**INTRODUCTION**

Activity based modeling approach is becoming popular in passenger transportation planning as a way to address shortcomings of the traditional four-step approach that includes: trip generation, distribution, mode choice, and traffic assignment. In contrast, freight transportation planning still relies mostly on the four-step approach which is even less suited for capturing the movement of goods. The use of four-step approach results in ignoring the critical role of individual firms, as the decision making agents of the shipping process, in a vast majority of the freight demand models. There are also other factors such as: deregulation of freight industries, increasing trend of supply chain management application by the business establishments, and the absence of disaggregate freight data that might have magnified the shortcomings associated with the use of the four-step modeling framework.

The deregulation of freight industries in the early 80’s promoted the companies to aggressively adopt the supply chain management techniques to survive in a competitive market (Rodrigue 2006). This was accelerated and also enabled by the globalization and the use of information technology. Eventually, businesses were able to find efficient ways to produce and/or distribute commodities. In fact, in the U.S., the share of the logistics-related expenditure of the GDP dropped from about 17% in 1980 to just above 10% in 2000 (AASHTO 2003).

The way that the logistic decisions are made within a production cycle directly affects the cost of production by influencing the transportation cost for the raw materials and the semi-finished goods. Similarly, the distribution cost for finished goods is optimized within a well-organized distribution system. This could bring the goods from the producers to the consumers at lower overall cost, causing a decrease in retail store prices (Rodrigue 2006). On the other hand, such strategic decisions could not be effectively captured in the model without having disaggregate data on individual firms’ shipping behaviors. Such data; however, is extremely challenging to collect, since most shipping managers are unwilling to share the company’s information on the shipping process. They also tend to have frantic schedules that not only diminish the likelihood of answering the survey, but also make the survey costly. Fortunately, there are periodical commodity freight surveys (CFS) and other public freight data that might have magnified the shortcomings associated with the use of the four-step modeling framework.

**ABSTRACT:** The framework of freight demand modeling has been the conventional four-step approach, in the past. Activity-based models that better capture the behavior and decision-making process are emerging in the passenger travel modeling to address the drawbacks of the four-step approach. This paper proposes an activity-based freight modeling framework and discusses its data needs. The framework consists of five modules: firm generation, supply chain replication, shipment forecasting, logistic decisions, and network analysis. Similar to the activity-based approach in the passenger travel modeling, in which individuals or households are the agents of the models, individual firms or a group of firms with similar characteristics, are the main players in the proposed framework. Public data sources that could be used in the U.S. to make this framework running are also introduced and a supplementary establishment survey is proposed as a cost-effective way to satisfy the data needs of the framework.

**KEYWORDS:** Activity-based approach, freight demand forecast, behavioral models.
data in some countries including the U.S., which provide aggregate level of freight transportation information. This can be used along with supplemental data collection efforts to fulfill disaggregate data needs at a reasonable cost.

This study attempts to introduce an activity-based freight movement microsimulation framework that incorporates the supply chain management concepts and highlight the critical role of individual firms as the primary decision makers. The development of such model would be a momentous achievement in freight modeling, since it should significantly improve forecasts and the evaluation of policies by incorporating individual firms’ adaptive reactions. To fulfill disaggregate data gap at a reasonable cost, an online establishment survey is also proposed and conducted in the U.S., a general overview of which is discussed in this study.

BACKGROUND

The four-step freight modeling framework consists of four sequential modules and are the primary approach for freight demand forecasting, especially by metropolitan and statewide planning agencies (Southworth 2003, Cambridge Systematics 1995). Generation and attraction of commercial trips are usually performed based on the zonal economic activity or employment (Anderson et al. 2007). Although information on the economic activity of an industry is difficult to obtain, there are some publications that provide an average rate of commercial trip generation and attraction for freight planners (Fischer and Han 2001). The distribution of commercial trips is also commonly carried out by a gravity model with shipping distance as the impedance (Auld 2007). Southworth (2003) has discussed different approaches of commercial trips distribution, including spatial interaction (SIA) method, in more details. Mode choice is a critical component of the framework and used to be estimated based on the shipping cost in the earlier models (Cunningham 1982). However, other determining factors such as velocity, reliability, and quality of service have also been considered in freight random utility models (Holguin-Veras 2002). A freight study in France (Jiang et al. 1999) also distinguished between long-term and short-term cost factors on modal selection and took into account the difference in private and for-hire shipping behaviors. Freight traffic assignment has been performed in a fairly straightforward way, especially for intercity traffic, which typically employ “all-or-nothing” method. Nonetheless, many four-step models have a fifth module to convert the commodity flow into vehicle flow, before performing the traffic assignment (Fischer et al. 2000). Urban freight traffic; however, is usually assigned to the cheapest or quickest path in conjunction with the base traffic when converted to passenger vehicle equivalent.

Baumol and Vinod (1970) are among the pioneers in modeling both mode choice and demands for links on a freight network. They utilized the same approach that had been developed for the analysis of passenger transportation. Their mode choice model considers the trade-off between the transportation cost, time, reliability, and safety, and also accounts for the carrier and commodity heterogeneity. Harker and Friesz (1986) also applied the conventional four-step approach with substantial modification to the supply and demand models. Proussaloglou et al. (2007) developed a freight demand model for Wisconsin in a four-step framework. Twenty five commodity groups were defined for 72 Wisconsin counties and 74 zones for the rest of the U.S. and North America. Employment data were used for trip generation and attraction for each commodity category. The 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 share for each commodity was used instead. Using the total daily truck trip table, freight traffic assignment was performed in conjunction with the daily and long distance passenger trips. A recent study sponsored by the American Association of State Highway and Transportation Officials in cooperation with the Federal Highway Administration (Transportation Research Board 2008) is a comprehensive source for the freight demand models in the U.S. that covers recent studies and data collection efforts.

Behavioral freight models are extremely scarce in the literature and a limited number of such studies could be found among the recent works. Companies have become increasingly customer-order-driven and new production systems such as Just-in-Time (JIT) are now common. In a comprehensive European freight study, de Jong and Ben-Akiva (2007) stated that almost all the existing freight transportation studies are missing supply chain and logistics components. They provided valuable insights in freight demand modeling by introducing some behavioral models in which the firms’ characteristics are incorporated in the 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. Hensher and Figliozzi (2007) also highlighted the importance of disaggregate behavioral freight models in mitigating traffic congestion and maintaining the efficiency and reliability of the freight transportation system.

Availability of data or lack thereof, is the main barrier in developing behavioral freight demand models and is a major challenge in almost all the freight studies. While secondary data might exist, access to such data is limited only to highly aggregated information that is not sufficient for the behavioral model development. On the other hand, disaggregate
freight data are so valuable that some companies are in the business of collecting such data. Furthermore, qualified potential participants in such surveys tend to have a high value of time and are often reluctant to reveal any data that may be considered as their business secrets. This significantly affects the participation rate and increases the cost of collecting disaggregate data on freight shipments compared to other types of surveys. There have been some local freight surveys in North America, such as an establishment survey in Alberta, Canada in which around 7,300 establishments were interviewed (Hunt et al. 2006). However, lack of disaggregate freight data in the U.S. appears to be a serious challenge for researchers and decision-makers, restricting the development of behavioral freight demand models.

**MODELING FRAMEWORK**

In order to make the model results more realistic and behavioral, the decision making agents should be effectively incorporated into the model. A firm, a shipment, or even commodity flow between zones are introduced as the observation unit in former freight models (de Jong and Ben-Akiva 2007). However, individual firms are the actual decision making units and could provide a more realistic and possibly more accurate modeling results. Incorporating the behaviors of the firms in the freight transportation model is the essence of the disaggregate freight models, and has been practiced by very few researchers including de Jong and Ben-Akiva (2007).

Freight Activity Microsimulation Estimator (FAME) is the proposed freight activity-based modeling framework with five basic modules (Figure 1), which will be discussed in detail in this section. 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 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, etc are made. Even though sophisticated firms make decisions on the physical infrastructure of the supply chain and logistics strategies simultaneously; in our proposed approach, 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 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 determine the shipping mode. Likewise, the way that the logistic decisions are made in the fourth step could affect the supply chain formation. Also, general cost of commodity transportation from the last module could be fed back into the second, third, and forth modules, and iterates until a stabilized set of commodity flows and costs are obtained. However, appropriate modeling framework should be chosen and modules should be coupled based on data availability and also the scope of the project.

**Module 1: Firm Generation**

The first and foremost step of the framework is to generate the firms and their essential characteristics. Although the characteristics to be considered are not limited and may vary in different studies, some critical ones such as industry type, location, and size of the establishment are of great importance. Other information such as annual turnover, square footage of the location, credit score, etc could be also helpful. Basically, the more information a modeler collects at this stage, the greater the flexibility in other modules that are developed subsequently. Usually a complete list of the firms with their detailed information is not available for the entire country, but number of establishments in a fairly smaller area with specific industry type and employee size may be publicly available. Synthesizing all the single firms in the U.S., with every single shipment that they may make in a fiscal year, will results in a gigantic dataset with a heavy computational burden. Therefore, some form of aggregation is inevitable. We propose to aggregate the firms based on firm-types. A firm-type is a collection of firms with similar location, industry type, and establishment size. This will not only keep the computational burden at a reasonable level, but also diminishes the need for highly disaggregate data that is not available in almost all the large-scale freight studies. In this module, critical specifications of the firm-types should be generated together with their annual incoming and outgoing commodity tonnage and value.

**Module 2: Supply Chain Replication**

This module is the most underdeveloped part of the framework and has a significant role in connecting the firm-types and identifying the sender and receiver of shipments. Supply chain configurations play a critical role in the model outputs since all the firms are part of such chains in the industry. Supply chain management seeks to improve competitiveness and thereby profitability of the industry by fulfilling
customer satisfaction (Stadtler 2005). This is achieved in two major ways: integration of organizational units, and coordination of flows (Stadtler 2005). Some basic factors such as the choice of partners and the leadership of the supply chain could be looked into at this stage of developing a behavioral freight model. However, other factors such as use of information technology could be investigated. It should be mentioned; however, that although some supply chain related decisions are made along with the logistics strategies, they are modeled separately in this framework mainly due to data limitations. The potential effects of supply chain decisions on logistic strategies or vice-versa could be captured by recursive loops in the framework. At this stage, we only focus on the supplier selection rules and the way that companies form a supply chain. Stadtler (2005) indicated that the locations and financial positions of the companies are among the most important elements in the supplier selection decisions. In addition to the aforementioned factors, supply chain formation is integrated with many other factors. For example, shipment characteristics, logistic decisions, and network condition should be considered in the formation of the chains. Some of these decisions are captured in the subsequent modules, and the recursive loops in the framework can be used to establish connections among them.

Within this module, basic rules for supplier selection process need to be developed. Agent-based models are typi-
cally used in such cases where agents (in our case firm-types) behave in accordance with a set of rules (Davidsson et al. 2005). Discrete choice models and artificial neural networks (Miao and Xi 2008) are also other alternatives for supplier choice process. Location, industry type, and size of each firm-type from the first module along with their annual incoming and outgoing commodities should be used as inputs to develop models or rules to perform the supplier selection process. Determining the dominant firm in the chain is another important task that could be performed in this module. Characteristics of the dominant firm usually affect the shipping decisions and behaviors. For example, a small manufacturing firm might have to keep a larger inventory if the distributing company controls the supply chain.

Module 3: Shipment Forecasting

Annual shipments in each supply chain that are estimated by the former module are broken down into single shipments at this step. In other words, shipment sizes or frequencies of shipment should be determined in this module. At the end of this module, every single shipment will be recognized. In addition, some commodity specifications of the shipments and the attributes of the supply chain in which the commodity is being traded will be modeled.

Integrating this module with the former one could make the result more realistic, as the shipment size determination is closely related to the inventory decisions. There is a trade-off between transportation costs and inventory costs that is influenced by shipment size. Smaller shipments tend to have higher transportation and stock out costs but lower inventory costs. Shipment size and optimal inventory policies are usually determined by the dominant firm in the chain, and the suppliers have to meet the requirements set by the dominant firm, regardless of whether it is optimal for them or not.

Size of shipments may be determined by solving optimization problems, especially if the detailed information about the supply chain and their specifications are available. In this approach, a general cost is introduced as a function of shipment size and minimized to find the best shipment size for the dominant firm in the chain. Such optimization problems could be very complex in real supply chains and require many simplifying assumptions. In the absence of detailed information, discrete choice model could be applied to divide the size of the shipments into a few categories and calibrating a random utility function for some observed shipments. Another alternative modeling approach for shipment forecasting is a rule-based or computational process model (Schmedding et al. 2006) which attempt to represent the decision making process itself, rather than modeling only the outcomes of the decision process as in an econometric model.

Module 4: Logistics Planning

Many significant logistic decisions could be made in this module including determining mode of transportation, use of consolidation center, distribution center, outsourcing, loading unit, etc. Again, having an interrelated framework with recursive loops between the second and fourth modules makes the modeling results more realistic and may be preferred in some studies. However, a basic form of the framework with no interrelation between the modules is taken at this early stage of developing a behavioral freight model. The critical part of this module is modal selection. As discussed earlier, many previous mode choice models have only considered general cost of good movement as the main determinant of mode choice. However, other factors such as type of firm, supply chain, and nature of geography have shown significant roles in mode choice decisions (Southworth 2003). A few researchers have tried to simultaneously determine the transportation mode and shipment size, and have argued the benefits of this approach as well (McFadden et al. 1985, Vernimmen and Witlox 2003). However, data limitations may prevent one to combine the third and forth modules. Different approaches may be chosen to determine modal split, among which econometric-base models are most common. However, machine learning approaches and rule-based models could also be useful in this module. Much of the theory of the rule-based models is based on work by Newell and Simon (1972) in the development of a production system. The production system is a model of cognitive behavior which states that individuals’ choices are based on their cognition of their environment. 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 strategies or to take actions, which is then added to the memory.

Information on every single shipment (e.g. weight, value, origin, destination, commodity type) along with some critical characteristics of the decision makers (e.g. location, access to each mode of transportation, establishment size) are provided from the previous steps and should be used in this module to determine the logistic decisions. Different models could be fed into this module, based on the scope of each study and data availability.

Module 5: Network Analysis

The effects of freight movement on the entire transportation network should be evaluated at this stage. Major characteristics of all the individual shipments are revealed from the pre-
vious modules, and many outputs of interest can be obtained depending on the scope of the study. A challenging task that should be performed in this module is to convert the commodity flow to vehicle loads. Variety of vehicle sizes in each mode, and the fact that a significant portion of vehicles are empty or partially loaded, make this task more challenging. Having the vehicle traffic volumes, traffic assignment should also be performed to determine traffic flow in each route. Freight traffic assignment is commonly performed in an all-or-nothing approach for intercity movements, while more sophisticated methods should be applied for urban freight traffic assignment in conjunction with the base traffic load of the network. Finally performance measures have to be calculated to provide estimates of travel time and mobility, safety, environmental impacts, etc. These values may be used to estimate the general cost of good movement, and fed back into the previous modules, and iterates until a stable solution is obtained.

DATA

An accurate, comprehensive, and reliable dataset is a very fundamental part of the framework, and the lack of such data could make the study fairly pricey or unfruitful. Cost and sensitivity of freight data can urge the researchers to use the public data and minimize the need for a supplementary survey. This section of the paper elaborates data needs of the framework and suggests some public data sources in the U.S. that could be utilized in developing the proposed modeling framework. A cost efficient establishment survey, which has been successfully conducted by the research team, is also briefly discussed to satisfy the data needs (Samimi et al. 2010). Broadly speaking, four categories of data are required for the proposed framework: aggregate freight movements, information on business establishments, information on individual shipments and supply chains, and specifications of the transportation networks.

A critical decision 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. Module one also needs some information about the establishments in each geographic zone to synthesize the firm-types. As mentioned earlier, location, employee size, and industry type of the establishments are necessary and other information such as square footage and annual turnover of the establishment could be helpful as well. County Business Patterns (CBP) is a publicly available dataset that serves this purpose (U.S. Census Bureau, Economic Planning and Coordination Division 2008). 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 due to confidentiality issue. 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 (Auld et al. 2009).

As part of the second module, annual value and tonnage of different commodity types that are traded between the zone pairs are needed. The Federal Highway Administration (2006) has utilized many freight data sources including but not limited to Commodity Flow Survey, Transborder Freight Transportation Data, and Rail Waybill samples to develop the Freight Analysis Framework (FAF). This set of data has the tonnage and value of each commodity type that is transported between all the zone pairs for each mode of transportation. Even though FAF is the most comprehensive public freight data, it has a few limitations that make it insufficient for some specific applications. One major drawback is the level of geographical aggregation. United States is divided into 114 domestic regions, 17 international gateways, and 7 international regions, that make this data difficult to use for regional studies. Although possible application of disaggregation methods to the FAF dataset has been examined to resolve this issue, no credible disaggregate FAF data has been introduced to the researchers yet. Other information that is needed in the second module is types and amount of commodities that are used and produced in every industry. The input-output account is a public dataset that provides this information in the US (Bureau of Economic Analysis 2008). It also provides some information on the values of the required commodities to produce a unit output of an industry. The values, however, are reported for the entire U.S. and geographical heterogeneity is not considered. There are county-level input-output data available from commercial vendors, but the data are imputed from the national data and the accuracy of the county-level data is unknown. Although the input-output account may have downsides in geographical and industry type aggregation level, considering the resources required to collect data on economic activities throughout the country, it provides rich information at

58 Transportation Letters: The International Journal of Transportation Research
A behavioral freight movement microsimulation model: method and data

A portion of the data should also 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. FAME 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.

**UIC National Freight Survey**

As discussed earlier, the primary missing data is the information on the shipments and supply chains. Those data are extremely costly to obtain mainly due to the sensitivity of the information and high time value of the qualified participants. The authors have carried out a cost-effective establishment survey in the U.S., whose general framework and method could be transferred to similar studies. The survey was conducted on-line, reaching shipping manager or someone with a comprehensive knowledge about the shipping decisions made within the company. This, of course, makes obtaining a satisfactory response rate a serious challenge due to their busy schedule. 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. The on-line survey could be generally performed in a more cost-effective way; however, the low response rate in this kind of survey could be a serious problem and should be appropriately addressed.

The UIC survey had three major sections. Some basic information about the establishment was asked in the first section. This includes the location, employee size, value of total annual shipments, potential use of each freight transportation mode, access to intermodal facilities, and warehousing situation of the establishment. The second section of the survey inquired about five recent shipments each respondent was familiar with. The information about the shipments includes origin, destination, mode of transportation, commodity type, value and weight of the shipment, haul time and cost of the whole shipping process, use of consolidation center, distribution center and/or warehouse, and decision making unit for each shipment. And finally, optional contact information of the respondent along with an evaluation of the survey was included. Dollar value and tonnage of the surveyed shipments in UIC survey and Commodity Flow Survey (2002) are compared respectively in Figure 2 and 3.
Figure 2. Share of each Mode in Commodity Weight.

Figure 3. Share of each Mode in Commodity Dollar Value.
These figures reveal an acceptable match between these surveys especially for truck and rail. A detailed explanation of the survey method, descriptive statistics and selection bias analysis are not in the scope of this paper and are discussed elsewhere (Samimi et al. 2010).

CONCLUSION

The state-of-the-practice of freight demand modeling lags the advancements of freight transportation behaviors in the real market by a significant margin. Obsolescence of the theory behind the conventional four-step approach in freight demand modeling on one hand, and the lack of a comprehensive set of disaggregate freight data on the other hand, have made the process of developing a behavioral freight demand modeling framework extremely challenging. This study is a part of an extensive effort to transfer the activity-based modeling that is becoming more common in the passenger travel demand analysis to the freight demand modeling and introduce an activity-based microsimulation framework (FAME) for this purpose. The proposed framework has five modules: firm generation, supply chain replication, shipment forecasting, logistic decisions, and network analysis. The unit of decision making in this microsimulation framework is firm-types each consisting of a group of firms with similar location, size, and industry classification. This paper also discussed public data sources in the U.S. that can be used at no cost to satisfy part of the data requirements for this framework, and also suggested a supplementary cost-effective establishment survey to satisfy the data needs.

ACKNOWLEDGEMENTS

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

REFERENCES


Hensher, D. and Figliozzi, M.A. (2007). “Behavioural insights into the modelling of freight transportation and distri-
bution systems” Transportation Research Part B, 41(9), 921–923.


