Freight Micro-simulation in the U.S.

Amir Samimi (corresponding author)*
Ph.D. Candidate
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor Street
Chicago, IL 60607-7023
Tel: (312) 996-0962
Fax: (312) 996-2426
Email: asamim2@uic.edu

Abolfazl (Kouros) Mohammadian, Ph.D.
Associate Professor
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor Street
Chicago, IL 60607-7023
Tel: (312) 996-9840
Fax: (312) 996-2426
Email: kouros@uic.edu

Kazuya Kawamura, Ph.D.
Associate Professor
College of Urban Planning and Public Affairs
University of Illinois at Chicago
412 S. Peoria Street
Chicago, IL 60607-7064
Tel: (312) 413-1269
Fax: (312) 413-2314
Email: kazuya@uic.edu

Paper submitted for:
Presentation in Transport Chicago Conference, 4 June 2010

ABSTRACT

Passenger transportation planning has been performed in the well-known four step framework for almost a half-century. The same approach has been employed by the researchers for freight movement modeling. However, the functionality of this approach is questioned for freight studies. The main reason for the misleading results is attributed to the absence of the crucial role of the firms, as the decision making agents. Similar to the passenger travel behavior studies, research on the freight movement modeling is becoming more behavioral, by incorporating the supply chain concepts in it. Drastic changes in the behavior of the decision makers in the freight transportation systems motivate the researchers to change the modeling approach and keep the reliability and efficiency of the freight transport systems. Firms' behavior in forming the supply chains is the missing component of almost all the current models. In this paper, a behavioral activity-based freight movement microsimulation has been introduced, and the data needs are elaborated. The proposed framework has five modules which are *Firm Generation*, *Supply Chain Replication*, *Shipment Forecasting*, *Logistics Planning*, and *Network Analysis*. The data needs are also discussed and the available datasets in the U.S. are explained. In order to gather needed information on the commodities shipping process, which otherwise would be unavailable, an affordable establishment survey is also included.

Keywords: Fright Modeling, Supply Chain, Freight Data Needs, Freight Survey

INTRODUCTION

In addition to the deregulation of freight industries in the early 1980's, an increase in globalization and use of information technology prompted the firms to apply supply chain management (SCM) concepts in the real market (*1*). That has led to more efficient and complex behaviors in the production and distribution cycles of commodities. For example, in the U.S., the share of the logistics-related components of the GDP decreased from about 17% in 1980 to just above 10% in 2000 (*2*). Long haul commodity flows increased when the firms sought better partners across the country or even the world to form the best possible chain. In order for the firms to survive in the competitive market, they had to keep the transportation costs as low as possible by using the knowledge of logistic professionals. While the structure of supply chains and freight transportation systems along with shippers' and carriers' behavior changed toward more complex patterns, the conventional four-step models became even more unreliable; mainly because of the obsoleteness of the theory behind them (*3*). As understood from the name, similar to the passenger travel demand models, this framework has four sequential modules: commercial trip generation, distribution, mode choice, and traffic assignment. It is not possible to capture the strategic decisions that individual firms make regarding their supply chain design and operations using a four-step model. For example, issues such as: how a supply chain is shaped, which firm has the dominant control over the chain, how the shipping decisions are made, whether or not the shipping task should be contracted out, how the warehousing facilities should operate, whether or not a consolidation and/or distribution center is needed, could not be effectively incorporated in the model, if the firms' characteristics and the way they behave are ignored in the framework.

Even in the passenger transportation modeling, the effectiveness of the four-step framework is questioned (*4*). In the past few decades, researchers have developed and advanced the Activity Based Modeling approach (*5*). In this emerging framework, the way that the individuals (or households) are making decisions on the type of activity, destination choice, mode choice, etc, are embedded in the model. It was partly motivated by the need to incorporate changes in travel behavior such as trip chaining. Unlike the freight transportation, there has been significant progress and investment in this area, and considerable amount of literature has been produced by the researchers. Nevertheless, there is a lot to be done in the area of freight planning and modeling. The primary objective of this study is to introduce an activity-based freight movement microsimulation framework that incorporates the SCM concepts, and propose strategies for gathering data for calibrating such model.

BACKGROUND

In this section, an overview of the past efforts on freight demand forecasting model is provided. It should be noted that this study focuses on the national level model, and thus the literature review does not include the considerable volume of the literature on urban freight models.

During the past two decades, the four-step freight modeling framework, primarily designed for the passenger transportation, has made the groundwork for freight demand forecasting. Trip generation component mainly uses zonal employment or economic activity in order to forecast the production and attraction of commercial trips. This approach has been studied by several researchers including Anderson et al (*6*). The second component, trip distribution, is usually performed using some form of gravity model in which the distance between the origin and destination is embedded in the impedance function (*7*). Mode choice is one of the most critical and controversial parts of the framework. In earlier models, mode choice decisions were mostly based on the shipping cost (*8*); however, reliability, flexibility, haul time, and quality of service entered the analysis when the random utility models became trendy (*9*, *10*). For the intercity freight traffic assignment, "all-or-nothing" method is the typical approach, though intracity freight traffic assignment is usually performed in conjunction with the passenger traffic assignment.

Baumol and Vinod (*11*) are among the pioneers in modeling both mode choice and demand for a freight network. They utilized the same approach that had been developed for the analysis of passenger transportation. The mode choice model considers the trade-off between the transportation cost, time, reliability, and safety, and also accounts for the carrier and commodity heterogeneity. In a seminal paper, Harker and Friesz (*12*) presented a network equilibrium model of a freight transportation system in which the generation, distribution, modal split and assignment of freight movements were performed simultaneously. They used the same framework as the four-step passenger transportation modeling, but modified it by introducing a profit maximization model in the supply side and a price equilibrium model in the demand side, as well as an economic mechanism that integrates both of them. In a recent study the Wisconsin Freight Model (*13*) followed a variation of the traditional four-step modeling approach, utilizing commodity flow data to develop trip generation and distribution models. Twenty five commodity groups were defined in 72 Wisconsin counties and 74 zones for the rest of the U.S. and North America. Employment data was used for trip generation and attraction for each commodity category. Trip distribution model was based on a gravity model that used average trip lengths in the impedance function. There was not a modal selection model and the existing mode shares by commodity were used. Using the total daily truck trip table, freight traffic assignment was performed in conjunction with the daily and long distance passenger trips.

Limited attention has been paid to the development of a model that fully captures the important role of firms as the decision-making agents. Hensher and Figliozzi (*3*) made a convincing argument for the need for innovative freight models to deal with increased congestion and to maintain the reliability and efficiency of freight transport systems. They believe technological improvements and logistics practices are main reasons that necessitate changes in the freight demand forecasting methodologies. Companies have become increasingly customer-order-driven and new production systems such as Just-in-Time (JIT) and Time Based Manufacturing are now getting popular. In a comprehensive European freight study, de Jong and Ben-Akiva (*14*) stated that almost all the existing freight transportation models are lacking the supply chain and logistic elements. They introduced a new perspective in freight modeling by bringing up the firms' characteristics into the model to resolve the issues. They disaggregated the inter-zonal freight flow into the firm to firm commodity flow and introduced a behavioral freight model. Although their paper did not propose new ideas in the trip generation and traffic assignment, a substantial step was taken toward a behavioral freight model. A recent study sponsored by the American Association of State Highway and Transportation Officials in cooperation with the Federal Highway Administration (*15*) is a comprehensive source for the freight demand models in the US that covers recent studies and data collection efforts.

Availability of data, or lack thereof, has always been a barrier and an inevitable challenge in the majority of freight studies. Secondary data exist, but access is limited to highly aggregated information that is not sufficient for model development purposes. Most freight data are proprietary information. Some firms actually are in the business of collecting and analyzing freight data, and understandably are not willing to share the information. Surveys can be difficult since shipping decisions are a critical part of corporate strategy, and decision-makers are mostly unwilling to participate in any freight survey for fear of jeopardizing their competitive edge. A few years ago a survey was performed in Alberta, Canada, to better understand the complexity of commercial vehicle movements (*16*). They interviewed more than 7,300 businesses in order to obtain an establishment-based dataset which is essential for the development of a behavioral model of freight movements. Lack of such data in the U.S. presents a serious obstacle for the researchers by restricting the scope and framework of the model that can be developed. Data limitations and solutions in the U.S. are explicitly elaborated in this study as well.

MODELING FRAMEWORK

In order to make the model results more realistic and behavioral, the decision making agents should be effectively incorporated in the model. The unit of observation in the freight transportation models can be a firm, a shipment, or even the commodity flow between zone pairs (*14*). Generally, aggregation leads to loss of information, hence the more disaggregate the unit of observation, the more flexible and possibly more accurate the model will be. Since firms are the decision makers, the appropriate unit of observation for a behavioral model should be at the firm level. Integrating the firms in the freight transportation model is the essence of the disaggregate freight models, and has been practiced by a few researchers (*3, 14, 17*). However, the crucial module, which is mostly ignored in the literature of freight movement modeling, is defining the supply chains by capturing the essential characteristics of the companies. Understanding the way that the supply chains are formed will provide a robust tool to mimic the decision making process in the model.

Proposed freight microsimulation framework has five basic modules, which will be discussed in detail in the latter part of the paper. 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 type, and amount of incoming and outgoing goods, and the design of the supply chains are replicated in the second module. In the third module, however, the shipment types will be defined based on the acquired information on the firms’ characteristics and the way that they trade commodities between each other. The forth module in which the shipping decisions are made is of great importance, because the decisions such as shipping mode, haul time, shipping cost, warehousing, consolidation, etc are made. Even though sophisticated firms make decisions on physical infrastructure of the supply chain and logistics strategies simultaneously; in our proposed approach, this module has been separated from the second, in order to make the modeling structure compatible with the available data. A two-step (machine-learning and discrete choice) disaggregate mode choice model is introduced by the authors in another study (*18*), which will be used in this module of the framework. Finally, in the last module, the impact of the freight shipments on the transportation network is investigated.

In an ideal modeling structure, the above-mentioned modules are interrelated and a recursive structure leads to more realistic results. For example, the results of the last module, *Network Analysis*, could help the model to better estimate the shipping mode. Likewise, the way that the logistic decisions are made in the fourth step could affect the supply chain in the second. While this study provides an overall picture of the freight transportation model and introduces the main components, the desired model structure should be chosen based on the data limitation and scope of the project. Compromising between the ideal and feasible model structure in long and short runs would result in several possible structures with different levels of sophistication. In the following subsections, the main goal of each of the five modules is elaborated.

Firm Generation

Firm Generation is a very fundamental part of the framework. The greater the information collected for the firms, the more accurate the other modules will perform. The information that enhances the model to better capture the firm's behavior is not limited. Firms' critical characteristics and the amount of incoming and outgoing goods should be known by the end of this step. Usually the complete list of the firms with their detailed information is not available for the whole country, but the number of firms, aggregate tonnage and dollar value of the incoming and outgoing commodities, annual turnover, etc could be obtained at the regional level. Disaggregating the information from the regional level into the firm level can be performed according to the accepted synthesizing methods. Iterative Proportional Fitting (*19*, *20*) is an acceptable way when some information such as the number of employees is available at the firm level. Also some distributions may be parameterized on available data and then individual firms could be drawn from these distributions (*17*).

Supply Chain Replication

The idea of having such a module in the framework arises from the fact that a vast majority of the firms are part of many supply chains in the industry. Stadtler (*21*) defined the supply chain management as the "task of integrating organizational units along a SC and coordinating materials, information and financial flows in order to fulfil (ultimate) customer demands with the aim of improving competitiveness of the SC as a whole". He also illustrated the main objective and the primary components of SCM concept in a house, known as House of SCM. Improving the competitiveness of the supply chain is on top of the house, laying over the customer service;

meaning that the most important mean to achieve the primary objective of SCM is customer satisfaction. The house has two pillars, 'integration of organizational units' and 'coordination of flows' each of which made up of three blocks. He also described the essential factors that should be considered in making each block. Some important blocks that have a significant overlap with freight demand forecasting are "Choice of partner", "Leadership", and "Network organization". Other blocks such as "Use of information technology" are of less importance at this early stage of developing a behavioral freight demand model.

In this module, only the supply chains will be formed and the logistic decisions are to be made in the forth module. Although supply chain management includes manufacturing process, staging facilities, logistics, etc; this module only deals with the selection of the suppliers. It is important to note that the proposed framework assumes that decisions on the suppliers and logistics are sequential, while in real life, they may or may not be made together. Most important information that this module should feed into the model are about the way the firms build a chain and their incentives to choose another firm as the partner and also the role of each firm in the chain. As indicated by Stadtler (*21*) geographical aspects and financial position of the firms are among the deciding factors in choosing the partner. Also, financial power or particular knowledge of products and processes help the firms to take more dominant power in the supply chain. Beside aforementioned factors, supply chain formation is integrated with many other factors, part of which is to be discussed in the coming modules. For example, shipment characteristics, logistic decisions, and network condition should be considered in the formation of the chains. This could be considered by the recursive loops in the framework.

In order to establish basic rules for supply chain formation, some definitive information about the firms and also the commodity flow pattern should be provided. Ideally, the tonnage and the dollar value of different commodity types that either enter or leave the establishment are available. In addition to type and amount of incoming and outgoing commodities, the general transportation cost of the goods should be considered. At this point, a rough estimate of this cost is sufficient; however, a more accurate estimation will be obtained in module 4 and 5. As soon as the supply chains are formed, the commodity flows between the firms will be known. Other than the commodity flow between the respective firms, some firms' characteristics in the chain will be input to the next modules in order to find the dominant firm in the supply chain determining the logistic decisions. Especially, the choice of shipping characteristics (e.g. lead time, cost, and mode), warehousing, and consolidation and distribution decisions are influenced by the dominant firm in the chain. For example, a small manufacturing firm might have to keep a larger inventory if the distributing company controls the supply chain. Interaction between the firms can be captured in an agent-based model (ABM), in which firms behave according to a set of rules. An agent-based modeling approach simulates the actions and interactions of individuals in a network and their effects on the system (*22*). Discrete choice models also can be used as an alternative for the ABM. In this case, a choice set should be defined for each firm, and the partners are to be chosen using a random utility functions (*17*). Furthermore, machine learning models such as artificial neural network (ANN) models can be utilized to capture the supply chain behavior (*23*).

Shipment Forecasting

After the total annual shipments in each supply chain are determined in the second module, the shipment size, or the frequency of shipments are to be set in this step. To obtain more realistic results for the shipments, this module should be interrelated with the supply chain replication. Since the total annual shipment is the product of shipment frequency and the shipment size, there is no need to formulate both in the model. Size of shipments directly influences the transportation costs and the inventory cost. Obviously, large shipments increase the capital cost for the inventories while small shipments require more frequent deliveries with higher transport and stock out costs. Dominant firms in the supply chains find the optimal inventory strategies to decide on the reorder point and consequently the frequency of shipment. The suppliers have to meet the requirements set by the dominant firm, no matter it is optimal for them or not. The main goal of this module is to break down the annual shipments within a supply chain into separate shipments. By the end of this module, every single shipment will be recognized. In addition, some commodity specifications in the shipment and the attributes of the supply chain in which the commodity is being traded will be perceived. This part of the framework could be modeled in the form of an optimization problem, or either a discrete choice or ANN model. Optimizing the shipment size needs more detailed information, while econometric or ANN modeling approaches could be taken in the absence of accurate data. In the optimization approach, a general cost function should be defined and minimized by the dominant firm in the supply chain, which can be challenging for complex supply chains. On the other hand, for the discrete choice or ANN model, a few types of shipment size should be defined and a random utility function should be calibrated based on the available data. Another alternative modeling approach for shipment forecasting is a rule-based or computational process model (*24*) which attempt to represent the decision making process itself, rather than modeling only the outcomes of the decision process as in an econometric model.

Logistics Planning

This module is the most crucial in the framework, and needs much more investigation in the future. Having the shipment characteristics and also the supply chain attributes, the logistic decisions are made in this module. Logistic decisions might include a variety of determinants based on the nature of the supply chain and commodity. The decisions that would ideally be covered in this step include the lead time, shipping cost, outsourcing, loading unit, use of consolidation and distribution centers, and mode of transportation. As this module is performed all the desired information about each single shipment will be available and ready for the network analysis module. Based on the scope of each project, a variety of models could be defined at this step. One of the most important and challenging models is modal selection. Different modeling approaches could be chosen at this step including econometric-based discrete choice or a machine-learning approach such as ANN. However, it appears that a rule-based or computational process model (CPM) can be also a useful tool in logistics planning. Much of the theory of the computational process model is based on work by Newell and Simon (*25*) in the development of the production system. The production system is a model of cognitive behavior which states that "individuals' choices are based on their cognition of their environment (*26*). This means that a cognitive process can be represented by a model which contains a decision-maker's memory, including knowledge of the market, the environment and the results of their interactions with others, rules which operate on that memory and some currently known information about the environment. This allows the decision maker to form some resulting thought or to take action, which is then added to memory.

Network Analysis

This is the final part of the framework in which the impacts of the freight transportation are observed in the entire transportation system. Depending on the nature of the study, variety of outputs might be of interest for the researchers. Traffic congestion, network safety, and environmental impacts are among the most common. However these pieces of information in addition to some shipping specifications might be used as the performance measures of a freight transportation network, a more comprehensive measure of performance is desired for such studies.

DATA

Having an accurate, comprehensive, and reliable dataset is a very fundamental part of the framework. This part of the paper discusses the data issue for the proposed freight transportation framework. After describing the ideal set of data for developing the model described above, some practical approaches are suggested to utilize the available datasets in the United States to satisfy the data requirements. However, freight data has always been an expensive piece of information, in some cases doing a complete or partial survey is unavoidable. An affordable establishment survey has been successfully conducted by the authors (*27*), to collect the required data. An in-depth selection bias was also performed on this survey to detect any potential distortion in the surveyed data (*27*).

Ideal Data Needs

Broadly speaking, four types of data are required for the proposed framework: information on the national level freight movements, business establishments, shipments and supply chains, and the freight transportation networks.

National Level Freight Movements:

In this part of the data, the commodity flow between each zone pair for each specific commodity type is sought. The dollar values, and tonnages figures are necessary, and shipping distances are desirable. One of the critical decisions that should be made at the very first step is the way that the commodities are categorized and how the zones are defined. Depending on the scope of the project, the zones might be defined at the county, metropolitan statistical area (MSA), state, or any other self-defined level. As with most models, the more refined the zones are, the better the model's ability to replicate the decision makers' behavior tend to be. However, the lack of data is a serious barrier against pursuing a high level of disaggregation. The same can be said for the commodity categorization.

In this step a dataset in which each record shows the zone of origin, destination zone, dollar value, tonnage, and shipping distance for a specific type of the commodity. A shipment-by-shipment data is not necessary at this step, and the annual aggregate numbers are sufficient for the framework. However, in order to account for the seasonality, the annual data should be disaggregated over smaller time intervals.

Business Establishments:

In each geographic zone, some information about the establishments is needed. Ideally, for each establishment, employee size, annual turnover, square footage, number of franchises, types and amounts of the commodities coming into and shipped out of the establishment by tonnage and dollar value are called for. The most challenging is obtaining accurate information on the commodities, since many firms work with a wide variety of goods.

Shipments and Supply Chains:

The reason that the supply chain data and the shipment data are put together under a single title is that one can construct, at a reasonable accuracy, the supply chain for a given firm based on the detailed information about the shipments. Since it is not possible to obtain information on all the shipments, some aspects of supply chain may not be revealed from a small sample of shipments. In this step the path through which each shipment has been transported should be recorded in detail. Ideally, a comprehensive dataset on the shipments is needed to capture the complexity of today's supply chains. This part of the data is the most undeveloped and expensive one and needs lots of investment and effort. The detailed specifications of the sending and receiving agents at different segments of the whole shipping process will provide some information on the firms that are forming the supply chain. For each acting agent in the whole shipping process, some information including the primary activity, employee size, annual turnover, establishment square footage, number for franchises, etc are of interest. In addition to information about the shipping agents, the shipment characteristics and shipping specifications are desired. The former include, for example: weight, value, dimensions, time sensitivity, commodity type, origin, and destination of the commodity, and the latter may be comprised of haul time, cost, mode, and damage risk of the shipping process.

Freight Transportation Networks:

In order to load the freight traffic flow over the network, detailed specifications of the network for each mode of transportation is needed. The freight network information might be divided into two parts. Firstly, the nodes such as ports, terminals, consolidation and distribution centers, and warehouses should be determined. Also, information regarding the commodities entering and leaving each node, and also the facilities and the modes available at the node should be obtained. Secondly, road, rail, water, and air networks' information should be obtained. Travel time and cost information about available modes for each link of the network is the most critical. In urban freight studies, the congestion effect should be observed and embedded in travel cost estimations.

Available Datasets

In this subsection, publicly available datasets that have a potential to fulfill data needs for the four broad types of information: the national level freight movements, business establishments, and freight transportation networks, are reviewed.

National Level Freight Movements:

The Federal Highway Administration's Freight Analysis Framework (FAF) contains four sets of commodity movements' data across the United States (*28*). The datasets have the commodity value and weights for domestic and foreign trade activities, organized by the origin and destination zones, mode of transportation, and type of the commodity. FAF draws heavily from survey date such as the Commodity Flow Survey, Transborder Freight Transportation Data, and Rail Waybill samples. Although FAF is the most comprehensive survey-based data that are available publicly, its level of disaggregation, specially defined 114 domestic regions, 17 international gateways, and 7 international regions, may not be sufficient for some applications. Commodity's and geographic zones' method of categorization make this data difficult to use for regional studies. Thus, the possible application of disaggregation methods to the FAF dataset should be examined. For a national level freight study, however, this dataset provides valuable and creditable information and lays a reliable and justifiable foundation for further steps.

Business Establishments:

County Business Patterns (CBP) is an annual series of data that provides the number of establishments, number of employees, and payroll data by industry classification (*29*). The data is provided for different geographic zones ranging from country to ZIP code levels and also from an aggregate to a fairly disaggregate level of industry type. The problem with the CBP's disaggregate dataset is that a considerable number of values are not released. However, the missing values could be approximated using the conventional methods, such as iterative proportional fitting (IPF). Since most of the aggregate numbers are provided for larger geographic areas and also for larger industry classifications, IPF is a promising approach to address the issue of missing values (*20*). The main problem with this dataset is that the number of employees and payroll data are not reported for each specific firm. Because of the confidentiality issues, these numbers are reported for a group of establishments having the same geography and industry type. When the number of establishments drops below a predefined value, the numbers are not reported although they may be available at a more aggregate level. Unfortunately, some useful information such as establishment square footage and number of franchises is not available in this dataset and the employee size and annual turnover need be approximated.

Another piece of information that is of great importance is the amount of different commodity types being transported to and from the establishment. Fortunately, there is another publicly available dataset that provides this information in the US. The input-output accounts (*30*) provide general information on the type of industries that either make or use a specific type of commodity. It also provides some information on the values of the required commodities to produce a unit output of an industry. There are two main problems using this dataset. First of all, the figures are the average values for the entire country and don't capture the geographical heterogeneity. Also, the pattern of commodity use and production is not the same for all the firms within a particular industry sector, and this pattern should be provided at more disaggregate levels. Another problem is that for the warehousing sector, the figures reported in the input-output table represent the amount of value-added operations performed at facilities, instead of the value of the goods being stored or transported through.

Although input-output account may have drawbacks, considering the resources required to collect data on economic activities throughout the country, it provides rich information at a reasonable cost. However, a supplemental survey will be discussed to partially resolve this issue.

Shipments and Supply Chains:
To the best of authors' knowledge, there is no creditable public data on these items in the United States. Since this information has a great influence on the final model results, an affordable and practical approach to do the survey and collect this data is presented in the next subsection.

Freight Transportation Networks:
North American Transportation Atlas Data (NORTAD) has been made available to the public by the Bureau of Transportation Statistics (*31*). Two network databases of federal-aid roads, railway mainlines, railroad yards, and major sidings in the U.S. are provided. In addition, some point databases of: public use landing facilities, major highway-rail intermodal freight facilities, waterway and marine terminals, and highway/rail transfer facilities in the United States are available. Furthermore, for the border crossings, there are two point databases including highway border crossing facilities between the United States and Canada or the United States and Mexico and also rail border crossing facilities between the United States and Canada or the United States and Mexico.

Freight Survey

As discussed earlier, the primary missing data is the information on the shipments and supply chains. Those data are extremely difficult to collect due to the reluctance of the firms to share the information of such critical importance. An affordable survey with a mandatory selection bias analysis was conducted by the same research team (*27*) to satisfy the data needs. This cost-efficient experience could be easily transferred to other freight microsimulation studies. In this survey, for each firm, a person with the comprehensive knowledge of the firm's supply chain and transportation activities, for example, the logistics manager, needs to be surveyed, which makes obtaining a satisfactory response rate a serious challenge due to their high value of time.

Depending on the budget and time limits, different methods could be used to conduct the survey. In many cases, the most expensive one with the richest information is an in-person interview. Approaches such as phone interviews, mail-in and mail-out forms, and online surveys are other options. Generally, a group of well-trained interviewers has a better chance of convincing the interviewees to participate, given their better communication tools and exclusive time with the participants. The on-line survey could be generally performed in a more cost-effective way and could take advantage of a variety of audio and visual stimuli to enhance the survey questions. Since the response rate in this kind of survey is extremely low, some information must be obtained from non-participants, in order to assess the presence and the severity of non-response bias. If such patterns exist, some instrumental variables should be applied to the models with essential heterogeneity (*32*).

Table 1 provides a list of potential questions that could be helpful in the data collection step of the proposed freight study framework.

TABLE 1 Freight Survey Sample

Establishment Information
Zip code of the establishment. Total gross floor area occupied by the establishment. Number of employees? Primary industry of the establishment. Potential use of each transportation mode. Access to rail-truck inter-modal terminals. Warehousing situation in the company.
Recent shipments (for every single shipment)
Origin and destination. Mode of transportation used for the shipment. Type, value, weight, and volume of the commodity. Cost and time of the entire shipping process. Whether the shipment was Inbound / Outbound / Import / Export / Containerized / Damaged Use of consolidation center, distribution center, or warehouse for the shipment. Decision making unit (sending firm / receiving firm / a 3PL) Whether the same transportation mode was preferred TWO years ago for a similar shipment.
Optional Contact Information
Company address, phone, and email. Respondent’s position in the company. Survey evaluation (Friendly / Neutral / Unfriendly).

The table consists of three following parts: establishment information, outbound shipment, and inbound shipment information. Questions 1 to 7 could be asked once and the other two groups of questions may be repeated for as many shipments as the participant is willing to provide the information for. Multiple choice and drop-down lists (for an on-line survey) should be provided for most of the questions to simplify the response process and reduce the amount of manual data entry. This table provides a general idea about the type of questions that should be included in a freight survey, however other recommendations and standards for survey design should be observed to boost the response rate.

Data for Validation

A portion of the data should be set aside for validation purpose. This part of the data is needed to validate the outputs of the models, and cannot come from the same source that was used to calibrate the models. The proposed framework contains different types of outputs, from individual firm information to aggregate modal splits. Generally, truck ADT on links, shipment volumes at terminals, and mode share for different final products may be used to assess the models' accuracy. Also, road-side intercept survey of trucks may be the most efficient way to obtain such data for validating truck shipments.

CONCLUSION

The four step freight modeling approach, originally developed for the passenger transportation models, is not able to capture the rapid changes in structure of supply chains and freight transportation systems. In order to resolve this problem, two main steps should be taken by both researchers and practitioners. The first step is to introduce a novel modeling framework that incorporates the new supply chain components and accounts for the firms' behavior as the decision making unit. In the second step, the data needed to build and calibrate the models of the proposed framework must be obtained from various sources.

In this exploratory study, a freight modeling framework was introduced and the data needs and potential modeling approaches were discussed. The proposed framework has five modules, including *Firm Generation*, *Supply Chain Replication*, *Shipment Forecasting*, *Logistics Planning*, and *Network Analysis*. The main characteristic of the framework is that the role of firms, as the decision making agents, has been taken into consideration. The ideal data at the national level was discussed and the available datasets in the U.S. were explored, in addition to a suggested survey to cover the data gap.

Compared to the passenger travel behavior studies, research on freight movements is still at its infancy and in a dire need of support of the government and academia. Freight studies are in the transition stage from the conventional four-step models to a well-defined behavioral based freight framework, and the data gap, as the biggest barrier, should be resolved.

Acknowledgment
The authors would like to acknowledge the National Center for Freight, Infrastructure, Research and Education (CFIRE) at the University of Wisconsin-Madison and Illinois Department of Transportation for funding this research.

REFERENCES
1. Rodrigue, J. P., *Commodity Chains and Freight Transportation*, Chapter 5, Concept 3,in The Geography of Transport Systems, by J. P. Rodrigue, C. Comtois and B. Slack, New York: Routledge, 2006.
2. AASHTO, Invest in America- Freight Rail Bottom Line Report. January 2003.
3. Hensher, D., and M. A. Figliozzi. Behavioural insights into the modelling of freight transportation and distribution systems. *Transportation Research Part B*, Vol. 41, 2007, pp. 921–923.
4. McNally, M. G., "The Activity-Based Approach" (December 1, 2000). Center for Activity Systems Analysis. Paper UCI-ITS-AS-WP-00-4.
5. Ettema, D.F., and H.J.P. Timmermans. Activity Based Approaches to Travel Analysis, Elsevier Science, 1997, Tarrytown, New York.
6. Anderson, M.D., A.D. Youngblood, and G.A. Harris. A Statewide Freight Flow Model for Alabama. Transportation Research Board 86th Annual Meeting. Washington D.C., 2007.
7. Auld, J.A. Analysis of National County-Level Heavy-Duty Freight Truck Emissions, M.Sc. Dissertation. University of Illinois at Chicago, USA, 2007.
8. Cunningham, W. H. J. Freight Modal Choice and Competition in Transportation: A Critique and Categorization of Analysis Techniques. *Transportation Journal*, Vol. 21, No. 4, pp. 66-75, 1982.
9. Holguin-Veras, J. Revealed Preference Analysis of Commercial Vehicle Choice Process. *Journal of Transportation Engineering*, Vol. 128, No. 4, pp. 336-346, 2002.
10. Norojono, O., and W. Young. A Stated Preference Freight Mode Choice Model. *Transportation Planning & Technology*, Vol. 26, No. 2, pp. 195-212, 2003.
11. Baumol, W.J., H.D. Vinod. An Inventory Theoretic Model of Freight Transport Demand. *Management Science*, Vol. 16, No. 7, pp. 413-421, 1970.
12. Harker, P.T. and T.L. Friesz. Prediction of Intercity Freight Flows, I: Theory. *Transportation Research B*, Vol. 20, No. 2, pp. 139-153, 1986.
13. Proussaloglou, K., Y. Popuri, D. Tempesta, K. Kasturirangan, and D. Cipra. Wisconsin Passenger and Freight Statewide Model: A Case Study in Statewide Model Validation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2003, pp. 120–129, 2007.
14. de Jong, G.C., M. Ben-Akiva. A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B*, Vol. 41, pp. 950–965, 2007.
15. Transportation Research Board. Forecasting Statewide Freight Toolkit, National Cooperative Highway Research Program: Report 606. Washington D.C., 2008.

16. Hunt, J.D., K.J. Stefan, and A.T. Brownlee. Establishment-Based Survey of Urban Commercial Vehicle Movements in Alberta, Canada: Survey Design, Implementation, and Results. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1957, pp. 75–83, 2006.
17. RAND Europe, Solving, INRO. The specification of logistics in the Norwegian and Swedish national freight model systems, model scope, structure and implementation plan. Report for the NTP and Samgods group, TR-225-SIKA, RAND Europe, Leiden, 2004.
18. Samimi, A., A. Mohammadian, and K. Kawamura. A Disaggregate Analysis of Rail-Truck Mode Choice Behaviors for Freight Shipments. Submitted to the 89th Annual Transportation Research Board Meeting, Washington, D.C., 2010.
19. Arentze, T., H. Timmermans, and F. Hofman. Creating Synthetic Household Populations: Problems and Approach. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2014, pp. 85–91, 2007.
20. Auld, J., and A. Mohammadian . Population Synthesis with Region-Level Control Variable Aggregation, in the Proc. of 87th Annual Meeting of the Transportation Research Board , (CD-ROM), Washington , D.C. January 2008.
21. Stadtler, H. Supply chain management and advanced planning - basics, overview and challenges. *European Journal of Operational Research*, Vol. 163, pp. 575-588, 2005.
22. Davidsson, P., Henesey, L., Ramstedt, L., Tornquist, J., and Wernstedt, F., An Analysis of Agent-Based Approaches to Transport Logistics, *Transportation Research Part C: Emerging Technologies*. Vol. 13 No. 4, pp 255-271. 2005.
23. Miao, X., and B. Xi. Agile Forecasting of Dynamic Logistics Demand. *Transport*, Vol. 23 No. 1, pp 26-30. 2008.
24. Schmedding, D., G.T. Liedtke, and J. Babani. The Rule-Based Freight Transport Simulation System InterLOG, in the Proc. of 85th Annual Meeting of the Transportation Research Board, (CD-ROM), Washington D.C., January 2006.
25. Newell, A. and H.A. Simon. Human Problem Solving. Englewood Cliffs, NJ: Prentice Hall. 1972.
26. Ettema, D.F., and H.J.P. Timmermans. Theories and Models of Activity Patterns. In Ettema, D.F. and Timmermans, H.J.P.(eds.): Activity-Based Approaches to Travel Analysis. Oxford: Pergamon Press. 1997.
27. Samimi, A., A. Mohammadian, and K. Kawamura. An Online Freight Shipment Survey in US: Lessons Learnt and a Non-response Bias Analysis. Submitted to the 89th Annual Transportation Research Board Meeting, Washington, D.C., 2010.
28. Federal Highway Administration. An Overview of the 2002 Commodity Origin-Destination Database: Methodology and Data. Freight Analysis Framework (FAF), Report No.1 (R1), 2006.
29. U.S. Census Bureau, Economic Planning and Coordination Division. County Business Patterns. Available at: http://www.census.gov/epcd/cbp/view/cbpview.html, accessed April 2008.

30. Bureau of Economic Analysis (BEA). Annual Input-Output Accounts of the U.S. Economy. Available at: http://www.bea.gov/industry/io_annual.htm, accessed May 2008.
31. Bureau of Transportation Statistics (BTS). North American Transportation Atlas Data. Available at: http://www.bts.gov/publications/north_american_transportation_atlas_data/, accessed June 2008.
32. Heckman, J. Varieties of Selection Bias. *American Economic Review*, Vol. 80, No. 2, pp. 313-318, 1990.