A hybrid microsimulation model of urban freight travel demand

Rick Donnelly, Marcus Wigan & Russell Thompson*

July 15, 2010

The distribution of freight within large urban areas is a highly complex process involving a vast multitude of interactions between an equally large number of agents. Shippers, carriers, intermediaries, warehousers, consumers (businesses as well as households), and the government all dynamically exchange goods and information in order for this system to work. Planners and policy-makers have long sought to understand and predict these interactions and their impact on the urban transportation system. They have perhaps naturally turned to modeling the interactions, most often using sketch planning techniques or adapting the sequential modeling approach used for person travel modeling. Not surprisingly, such simplifications have not fared well, as they fail to capture some of the key dynamics of urban freight. In particular, they miss the high prevalence of trip chaining and the widespread use of distribution centers in the supply chains that generate freight movements. As these and other patterns have come to define urban freight existing models have been seen as increasingly deficient (Taylor & Button 1999, Hunt & Steffan 2007).

Our dissatisfaction with the status quo in freight modeling inspired the idea for an agent-based modeling approach capable of representing these complex interactions, their diversity, and their inherent variability. The chance to do so presented itself in Oregon, and the initial framework was originally developed as part of their Transportation and Land Use Model Integration Program (TLUMIP) almost a decade ago. The goal was to translate economic flows between firms, as well as firms and households, into freight flows by mode of transport. While the approach chosen was well founded in theory and current thinking, the data required to comprehensively estimate such a model were unavailable. Each known source of data only provided a narrow glimpse of the overall freight picture, and a means for fusing them into a holistic dataset suitable for model estimation seemed forlorn. Thus, it was decided to use a microsimulation approach that separately modeled each of the important decisions that characterize the demand for freight transport.

The initial model was known simply as CT. Another model in the TLUMIP suite estimated economic growth and allocated it to production, exchange, and consumption between firms and households. Many of these flows were mapped to commodities that were modeled in CT, which translated them into discrete daily shipments carried by specific vehicles. The shipments, which includes attributes such as departure and dwell times as well as origin and destination, were organized into truck tours. The stops on each tour were optimized using the well-known traveling salesman problem and then assigned to the network. The land use and regional economic component has since evolved into the PECAS modeling framework, which is being deployed both in Oregon (replacing its predecessor in the TLUMIP framework) as well as in other cities and states.

*Rick (donnellyr@pbworld.com) is a principal consultant at Parsons Brinckerhoff and Senior Fellow at the University of Melbourne. Marcus is a Professorial Fellow at the University of Melbourne, and Russell Thompson is a Research Fellow at Monash University.

The CT model was subsequently adapted to operate in a standalone manner, internally generating the economic drivers of demand rather than simply processing exogenously supplied production and consumption levels and flows. A deeper analysis of several data sources was completed, to include truck tour surveys completed earlier in Denver and Houston, and the model was overhauled to reflect the better understanding gained. Some new data sources, not available when CT was initially developed, were further mined. Donnelly (2007) used the resulting model as the basis for academic research. It explicitly models several important elements of freight demand and the factors influencing them through the interaction of semi-autonomous agents. Individual firms, importers, exporters, and carriers are all represented as agents of varying complexity that interact on several levels. An overall simulation environment provides global information exchange, access to information such as travel times and vehicle availability, the regional economy, and transportation networks, gateways, and intermodal and distribution centers. Some of the key dynamics modeled are summarized in Table 1. The individual components of the model include:

- *Economic drivers:* The total level of freight demand is modeled as a function of gross urban product by economic sector, which is the portion of gross state product attributable to the modeled urban area. The relationship between economic output and commodity generation implied in the Commodity Flow Survey (CFS), the PECAS model, and employment patterns are used to estimate the overall value — in annual dollar terms — of commodities produced within the region.

- *Modal alternatives:* Much of the freight locally generated is exported to receivers or intermediaries outside the modeled region. A mode choice allocation is carried out by commodity, as the value-to-weight ratio varies considerably by mode chosen. The total tonnages by mode and commodity are allocated to individual firm agents as functions of their share of employment in the applicable sector(s).

- *Trans-shipment:* The probability of the commodities produced or consumed by each firm being handled as a trans-shipment is calculated early in the model. This probability is a function of the commodity, the distance traveled, and indirectly, the primary mode of transport chosen. These probabilities were derived from the Canadian National Roadside Survey, the only known source of comprehensive data on trans-shipment by commodity. This choice is highly interrelated with the allocation of the flows to importer or exporter agents. Distribution centers, warehouses, and intermodal terminals are three common types of trans-shipment facilities.

- *Exports:* Some firms export some or all of their output. Exports are allocated to export agents, which include airports, rail yards, distribution centers, and abstract representations of freight forwarders. Trips from the local firms to the export agents (drayage) or external gateways are generated by the model. Input-output (IO) tables and coefficients are used to determine the share of goods consumed locally versus exported.

- *Imports:* Some of the internal demand will be satisfied by imports from other regions, which are handled by import agents operating essentially in the reverse of exporters. The importers compete with local firms to meet the requirements of each firm agent. As with the exporters, imports are drayed from their point of entry into the region (airport, rail yard, highway, etc.) to the import agent, which might also function as a distribution center. Input-output tables and coefficients define the share of goods imported versus those produced locally.

Table 1: Typology of model components

Mathematical equations (deterministic outcomes)
Estimation of gross urban product
Translation of gross urban product to (value of) commodities
Translation of value of commodities from annual value to weekly tons
Tour optimisation using traveling salesman problem (TSP) algorithm
Traffic assignment (EMME/2 multi-class assignment by period of the day)
Sampling from statistical distributions or generated by rules (stochastic outcomes)
Decision whether to ship when total value falls below threshold
Generation of discrete shipments from total tons shipped
Discrete choice of destination firm and its distance from shipper
Firm's choice of carrier
Incidence of trans-shipment (including distribution centers)
Choice of import and export agents
Carrier's choice of vehicle
Number of hauls (tours) per day
Selection of routing inefficiency factor

- *Shipment generation:* The total tonnage produced by each firm is converted into daily tonnage using observed and reported seasonality trends. Some firms do not ship each day, and so randomly choose whether to ship on the simulation day. The daily tonnage is divided into discrete shipments by sampling from observed distributions of shipment size by commodity and mode of transport. These were gleaned from survey data from Oregon and Michigan (for trucks) and the STB Carload Waybill Sample (for rail).

- *Destination choice:* A destination choice model allocates each shipment to a specific receiving firm or exporter based on its distance and economic sector. Input-output coefficients are used to define the relative attractiveness of firms, as are observed or asserted trip length frequency distributions.

- *Carrier and vehicle choice:* A carrier type (private or for-hire) is assigned to each firm based on its size and economic sector. Data from the Commodity Flow Survey (CFS) and Vehicle Inventory and Use Survey (VIUS) were used to choose between the carrier types. In addition, a vehicle type is selected based upon the chosen carrier type, data for which are available from the same sources. The vehicle choice is further conditioned by observed shares of trucks by types from traffic counts within the region and on its periphery.

- *Tour optimization:* Finally, the shipments are allocated to specific vehicles based upon their capacity and observed loading characteristics (collected from truck intercept surveys). The tours are optimized to minimize total travel time using the well-known traveling salesman problem (TSP) methodology. For firms requiring more than one truck the deliveries are first partitioned by general direction of flow before optimization is carried out. It is assumed that private trucks do not accumulate shipments from other shippers, and return home empty. For-hire trucks, on the other hand, can add shipments from nearby firms dynamically as the simulation progresses, and can wait for several shipments before initiating their tours.

Table 2: Summary of data required for model application

Source	Data requirement(s)
Commodity Flow Survey (CFS)	Value-to-ton ratios
	Mode shares by commodity
	Long distance trip lengths
Vehicle Inventory and Use Survey (VIUS)	Average weekly miles by commodity and truck class
	Distribution of carrier type by commodity
	Distribution of truck type by commodity
	Average stops per week
Truck intercept surveys	Average and total shipment weights by truck type
Employment by firm	Attribution of Firm agents
	Discrete destination choice
Make and use coefficients	Shipment generation
	Discrete destination choice
Truck counts	Attribution of Import and Export agents
	Model assessment and validation

Tours are further constrained by operator shift limits and typical dwell times at each location. All deliveries and return to the base at the end of the tour must be accomplished within the available times. Because the model knows the travel and dwell times associated with each shipment the beginning time associated with each event is appended to each trip within each tour. The resulting tour list can be used directly in a network simulation, or aggregated into trip matrices by time period for use with traditional traffic assignment models. The latter was used during development of the CT and MOSAIC models.

The agent behavior is represented within the model by sampling from observed statistical distributions, rule-based decision-making, and spatial allocations. Almost all of the distributions, such as number of stops per tour, average trip length, mapping of commodities to and from industries, and average shipment sizes, were gleaned from survey or published data. Some could be asserted based on theory or expert advice, and almost all can be adjusted in a straight-forward manner by time series data or advice from experts or a Delphi panel. In some cases a logit-based discrete choice model was estimated in place of sampling from observed distributions, but both performed equally well in application. However, the latter proved far easier to adjust for scenario or sensitivity testing, especially when the studied behavior fell outside of the range of the data available for model estimation.

The MOSAIC model was implemented using data from Oregon, and extensively tested and validated as part of its development. The key data required to implement the model are shown in Table 2. Because the model is stochastic the question arose about the number of runs needed in order to obtain stable average results. At the level of the individual firm there is of course wide variation between model runs, but at the major corridor and overall system level the variation in key performance measures, such as vehicle miles and hours of travel, was slight. The degree of acceptable variability was of course dictated by the scenario being tested, but 20 to 30 runs produced average outcomes with highly stable indicators. The outcome of many other (most?) simulation models are combined to produce the average statistics from them. By contrast, the MOSAIC run with key indicators closest to the mean values from all runs is chosen as the average outcome, and is used as the modeled result.

The requirement to run a model repeatedly is often perceived as a disadvantage, especially if it has run times that approach those of contemporary urban travel demand models. The ability to overcome this limitation with a very fast model was a key design goal of MOSAIC. In Portland each run takes about 8 minutes, and in more recent tests in Melbourne each run takes approximately 14 minutes. Over 60 percent of the execution time is devoted to destination choice, which can be distributed across several computers. By using a parallel version of MOSAICthat uses all of the processor cores in recent computers the run times have been reduced by approximately two-thirds, allowing 20 runs to be carried out in less time than most person travel models execute once.

A model is useful only if it can convincingly answer questions posed to it. How well a model replicates reality and corresponds to accepted thinking dictates how convincing it will be. Unfortunately, most travel demand models undergo only cursory testing and validation at best. Owing to its novel formulation and development by fusing disparate data the MOSAIC model was subjected to extensive testing and validation. Sensitivity testing of the key model parameters was also carried out, which provided interesting insights into the relative importance of the input data. Another important outcome of sensitivity testing was to demonstrate that the model performed as expected over a wide range of values, sometimes implausible, of each parameter or data stream. The model crashed when the average shipment sizes largely exceeded truck capacities, but otherwise it behaved as expected and revealed no pathological behavior. Tests of the effects of stochastic variability in outputs were also carried out, using indicators such as vehicle miles of travel, total tons shipped, and total tons handled by trans-shipment centers.

The model was compared to observed validation data, to include the number of stops per tour, average trip length by truck type, and comparison of observed versus modeled flows. None of these data were used to constrain or condition the model, making them unbiased and reliable indicators of model performance. The model replicated each of these patterns acceptably. However, the validation framework was more extensive than a simple comparison against known values. Several tests of model structure were carried out first, comparing empirical and structural elements of the model to concepts and behavior reported in the literature, mathematical proofs, and external review by domain experts. This was followed by several behavioral tests, which aside from comparing the model to the targets described above included sensitivity testing of outcomes. It is envisioned that this rigorous validation process will be continued as the model evolves and is tested in other locations.

It would be reckless to assert that the MOSAIC model is a mature framework posed to replace traditional trip-based urban truck models. It's successful application in Oregon represents a proof of concept, but considerable work remains to be done and much yet to be learned. Other tour-based models offer equally imaginative and innovative features. The point to be made here is that such richer and more behaviorally sensitive models are clearly within reach, and are poised to enter the mainstream of urban freight modeling. With them will come much greater opportunities for understanding and facilitating the safe and efficient movement of goods in urban areas.

References

Donnelly, R. (2007), A hybrid microsimulation model of freight flows, *in* E. Taniguchi & R. Thompson, eds, 'City Logistics V', Institute of City Logistics, Kyoto, pp. 235–246.

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

Taylor, S. & Button, K. (1999), Modeling urban frieght: what works, what doesn't work?, *in* E. Taniguchi & R. Thompson, eds, 'City Logistics I', Institute of City Logistics, Kyoto, pp. 203–217.