methods into a single framework that can be applied to all commodities for both long-haul and short-haul goods and truck movements.
- Leverage existing model component structures to assist in estimating models, given resource and time limitations, on this research wherever possible.

The unique part of the framework is the combination of supply chain and tour-based methods into a single framework rather than a hybrid approach where supply chain modeling is used for certain industries or commodities and tour-based methods are used for the remaining industries or commodities. During the development of the framework, this hybrid approach was determined to be limiting because it would only capture a portion (albeit the majority) of the behavior for any particular industry or commodity. In fact, some industries contain both aspects of behavior and should be represented individually. Many data collection efforts conducted by MPOs to date are designed specifically to support the hybrid approach by collecting different surveys to support each tour-based and supply chain methods. This makes it difficult to make the connections between long-haul and short-haul goods movement.

In addition, the framework adds a model component to specifically identify the type of supply chain that is represented by long-haul commodity movements. This model component is difficult to estimate given that most surveys to date only ask questions about a portion of the supply chain and does not track the entire chain. When data collection methods are adapted to collect more data on the full supply chain, this model component can be expanded to more accurately represent the full supply chain.

6.2 Implementation and Demonstration

The freight forecasting framework was developed for two commodity groups in order to test the implementation of this framework in a real world setting:

- Manufactured Goods (SCTG 32-35, 38-40)
- Food Products (SCTG 1-9)

Each model within the framework was estimated from available data sources for these two commodity groups. These two commodity groups were chosen based on available sample sizes and on representing the different logistics in each group. These models were applied using Chicago regional and national data sources and open source software developed for this project.

Chicago Metropolitan Agency for Planning conducted a study in parallel to this research which implemented several components of the freight forecasting framework from existing research in the Chicago region (Cambridge Systemstics, 2011):

- **Firm Synthesis and Goods Demand** models were adapted from the Freight Analysis Microsimulation Estimator (FAME) models, originally developed by the University of Illinois Chicago (Samimi et al, 2010)

- **Production Mode and Intermodal Transfers** were adapted from the Norway and Sweden models (de Jong et al, 2007)

The remaining model components were estimated from available data sources specifically for this project:

- **Distribution Channel** models were estimated using FAME survey data
- **Shipment Size and Type** models were estimated using Texas Commercial Vehicle Survey data
- **Vehicle Touring Models** (including Vehicle Choice, Tour Patterns, Stop Duration and Location, and Delivery Frequency/Time of Day) were estimated using Texas Commercial Vehicle Survey data

The implementation and demonstration of these models produced results, but these were not calibrated or validated against observed data in the Chicago region. The implementation demonstrated that the software and results produced consistent and valuable information about freight and goods movement in a regional context. In addition, the sensitivities in a regional freight forecasting model are available in this approach.

6.3 Next Steps

This research provided a new approach to regional and national freight forecasting and demonstrated that the approach could be developed in a practical and useful manner for metropolitan planning organizations and state departments of transportation in the U.S. The approach overcomes several deficiencies in current freight forecasting methods.

One result of the research was the development of data collection recommendations to support the estimation, calibration and forecasting of the framework for future use. Many of the data needed are readily available, but estimation data are still needed for both the national and regional aspects of the framework. The national data collection should be considered by the U.S. Department of Transportation to support this type of forecasting across the country. The regional data collection should be considered by any region interested in applying the freight forecasting framework. This regional data collection may be considered at the state level, similar to the Texas data collection program, to provide more robust sample sizes and consistency for several regions in a state.

The model components in the freight forecasting framework should be re-estimated when new data are available, especially the production mode model, which was not re-estimated as part of this research, and the distribution channel model, which was limited in scope by the available data in the U.S. Ideally, each model component would be calibrated to local conditions and validated as a system for a region. Forecasts using a calibrated and validated modeling system could be reviewed for reasonableness and sensitivity tests would provide insight on the details of modeling system and may contribute to some adjustments. Forecasts of individual projects or future alternatives would also provide useful information on any aspects of the system that could be improved.

The freight forecasting framework contained herein is a start to implementing advanced modeling techniques to forecast goods movement and commercial vehicles for regional planning purposes. Further efforts to improve this framework with new data, model improvements, and forecasts would be welcome. The open-source aspect of the software allows any transportation agency to use and contribute to this framework.

7.0 BIBLIOGRAPHY

Arunotayanun , Kriangkrai (2009). **Modelling Freight Supplier Behaviour and Response**, Ph.D. Thesis, Imperial College London, United Kingdom.

Bassok, A., Outwater, m.; McCormack, E.; Ta, C. (2011). **Use of Truck GPS Data for Freight Forecasting**. Presented at the Transportation Research Board 90th Annual Meeting, pending publication in the Transportation Research Record, Washington DC.

Boerkamps, J. H. K., van Binsbergen, A.J., and Bovy, P.H.L. (2000) **Modeling Behavioral Aspects of Urban Freight Movement in Supply Chains**, TRR 1725, 2000.

Boerkamps, J., and van Binsbergen, A. (1999) **GoodTrip – A New Approach for Modeling and Evaluation of Urban Goods Distribution**, Delft University of Technology, and The Netherlands Research School for Transport, Infrastructure and Logistics (TRAIL).

Cambridge Systematics (2010). **Regional Freight System Planning Recommendations Study**, prepared for Chicago Metropolitan Agency for Planning.

Cambridge Systematics, Inc. (2011) **A Working Demonstration of a Mesoscale Freight Model for the Chicago Region Final Report and User's Guide,** prepared for the Chicago Metropolitan Agency for Planning.

Cambridge Systematics, Inc., with NuStats and Northwest Research Group (2007) **MAG Internal Truck Travel Survey and Truck Model Development Study Final Report**, Maricopa Association of Governments.

Canal, B., Costa, G., Franc, P., Gouvernal, E., Guilbault, M., Hémery, C., Lebelle, N., and Rizet, C. (2006). **Enquete ECHO - Premiers résultats d'analyse**, INRETS, Protocole d'accord du Ministère de l'Equipement, Des Transports et du Logement no 5803/02, Phase 6 Rapport final, (in French).

Cavalcante, R., and Roorda, M.J. (2010) **A DISAGGREGATE URBAN SHIPMENT SIZE/VEHICLE-TYPE CHOICE MODEL**, presented at the 89th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 10-14, 2010.

de Jong, G., Ben-Akiva, M. (2007) **A micro-simulation model of shipment size and transport chain choice**, Transportation Research Part B 41 (2007) 950–965.

Donnelly, R., Wigan, M., and Thompson, R. (2010), **Hybrid Microsimulation Model of Urban Freight Travel Demand**.

Fischer, M.J., Outwater, M.L., Cheng, L.L., Ahanotu, D.N., and Calix, R. (2005) **Innovative Framework for Modeling Freight Transportation in Los Angeles County, California**, TRR 1906, 2005.

Gliebe, J., Cohen, O., and Hunt, J.D. (2007) **Dynamic Choice Model of Urban Commercial Activity Patterns of Vehicles and People**, presented at the 86th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2007.

Holguín-Veras, J., and Patil, G.R. (2005) **OBSERVED TRIP CHAIN BEHAVIOR OF COMMERCIAL VEHICLES**, presented at the 84th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2005.

Hunt, J. D. (2006) **Tour-Based Microsimulation of Urban Commercial Vehicle Movements in Calgary, Alberta, Canada: Case Example**, Freight Demand Modeling: Tools for Public-Sector Decision Making, Summary of a Conference, September 25-27, 2006, Transportation Research Board., p62.

Hunt, J. D., Stefan, K.J., and Abraham, J.E. (2003), **Modeling Retail and Service Delivery Commercial Movement Choice Behaviour in Calgary**, presented at the Tenth International Conference on Travel Behaviour Research in Lucerne.

Hunt, J. D., Stefan, K.J., and Brownlee A.T. (2006) **Establishment-Based Survey of Urban Commercial Vehicle Movements in Alberta, Canada: Survey Design, Implementation, and Results**, presented at the 85th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2006.

Hunt, J.D., Stefan, K.J., Brownlee, A.T. (2006). **An Establishment-Based Survey of Urban Commercial Vehicle Movements in Alberta: Survey Design, Implementation and Results**. Transportation Research Record 1957, 75–83.

MLIT (Ministry of Land, Infrastructure and Transport) in the Japanese Government (1982, 1994, 1999, 2004) **Tokyo Metropolitan Goods Movement Survey**.

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.

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.

Nepal, S., Farnworth, S., Pearson, D. (2007c) **2006 San Antonio Area Commercial Vehicle Survey Technical Summary.** Texas Transportation Institute, Report prepared for Texas Department of Transportation Travel Survey Program.

Prozzi, J., Mani, A., Harrison, R. (2006) **Development of Sources and Methods for Securing Truck Travel Data in Texas.** Texas Department of Transportation, Report No. 0-4713-P2.

Prozzi, J., Wong, C., Harrison, R. (2004) **Texas Truck Data Collection Guidebook**. Report 0-4713-P4.

Roorda, M.J., McCabe, S., and Kwan, H. (2007) **A Shipper-Based Survey of Goods and Service Movements in the Greater Golden Horseshoe (GGH), Report I: Survey Design and Implementation**, Ministry of Transportation of Ontario, Region of Peel.

Roorda, M.J., McCabe, S., and Kwan, H. (2007) **A Shipper-Based Survey of Goods and Service Movements in the Greater Golden Horseshoe (GGH), Report II: Preliminary Analysis of Results and Comparison of Survey Techniques**, Ministry of Transportation of Ontario, Region of Peel.

Ruan, M., J. Lin, K. Kawamura (2011) **Modeling Commercial Vehicle Daily Tour Chaining,** presented at the 90th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2011.

Ruan, M., J. Lin, K. Kawamura (2011) **Modeling Commercial Vehicle Daily Tour Chaining,** presented at the 90th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2011.

Ruan, M., Lin, J., and Kawamura, K. (2010) **Modeling Urban Commercial Vehicle Daily Tour Choice Using the Texas Commercial Vehicle Survey Data**, presented at the 89th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 10-14, 2010.

Samimi, A., Mohammadian, A., Kawamura, K., (2009) **Behavioral Freight Movement Modeling, Resource Paper W2: Behavioral Paradigms for Modeling Freight Travel Decision-Making**, the 12th International Conference on Travel Behavior Research, Jaipur, India.

Samimi, A., Mohammadian, A., Kawamura, K., (2010) **A Behavioral Freight Movement Microsimulation Model: Method and Data**, Transportation Letters: The International Journal of Transportation Research, DOI 10.3328/TL.2010.02.01, Volume 2, pp 53-62.

Stefan, K. J., McMillan, J. D. P., and Hunt, J. D. (2005) **Urban Commercial Vehicle Movement Model for Calgary, Alberta, Canada**, presented at the 84th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2005.

Tavasszy, L.A., Smeenk, B., And Ruijgrok, C.J. (1998) **A DSS For Modeling Logistic Chains in Freight Transport Policy Analysis**, Int. Trans. Opl Res. Vol. 5, No. 6, pp. 447±459, 1998.

TRB (2003). **A Concept for a National Freight Data Program**, TRB Special Report 276, Transportation Research Board of the National Academies, Washington, D.C.

TRB (2011). **How We Travel: A Sustainable National Program for Travel Data**, TRB Special Report 304, Transportation Research Board of the National Academies, Washington, D.C., 2011.

Wang, Q., and Holguín-Veras, J. (2009) **Tour-based Entropy Maximization Formulations of Urban Freight Demand**, presented at the 88th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2009.

Wisetjindawat, W., And Sano, K. (2003) **A Behavioral Modeling In Micro-Simulation For Urban Freight Transportation**, Journal of Eastern Aisa Society for Transportation Studies, Vol 5, October 2003.

Yang, C.H., Chow, J.Y.J., Regan, A., (2009) **State of the Art of Freight Forecasting Modeling: Lessons Learned and the Road Ahead**, presented at the 88th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., January 2009.

Appendix A. Freight Forecasting Framework Software

This appendix provides a step-by-step description of the code for implementing the supply chain and tour-based freight forecasting model developed for the CMAP region. The code uses County Business Pattern (CBP), Input/Output, and network skim data that had been pre-processed for use in CMAP's Mesoscale freight model. The code has the following major steps:

- Firm Synthesis
- Supplier Selection
- FAF3 Flow Apportionment
- Business Location Assignment
- Distribution Channel
- Shipment Size and Frequency
- Mode-Path Selection
- Vehicle choice and Tour Patterns
- Stop Sequence (for peddling tour pattern)
- Stop Duration
- Time of Day
- Prepare Trip Tables

The implementation of each major step in the code is described below. Every piece of the code with each major step has a corresponding description below.

A.1 Firm Synthesis

This is the first step in the model and involves synthesis of individual firms or businesses using aggregate control data from Census County Business Pattern (CBP) data. The CBP data should already be processed and output in a *.csv file that contains the CBPZone, FAFZone, NAICS code and number of firms in each of the eight firm size categories. The eight firm size categories are as follows:

- 1 - 19
- 20 - 99
- 100 - 249
- 250 - 499
- 500 - 999
- 1,000 - 2,499
- 2,500 – 4,999
- Over 5,000 employee

For the CMAP region, agricultural and construction employment data were added to the CBP data using local sources. The firm synthesis step involves the following sub-steps:

1. The employment data are read in from a CSV file to an R data frame, which is then converted to an R "data.table". The code uses a data.table instead of a standard R-data frame where possible to improve the performance of the model.

2. Employment data is aggregated by NAICS, CBPZone, FAFZone, and firm size category (esizecat) to obtain the number of firms in each of the combinations.
3. A previously prepared correspondence between NAICS6 (6-digit NAICS) and the NAICS used in IO tables (NAICSIO) is used to map NAICSIO onto the firm list. The correspondence file is "n6io_short.csv".
4. The dataset is then enumerated (one record per firm) where each firm is identified by NAICS, CBPZone, FAFZone, and esizecat. There are about 8.4 million firms.
5. A correspondence between NAICS and SCTG (naics_uniqsctg.csv) is used to attach a unique SCTG commodity produced to each firm in the firms database.
6. Some industries produce more than one commodity (such as Wholesale). To account for this, the commodity for each such firm is simulated based on the probabilities of the multiple commodities that it could produce.
7. The NAICSIO and SCTG commodity code are combined to create a combined attribute - N6IO_SMake.
8. Firms are then aggregated by CBPZone, FAFZone, N6IO_SMake, and esizecat and re-enumerated so that there is one record per firm.
9. Two identifier variables are created. A firm type id (ftid) combining CBPZone, N6IO_SMake, and esizecat. A business id (busid) is also created for each firm by combining the ftid and its serial number within the firm type.
10. A producers/suppliers database is created by filtering out the firms that don't have a missing SCTG code.
11. One producer firm is flagged as "mustkeep" within each N6IO_SMake-CBPZONE category. This is to ensure that there is at least one producer/supplier firm representing an industry-commodity category in each zone.
12. A producers firm type database is also created where only one producer for each unique firm type (ftid) is retained. This database is later used to select the supplier in the supplier selection model. A firm type is chosen as a supplier first to increase computational efficiency. The specific firm within the firm type is chosen in the Business Location model.
13. A consumers/users database is created next. This consists of all the firms in the firm database merged with processed I/O data (user5top_array.csv) based on the NAICSIO of the consumer.
14. The processed I/O data (user5top_array.csv) identifies the top 5 SCTG commodities (in terms of value) consumed and the corresponding supplier NAICSIO for each consumer industry NAICSIO. It was thought that top 5 would include most of the commodities consumed a particular firm and hence this limit was set to decrease computational burden.
15. Similar to what is done during firm synthesis, the SCTG commodity for suppliers who could produce more than one SCTG commodities is simulated using probability thresholds. This is done for all the top 5 suppliers that are being considered.
16. It is being assumed that a certain percentage (in this case, 30%) of consumers would work with a wholesaler instead of directly with a producer. Therefore, some suppliers to consumers who themselves are not wholesalers, are probabilistically mutated to an appropriate wholesale supplier (NAICSIO) based on the SCTG commodity being consumed.
17. Just as in the case of producer firms, certain consumer firms are flagged "mustkeep" so that at least one unique industry-firm size combination is retained in every CBPZone.
18. Finally, two datasets are created in this step. A makers dataset is created from the producers firm type database and consists of firms that are located in a Chicago zone or if they have more than a 1,000 employees or if they have been flagged "mustkeep". A users dataset is created from

the consumers database and consists of firms that are located in a Chicago zone or if they have more than a 500 employees or if they have been flagged "mustkeep" or if they fall in a set of 5% of randomly selected consumers.

A.2 Supplier Selection

At this stage there is a list of supplier (or maker) firms and a list of consumer (or user) firms. In this step, for each consumer in the users dataset, a supplier is selected from the makers dataset. Note that selection of a supplier does not mean the selection of an exact business but that of a firm type (CBPZone, N6IO_SMake, and esizecat). The following sub-steps are involved:

1. A pre-processed FAF OD flow dataset (f7_array.csv) is read in. It contains a list of supplier CBPZones and corresponding tonnages for each consuming CBPZone by SCTG commodity.
2. A new variable – NSCa – is created in the users dataset which is a combination of N6IO_SMake (industry and commodity required by the consumer) and CBPZone (location of consumer firm).
3. Great Circle Distances (GCDs) between CBPZones are read in from a pre-processed skim file (ornlfull.csv). The skim values are obtained from the Oak Ridge National Laboratory (ORNL) skim files.
4. The users dataset is restructured from the wide format to a long format i.e., the top five suppliers for each consumer that were columns are now transformed into rows.
5. A new dataset is created from the users dataset that has one record per unique NSCa - NSCa_Uniq.
6. The NSCa_Uniq, makers and FAF OD flow datasets are all merged together such that (i) the supplier industry matches in the NSCa_Uniq and makers datasets, (ii) the NSCa_Uniq SCTG commodity and consuming zone match those in FAF dataset, and (iii) the supplier CBPZone matches in the makers and FAF datasets.
7. The first condition is to filter out those industries (and hence makers) that are not required by any of the consumers (represented by NSCa_Uniq dataset). The remaining two conditions ensure that a maker is matched to a consumer only if there are flows observed in the FAF data between the consumer CBPZone and maker CBPZone for the particular commodity being traded.
8. The merge creates a set of potential makers from which a supplier for each consumer will be selected.
9. The potential makers dataset could be directly merged with the users dataset in a many-to-many merge using the NSCa field. It would result in joining all the possible supplier choices to each consumer in the users dataset. This merge cannot be achieved under the current memory limit.
10. The potential makers dataset is split into two. One that has less than 10 suppliers for a unique NSCa (combination of required supplier industry and consumer CBPZone), makers_lt10, and other that has more than 10 suppliers for a unique NSCa, makers_gt10.
11. A map is created between the NSCas for which there are more than 10 potential makers and their corresponding row numbers (in makers_gt10).
12. The firms in the users dataset that have less than 10 potential suppliers are directly merged with makers_lt10 to create users_lt10 that has all the supplier choices for each user firm.
13. For the users that have more than 10 choices of potential suppliers, a process loops through all the unique NSCas and randomly picks suppliers from makers_gt10 that are 10 times the number of user firms for a particular NSCa. For example, if there are 15 user firms for a particular NSCa, then this process randomly chooses a 15*10=150 such supplier firms (from

makers_gt10) and creates the choiceset for all the 15 user firms in one go. This process creates users_gt10 dataset.

14. users_gt10 is appended to the users_lt10 to create a dataset that has all consumers firms and the supplier choice sets corresponding to each consumer firm. This dataset is evaluated for supplier selection and a single supplier is selected for each consumer firm.
15. The maker CBZone is calculated from the maker ID (which is the ftid of the maker) and GCDs are merged in for each user CBPZone and maker CBPZone.
16. There is a utility-like value given to each of the suppliers in the choiceset based on their satisfying various criteria such as employee size cat, CBP zone ranges, and GCD between consumer and supplier. A random constant is then added to this value. The following table shows the various parameters used:

Consumer Business Size (Number of Employees)	Coefficient							
	Producer Business Size (Number of Employees)			Great Circle Distance between Consumer and Producer (Miles)				
	1 to 99	100 to 499	500+	Over 1,509	596 to 1,509	150 to 595	1 to 149	0 (Intracounty)
1 to 99	0.2	0.2	0.4	-0.4	-0.3	-0.2	0	0.1
100 to 499	0.2	0.6	0.6	-0.2	-0.1	-0.05	0	0.1
500+	0.4	0.6	0.6	-0.1	-0.05	0	0	0.1

17. For each consumer, the supplier with the highest utility score is selected from within the choiceset. This gives the buyer-supplier pairings between the firm that have been synthesized, the pairs dataset.

A.3 FAF3 Flow Apportionment

After pairing buyer and supplier firms, the annual flow between each of the pairs is estimated in this step. FAF3 flows are apportioned to each buyer-supplier pair based on a consumption estimate for the supplier firm which is in turn determined by the firm size. This is done through the following steps:

1. The number of employees in each consumer\buyer firm is calculated as the mid-point of the range of the firm size category (esizecat) in the pairs dataset.
2. A pre-processed dataset created from I/O tables is read in (iomu6.csv). It contains the value of commodity consumed per buyer employee by a combination of supplier and buyer I/O NAICS\industry.
3. The pairs dataset is then merged with iomu6 to attach consumption values per employee to all the buyer-supplier pairs.
4. Buyer and supplier FAF zones are added to the pairs dataset using the respective CBPZones.
5. A pre-processed FAF3 flow dataset is read (f9.csv) and merged with the pairs dataset by origin and destination FAF zones and commodity SCTG.
6. Consumption value for each buyer firm is estimated as the product of the value consumed per employee (from I/O tables) and the firm size.

7. This consumption value is then aggregated across origin and destination FAF zones and commodity SCTG and merged back into the pairs dataset.
8. The proportion of flow between each buyer-supplier pair is calculated as the ratio of individual consumption value and the aggregated value.
9. This proportion is multiplied by the FAF3 tonnage and value (merged earlier from f9.csv) to obtain the annual tonnage and value flow between a firm pair.
10. In the final step, total tonnage between the firm pairs is factored appropriately to match the total tonnage in the FAF3 dataset by commodity SCTG.
11. FAF tonnages and the estimated tonnages are aggregated by SCTG. SCTG level factors are obtained and applied to estimated tonnages to match the observed (FAF) tonnages.

A.4 Business Location Assignment

Until this point, the geographic identifier for all the firms is CBPZone (which is based on county boundaries). In this step, the firms are located within a higher resolution geography, mesozones.

1. From the database of all firms, select businesses in the CMAP region based on the appropriate CBPZONE range.
2. There are a few counties which correspond to only one mesozone. The firms in these counties are directly assigned a mesozone.
3. The other counties are comprised of more than one mesozone. Firms in these counties are assigned to mesozones based on employment ranking by industry.
4. A pre-processed dataset contains the percentile ranking of each of 21 NAICS categories by mesozone based on employment numbers in each of those industries (mz_emp_rankings.csv). Higher employment implies a higher percentile rank.
5. This rankings dataset is merged with the firms database by county.
6. For each of the 21 NAICS categories considered, candidate mesozones are identified based on firm size and the ranking of a particular NAICS in a mesozone. The probability of a mesozone getting assigned to a particular firm increases with the rank of the firm's NAICS in the mesozone and the number of employees in the firm. For example, if a firm belongs to the manufacturing industry and has a firm size greater than 5,000 then all mesozones which have manufacturing ranked 9th or 10th are candidates for the particular firm.
7. Once candidate mesozones are assigned to each firm, one of the candidates is randomly selected as the firm's mesozone.
8. Firms not in the Chicago region are assigned a mesozone number based on their CBPZone number and combined with the Chicago firms dataset to create a full firms database with mesozones attached.

A. 5 Distribution Channel

This step simulates the distribution channel used between each pair of firms. Four alternatives for distribution channel are simulated: direct, one stop type, two stop types, and three stop types where stop types are warehouse, distribution center, or consolidation center. Separate models were estimate for food and manufacturing products. For other commodities, the manufactured products model is applied.

1. The model specification files for both food and manufactured products are read in (distchannel_food.csv and distchannel_mfg.csv).

2. Extra variables required by the two model specifications are created from the characteristics of the buyer and supplier firms and the shipment.
3. A logit simulation function simulates the distribution channel for firm pairs involving food products and firm pairs involving all other commodities.

A. 6 Shipment Size

Simulation of the shipment size is also based on a logit model. There are three alternatives: less than 999 lbs, 999-9999 lbs, and greater than 9999 lbs. The model that was estimated for food was used to simulate the shipment sizes for all commodities for demonstration purpose.

1. The shipment size model specification file (ShipSize_food.csv) is read in and extra variables required by the specification are created.
2. The logit simulation function is applied to determine the range of the shipment size used by each of the firm pairs.
3. Once the range is known, another function is used to simulate the exact shipment weight based on probability thresholds from the Texas survey for each shipment size category.

A.7 Mode-Path Selection

The mode-path selection step determines the mode and path of the shipments between each firm pair. There are 54 mode-path combinations out of which four are direct (two truck and two rail mode-paths).

1. The buyer mesozone is merged in from the updated firm database. The firms in the database were assigned a mesozone in Step 4.
2. Until this point, the supplier is a firm type. Based on the number of firms in each firm type, a firm is randomly selected to represent the exact supplier firm.
3. The seller\supplier mesozone is then merged in from the firms database based on the business ID (busID) of the supplier firm.
4. Pre-processed skims (emme3_skims.csv) are read in. The dataset has times and costs associated with all the 54 mode-path choices between all mesozone pairs.
5. The firm pairs dataset is now split into two based on the distribution channel: direct and indirect.
6. A commodity category and SCTG code correspondence (sctgcat.csv) is read in and merged into the direct and indirect firm pair datasets.
7. Annual logistic costs for a firm pair is calculated stochastically based on skims, mode-path capacities, storage costs, damage\loss costs etc for all the available mode-path alternatives. Four mode-paths are evaluated for direct distribution firm pairs and 50 for the indirect distribution firm pairs.
8. The least cost mode-path is selected for all the firm pairs. For a few pairs for firms with a direct distribution channel, direct skims are unavailable. In those cases, the indirect mode-path skims are used and an indirect mode-path is selected. The distribution channel is also changed from direct to an indirect on based on the proportions of occurrence. The opposite is done if indirect skims are unavailable for a firm pair with an indirect distribution channel.
9. After the mode-path for each firm is selected, the daily frequency of delivery is determined. The annual tonnage between each firm pair is divided by the shipment weight to get the number of annual shipments and the number of annual shipments is divided by 310 to get daily frequency.

10. Not all firm pairs have daily deliveries\pickups. A random number between 0 and 1 is generated and firm pairs for which the random number is less than daily frequency are assumed to have deliveries on the day being simulated. This also ensures that firm pairs with daily frequency greater than one have daily deliveries\pickups. From this point onwards only the selected firm pairs and corresponding shipments are simulated to represent a typical day scenario.
11. A local warehouse (within the CMAP region) is randomly selected for indirect distribution channel shipments. This warehouse would represent the last transfer stop form incoming shipments and first transfer stop for outgoing shipments.
12. Based on the mesozone of the warehouse and whether or not a shipment is incoming or outgoing, the origin and destination mesozones of all the shipments that are to be simulated in a day are determined.

Buyer Location	Distribution Channel	Origin	Destination
Buyer in Chicago	Direct	Seller mesozone	Buyer mesozone
Buyer in Chicago	Indirect	Warehouse mesozone	Buyer mesozone
Buyer outside Chicago	Direct	Seller mesozone	Buyer mesozone
Buyer outside Chicago	Indirect	Warehouse mesozone	Seller mesozone

13. If distribution channel is direct, origin mesozone is the mesozone of the seller, if not it is the warehouse mesozone (in case of a drop-off a truck will start from the warehouse with a shipment and in case of a pick-up, truck will start from the warehouse without a shipment).
14. If distribution channel is indirect and buyer is outside Chicago, destination mesozone is the seller mesozone (to simulate a pickup), if not it is the buyer mesozone.

A.8 Vehicle Choice and Tour Patterns

This step marks the beginning of the regional model system. There are six alternatives for this model which are combinations of 2 tour patterns (direct and peddling) and three vehicle types (2 axle, 3-4 axle, and semi/trailer). The model is influenced by commodity type, destination type, pickup\drop-off weight and county employment in the destination zone.

1. The model specification file is read in (tour_vehicle.csv).
2. Additional variables required by the model are created.
3. County employment data is read in (CBPemp.csv) and merged with the dataset of daily shipments to be simulated.
4. The logit prediction function is applied to predict the tour pattern and vehicle type choice for each daily shipment.
5. Shipments using a direct distribution channel are forced to have a direct tour pattern keeping the vehicle type choice same as that output by the simulation\prediction.

A. 9 Stop Sequence

The stop sequence step involves several stages. First, the category of number of truck tours for each shipment is determined. Each shipment record now represents a stop to be made on a truck tour. The stops for the indirect shipments are then clustered according to the number of truck tours category. Finally, the clustered stops are sequenced using a greedy algorithm.

1. There are three alternatives for the category of number of truck tours: all (stops) in one tour, two tours, three tours, and four tours. All in one tour means the shipment belongs to a tour category in which a truck covers all the stops assigned to it in a single tour. The purpose of this is to be able to determine the number of tours for all the stops serviced from a particular warehouse.
2. The number of tours model specification is read in (NumberofTours.csv) and additional required variables are created. The model requires pickup\drop-off weights, buyer and supplier industry categories etc. The model is applied to the daily shipments dataset.
3. The shipment stops are now clustered based on the number of truck tours category predicted. This is done only for indirect shipments since direct shipments are serviced by a direct tour with a single stop.
4. A separate vehicle type variable is created from the tour pattern vehicle type choice.
5. The count of shipment stops is obtained by warehouse, number of tours category, and vehicle type. Vehicle type is also used here in the aggregation so that stops grouped\clustered together have the same vehicle type since they are assumed to be serviced by the same vehicle.
6. It is ensured that the count of shipment stops is at least as much as the number of tours in the corresponding category. For example, if a particular shipment falls in a 4 tour category, there should be 3 more shipments that are in the 4 tour category assigned to the same warehouse for consistency. If that is not the case, the number of tours category is modified to be consistent.
7. The count of shipment stops is obtained again by warehouse, number of tours category (modified), and vehicle type.
8. The X and Y coordinates for each mesozone are read in (MZ_CEN.csv) and merged into the shipments data by destination mesozone (stop mesozone).
9. A function clusters the shipments for each warehouse by the number of tours category and vehicle type. For example, if a warehouse has 5 indirect shipment stops assigned to it and all of them fall in the two tours category using 2 axle truck, the clustering function would cluster the five stops in to two clusters using the Euclidean method.
10. After all the shipments stops are clustered into specific tours, unique tour and trip IDs are assigned to all the records.
11. Skims at the mesozone level are read in (mz_skim.csv).
12. For sequencing all the stops within a tour, a greedy algorithm is applied. It assigns the first stop as the one that is closest to the base (warehouse). It then keeps adding the next closest stop from the previous stop to the trip sequence until all of the stops are sequenced. After the final stop, the truck returns to the base.

A. 10 Stop Duration

The stop duration multinomial logit model predicts the durations of all stops for both direct and indirect distribution channel shipments\stops. There are six alternative stop durations: less than or equal to 15 minutes, 15-30, 30-45, 45-60, 60-75, and greater than 75 minutes.

1. The stop duration model specification file (StopDuration.csv) is read in and additional variables required are created in the shipments datasets.
2. The logit simulation function is used to predict the stop duration categories for both direct and indirect shipment datasets.
3. The exact stop duration in both datasets is assigned as the mid-point of the respective stop duration category:
    a. 0-15 min = 0.25 hrs
    b. 15-30 min = 0.375 hrs
    c. 30-45 min = 0.625 hrs
    d. 45-60 min = 0.875 hrs
    e. 60-75 min = 1.125 hrs
    f. >75 min = 2 hrs

A. 11 Time of Day

The time of day (TOD) multinomial logit model simulates the time of departure of the first trip in a particular truck tour. There are 5 alternatives for this – before 6AM, 6-8 AM, 8-9 AM, 9-10 AM, and after 10 AM.

1. The time of day models specification file (TOD.csv) is read in.
2. Other extra variables required such as total tour and stop durations are calculated for both direct and indirect tours.
3. Time of day categories are simulated for direct and indirect tour datasets. Based on the category of time of day an exact time is assigned to the tour start as follow:
    a. Before 6 AM = 5
    b. 6-8 AM = 7
    c. 8-9 AM = 8.5
    d. 9-10 AM = 9.5
    e. After 10 AM = 10.5
4. Now for indirect\peddling tours, the start times for each of the individual trips are calculated by adding the travel duration to current stop and stop duration at the current stop to the start time of the previous trip. Since the tour start time or the start time for the first trip is already simulated, start times for all other trips in the tour can be calculated.
5. There are certain tours which might have some trips starting very late in the day. Since it is unlikely that such trips occur, these are tagged on to a new vehicle\driver. A loop is run on the trips dataset until no trip starts after 10 PM.

A. 12 Prepare Trip Table

This least step in the model aggregates all of the trips by origin and destination mesozones to estimate a trip table.

1. Text descriptions of vehicle types and times of day are added to direct and indirect tours dataset.
2. Both the direct and indirect tours datasets are combined together.

3. The combined dataset is aggregated by origin mesozone, destination mesozone, vehicle type, and time of day to produce the final trip table.