ESTIMATING FREIGHT FLOWS FOR METROPOLITAN AREA HIGHWAY NETWORKS USING SECONDARY DATA SOURCES

For Presentation

National Urban Freight Conference
Long Beach, CA

February 2006

Author 1 (Contact Author)
Genevieve Giuliano
School of Policy, Planning, and Development
University of Southern California
Los Angeles, CA 90089-0626
Email giuliano@usc.edu

Author 2
Peter Gordon
School of Policy, Planning, and Development
University of Southern California

Author 3
Qisheng Pan
School of Public Affairs
Texas Southern University

Author 4
Jiyoung Park
School of Policy, Planning, and Development
University of Southern California

Author 5
Lanlan Wang
School of Policy, Planning, and Development
University of Southern California

ABSTRACT

We present ongoing work on developing an automated integration system for freight flow analysis and planning. To overcome the limitations of current estimation methods for commodity flows, we use reliable secondary sources, including small-area employment data, and derive estimates in a plausible way by means of a computational workflow. When available, we extract the data automatically from online sources, so that the system maintains the estimations continuously updated. In this paper we provide an overview of our modeling approach and the major data sources used. We apply the model using data from the Los Angeles region, and compare our assignment results with available screenline counts. The results are encouraging. Our approach will eventually allow planners and policymakers to make more informed decisions by utilizing the most recent data from many sources and enhancing the ability to explore different scenarios.

Estimating Freight Flows for Metropolitan Area Highway Networks

Using Secondary Data Sources

INTRODUCTION

Economic restructuring and globalization have vastly increased the volume of commodity flows by all transport modes. In the US, intercity ton-miles have increased approximately with GNP, but truck and air transport have increased faster than other modes. In 2001, for example, of the total bill of $579.6 billion the US spent on freight transportation, trucks carried 80.6%, i.e. $467.3 billion[1]. Total US ton-miles of freight increased from 3,584 billion in 1990 to 4,357 billion in 2003. Over the same period truck ton-miles increased from 854 to 1,264 billion, and air ton-miles increased from 10.4 to 15.1 billion[2].

Increased freight flows have had significant impacts on metropolitan areas. Traffic at major freight generators (ports, airports, rail yards, warehouse/distribution nodes) has greatly increased, adding to congestion and impacting surrounding neighborhoods. Increased train traffic interrupts road traffic and often conflicts with demands for passenger commuter service. Increased truck traffic is associated with added highway congestion, more delay due to accidents, and accelerated highway deterioration. In addition, rapid changes in economic linkages are leading to ever changing flow patterns and the spatial restructuring of metropolitan areas (USDOT 2005, Gordon, et. al. 1998, Giuliano 1998).

As freight flows and their impacts increase, transportation planners, managers and operators have a greater interest in developing better methods for tracking and monitoring

[1] http://www.bts.gov/publications/national_transportation_statistics/2005/html/table_03_07.html
[2] http://www.bts.gov/publications/national_transportation_statistics/2005/html/table_01_46b.html

commodity flows, and for analyzing these flows as they impact transportation nodes and networks. Yet, current freight flow estimation and analysis methods have several problems, some related to data and some related to the estimation methods themselves. This paper presents a different approach for commodity flow estimation in metropolitan areas. We use a regional input/output model and combine it with available import/export commodity flow data from secondary sources to estimate detailed commodity flow matrices. Additional computations allocate flows to modes and ultimately assign flows to the transportation network. Our work also includes data integration and automation techniques to make possible continuously updated and detailed freight flow estimates (Ambite and Weathers, 2005; Ambite, Jinwala and Kapoor, 2006). The estimation methodology and data sources utilized are the focus of this paper.

The remainder of this paper is organized as follows. The second section describes current freight flow methods and their problems. The third section presents the conceptual framework for our model. We discuss the motivations for our "bottom-up" approach and describe the model as developed to date. We present our plans for constructing and testing a continuously updatable freight flow estimation model in the fourth section. The last section presents some results from applying the model, and the final section discusses conclusions.

CURRENT METHODS FOR FREIGHT FLOW ESTIMATION

There is an extensive literature on urban transportation network modeling, and the state-of-practice is well advanced (Wilson,1970; List and Turnquist,1994; Willumsen,1978; 1984). However, such models (except List and Turnquist), do not explicitly treat freight flows. A common state of practice method of modeling truck flows in metropolitan area analysis, for example, is to use rule-of-thumb fixed factors based on passenger vehicle flows and observed truck counts at a small number of locations on the highway network.[3] Rail freight flows are not usually modeled at the metropolitan level.[4] This simple approach may be adequate when trucks account for only a small percentage of urban traffic, and when regional planners and policymakers are relatively unconcerned about intra-metropolitan freight traffic.

The situation has now changed, particularly in large metro areas like Los Angeles, home of the largest container ports (trade in 2004 was $243.44 Billion[5]) as well as the second largest air freight airport in the US. Globalization, restructuring of goods supply chains, and changes in warehousing practices have resulted in large overall increases in freight traffic, and extremely large increases associated with ports and airports. In the Los Angeles region, for example, heavy-duty truck (HDDT) miles (i.e. those trucks with five or more axles) have more than doubled since 1982, a growth rate greater than population, employment or total vehicle miles traveled. It is estimated that terminal activity at the Port of Long Beach alone generates an average of more than 403,000 annual HDDT trips (Starcrest Consulting, 2004). Increased freight traffic has, in turn, led to a number of problems: highway congestion, traffic accidents, roadway deterioration, air pollution, noise, risk due to hazardous materials transport, etc. Consequently freight flows have become a major concern for planner and policy-makers, and a need has been recognized for better metropolitan freight planning and analysis tools.

There is also a growing interest in metropolitan freight flows among urban researchers. The growing importance of international trade leads to questions of costs and impacts of commodity flows on regions and local areas; relationships between supply chains, flows and firm

---

[3] We have identified a number of papers with detailed proposals for further research in this area and assume that these are now in progress. See, for example, Houlguin-Veras, et al. (2001), deJong, et al (2002), Peacock and Demetsky (2003), Xu and Hancock (2004).

[4] Commodity flows are typically modeled at the inter-regional or inter-state level.

[5] http://www.wisertrade.org, port level total trade value for Los Angeles Port and Long Beach Port.

location behavior; costs and benefits of international trade, and impacts of goods movement on urban form and land values.

There are a wide variety of approaches to freight modeling. They range from simple techniques (e.g. estimating regression models between freight and passenger flows; assuming that freight follows the same patterns as passenger flows), to adaptations of passenger models to freight transportation analysis. In very few cases, specific intra-regional freight models have been developed; these typically rely heavily on unique data sources and hence are not transferable or easily updated.

Data Problems

Current methods of freight flow estimation are limited by lack of data that is appropriate, accessible, and reliable. Ideally, one would like to have accurate data on commodity flows by industry sector, mode, origin and destination at a geographic scale sufficiently fine to identify flows on specific routes or at specific locations. Since a large part of flows within a region either originate or are destined to locations outside the region, the regional import/export component is critical. Such a comprehensive data source does not exist, leaving analysts with two choices: develop an estimation method based on available data, or collect the necessary data either directly from freight transporters or indirectly through third party data providers. Surveys of trucking companies, railroads, air transport firms, etc. are costly, and private firms are often unwilling to provide proprietary information. Moreover, freight flows vary over time, and hence would require repeated surveys. Third party providers are costly, and the user has little information on how data are compiled and generated, and hence on data validity.

Reliance on conventional secondary data sources has its own problems. Metropolitan level analysis requires fine geography; most existing data are at a regional scale or higher. For example, waybill sample data on rail commodity flows is collected by the US Surface Transportation Board, but the data are limited to metropolitan area totals. With respect to commodities, there are various classification systems, units (dollars, tons), varying levels of aggregation, more information on import/export flows, little information on intra-regional shipments; more data on port, air import/export, little data on truck, rail imports/exports.

There are also problems associated with how to account for empty trucks, warehouse/secondary processing activities, intermodal exchanges within any region and how to account for data collected at different times, and over different time intervals. While we cannot solve all these problems, we have developed ways to circumvent many of them.

Standard Approaches and Methodological Problems

The Quick Response Freight Manual released by the US Department of Transportation in 1996 provided simple techniques and transferable parameters for developing commercial vehicle trip tables (USDOT 1996). Truck trip generation rates were estimated from the number of jobs in the employment sectors associated with commodity shipments. The default rates provided by the manual were taken from a survey in Phoenix, Arizona. After calculating truck trips from employment data, it is straightforward to construct a truck trip table and assign the trips by following the conventional UTPS (Urban Transportation Planning System) four-step models. This method is easy to implement. However, it is a nagging problem that the default parameters like truck trip generation rates are not easily transferable between different regions.

In the Los Angeles area, the Southern California Association of Governments (SCAG) developed a Heavy Duty Truck (HDT) Model in 1999 to forecast HDT travel patterns, traffic volumes as well as Vehicle Miles Traveled (VMT) for the entire SCAG region for base year 1997. SCAG continued to incorporate their newly updated HDT model into their regional transportation model. The most recent one is 2000 base year model that was summarized in SCAG's *Year 2000 Model Validation and Summary Report* in 2003. The forecasts for HDT activities were based on truck HDT trip generation rates developed through surveys, regional economic data and commodity flow data, and the activity at special generators, such as airports, seaports and intermodal transfer facilities. After trip generation, the HDT trips were distributed using gravity models with the friction factors estimated from truck trip diary survey data. At the end, the HDT trips were assigned to a regional highway network, and VMTs were estimated for emissions analyses (SCAG 1999). As with the Quick Response Freight Manual, SCAG's HDT model also used employment data to estimate internal truck trip generation rates on the basis of shipper-receiver survey data. However, the survey sample size was small and the survey was conducted over a short period of time due to limited funds. SCAG staff report that they have been conducting truck surveys, to improve their truck trip generation model, external model, special truck generators, to review truck PCEs, and to validate and to improved the HDT model

The California Department of Transportation (Caltrans) has released three versions of its Intermodal Transportation Management System (ITMS) for statewide transportation planning since 1996. ITMS estimated freight movement by different modes based on data from the STB's annual Rail Waybill Sample file,, Reebie Associate's Transearch database, Dri/McGraw Hill's US economy forecast, and the Port Import Export Reporting Service (PIERS). The ITMS traffic analysis zones are based on zip code areas. Because the Transearch database provides commodity flow information for different modes, ITMS has been working on converting commodity flows into truck trips and assigning the trips to the state-wide transportation network.

The freight model part of ITMS is still under development. Problems with these approaches as well as the various data limitations motivate our approach to estimating detailed freight flows[6].

CONCEPTUAL FRAMEWORK FOR INTEGRATED MODEL

Hedges (1971) proposed that an adequate model should: (a) have solid behavioral foundations, (b) be multi-modal, (c) be able to analyze interactions between passenger and freight trips and (d) be able to take feedback from policy changes. We would add that an adequate model should be (e) detailed enough to capture small area impacts, (f) use widely available (non-proprietary) and frequently updated data.

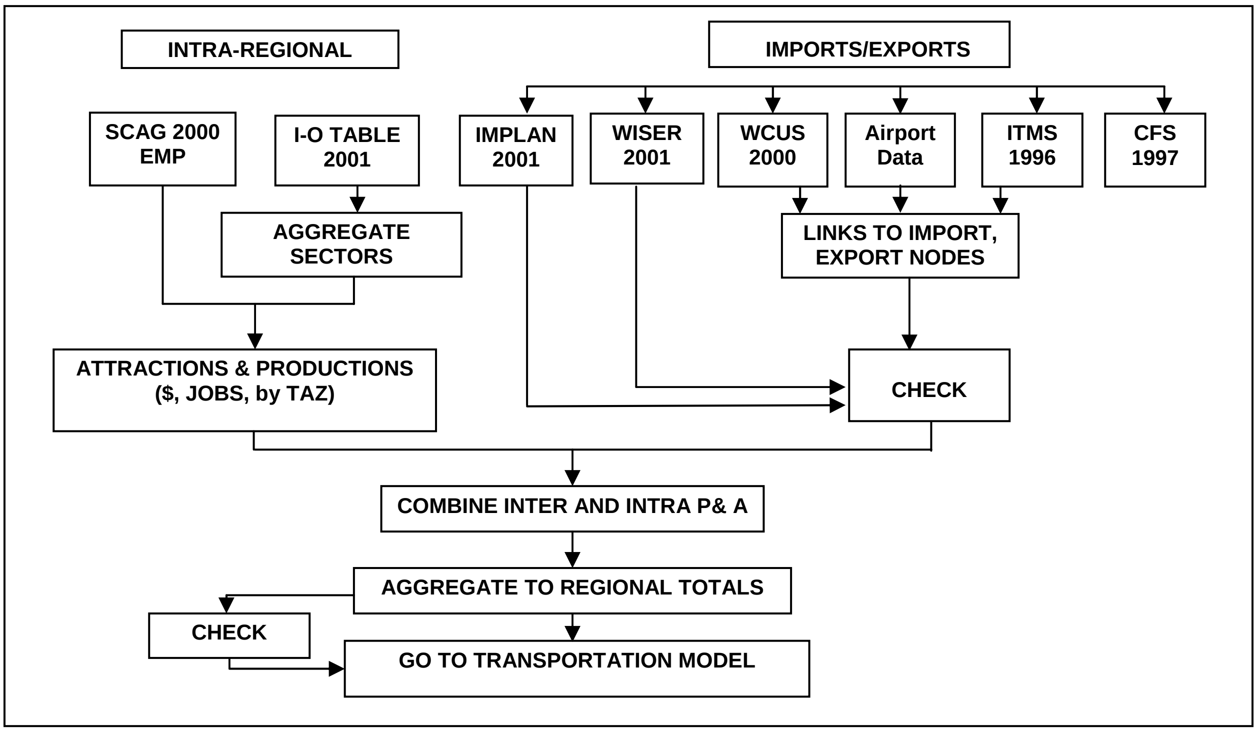
This work builds on a suggested approach to the problem by Gordon and Pan (Gordon and Pan 2001). The major research steps involved are the following:

1. Estimate commodity-specific interregional and international trip attractions and trip productions for those locations where airports, seaports, rail yards or regional highway entry-exit points are located.
2. Utilize a regional input-output transactions table to estimate *intraregional* commodity-specific trip attractions and trip productions, and allocate these to small-area units.
3. Create a regional commodity origin-destination matrix using estimates from (1) and (2).
4. Load the O-D matrix onto a regional highway network with known passenger flows.

[6] Booz•Allen & Hamilton Inc. (2001) *California Intermodal Transportation Management System* (ITMS), http://www.dot.ca.gov/hq/tpp/offices/oasp/itms/documentation.pdf

Our approach is illustrated in Figure 1. For interregional flows, we use a series of data sources to generate trip attractions and productions for the major import/export nodes. A regional input/output model and small area employment data are the basis for generating intraregional trip attractions and productions. Commodity attractions and productions are combined into an origin-destination matrix. These flows are converted from dollars to tons to passenger car equivalents (PCEs), and then assigned to the highway network.

FIGURE 1 Overview of Data Process Steps



In what follows, we describe a prototypical application of our approach to the Los Angeles metropolitan area (the five-county CMSA). Our approach has some important advantages. First, secondary data sources are widely available and regularly updated, hence the approach is easily transferable across metropolitan areas. Second, our approach avoids use of proprietary data (with two exceptions) and data obtainable only through metropolitan level surveys. Hence data costs are low, relative to other more conventional approaches.

However, using secondary data has some significant disadvantages. Because these sources were developed for other uses, no single source is sufficient for our purposes. Using multiple sources requires consistency across sources. We have found that freight data from the most important secondary data sources are described via various (often independent) classificatory systems and definitions. Much of our work has been devoted to reconciling data from these various sources. We have created a set of bridge tables for different economic classifications, which enable us to translate across classification systems.

MODEL DATA SOURCES AND RECONCILIATION

In order to estimate total freight shipments on the metropolitan highway network, we must account for shipments (in tonnage) to and from all sources, both within the region and to and from the region. We use IMPLAN data for commodity information and to inform the distribution of freight supply and demand within the region. We also require information on shipments by mode in order to generate a truck shipment O-D matrix. These requirements determine our selection of data sources. In this section we describe the main data sources used in each; following sections describe our estimation process.

Inter-regional Commodity Flows

The first research step requires data on inter-regional commodity flows through airports, seaports, rail yards or regional highway entry-exit points. We use several data sources in order to estimate the required flows.

Commodity Flow Survey

The Commodity Flow Survey (CFS) provides commodity flows by sector for US regions, states and Metropolitan Statistical Areas (MSA)[7]. The CFS uses the Standard Classification of Transported Goods (SCTG) code. Flows are provided in both dollar value and tonnage, and are given by mode. Level of detail varies by geographic unit. The 1997 CFS includes a total sample of over 5 million shipments. For each sampled 1997 CFS shipment, zip code of origin and destination, 5-digit Standard Classification of Transported Goods (SCTG) code, weight, value, and modes of transport, are given[8].

IMPLAN

IMPLAN is a proprietary data source, provided by the Minnesota IMPLAN Group, Inc since 1993.[9] The 2001 IMPLAN provides county level input/output data for 509 industrial sectors. IMPLAN also includes county-level inbound/outbound flows, as well as state and national-level foreign imports and exports information.

WISER and USITC

The World Institute for Strategic & Economic Research (WISER) and the United States International Trade Commission (USITC - http://dataweb.usitc.gov) provide monthly exports and imports by customs district by detailed Harmonized System (HS) commodity codes or annual exports and imports by port by Standard International Trade Classification (SITC).[10]

---

[7] Bureau of Transportation Statistics: http://www.bts.gov/programs/commodity_flow_survey

[8] http://www.bts.gov/programs/commodity_flow_survey/detailed_description/index.html. Survey data used 5-digit SCTG codes. CFS final report used 2-digit or aggregated 2-digit SCTG codes.

[9] http://www.implan.com/what.html

[10] See WISERTrade: http://www.wisertrade.org/home/index.jsp

WISER is a proprietary data source. We use it for modal data that are not available from the CFS. Data are provided in dollar and tonnage units.

<u>WCUS</u>

The Waterborne Commerce Statistical Center (WCSC) provides annual commerce data including both domestic and foreign trade for major U.S. ports, called Waterborne Commerce of United State (WCUS) data.[11] The codes of WCUS data conform to the hierarchical structure of the SITC Revision 3 commodity codes and the HS coding system, which allows comparisons with commodity movements in the U.S. and other countries. Data are provided in tonnage units.

<u>Other Sources</u>

The data sources described above do not have all the information needed for our estimates. We use data from the metro area's major airports and RAND Corporation for air cargo tonnage.[12] We also use the California Department of Transportation's ITMS (Intermodal Transportation Management System) to obtain data on truck and rail tonnage by import/export node. These sources do not provide sector-level information.

Intra-regional Commodity Flows

Data sources for intraregional commodity flows are straightforward: a regional input-output (I-O) transactions table and small area employment data. The I-O table is developed from the IMPLAN economic impact modeling system and provides up-to-date inter-industrial

---

[11] WCUS: http://www.iwr.usace.army.mil/ndc/data/dictionary/ddwcus.htm
[12] RAND Corporation: http://ca.rand.org/stats/economics/airport.html

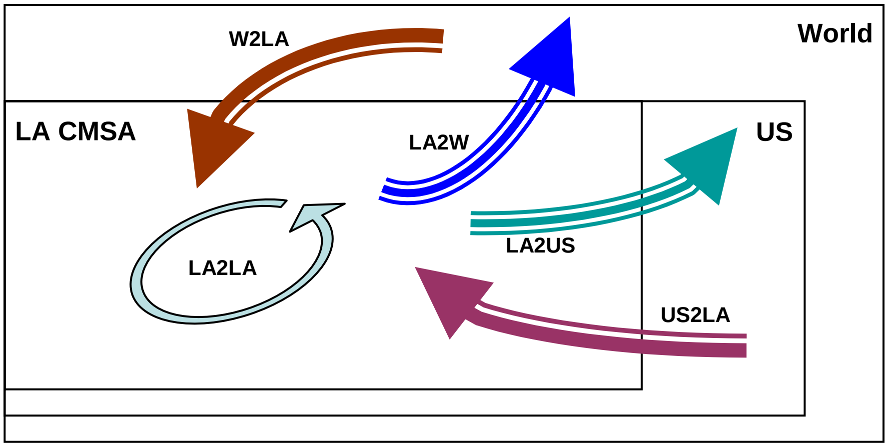
transactions by industry sector.[13] Inter-industrial transactions are described in dollars. Small area employment data are derived from state Employment Development Department (EDD) records and provided by the local metropolitan planning organization (MPO). Yearly jobs are classified in Standard Industrial Classification code. Small area employment data is the basis for converting inter-industrial transactions to inter-zonal commodity flows.

ESTIMATIONS

Interregional commodity flows

We divide the world into three regions, the metropolitan area (in this case the greater Los Angeles five-county area, hereafter referred to as LA), the rest of the US (hereafter referred to as US) and the rest of the world (hereafter referred to as W). This results in five freight flows as illustrated in Figure 2. The five flows are required in order to utilize data from the various sources noted above and maintain sum totals consistent with the IMPLAN data. Note that "LA2LA" corresponds to the intraregional flow to be discussed in the next section.

Figure 2: Conceptual Freight Flows



Legend: LA2LA = LA to LA; LA2US = LA to US;
LA2W = LA to rest of world; US2LA = US to LA; W2LA = rest of world to LA

[13] Minnesota IMPLAN Group;
http://www.implan.com/index.php?page=index&Base_Session=b3f81c27d3c3a8efe3fbf9c347fd6bfe

We use the IMPLAN data to derive each of these flows by industry sector. IMPLAN does not provide all 5 flows directly.[14] For this research, we aggregate the IMPLAN 2001 information to 9 commodity sectors that match those of the Commodity Flow Survey (CFS 1997; SCTG sectors) data available for the LA area. Once the flows have been estimated for each of the nine commodity sectors, we use the CFS data to allocate flow proportions to four modes: air, rail, truck, water. We create five 9x4 matrices of proportions corresponding to the five trade flows of Figure 2. Because the CFS data is categorized as either "inbound" or "outbound" (e.g. we do not know where "inbound" is coming from or "outbound" is going to), and because the LA2LA flow is embedded in both inbound and outbound data, we cannot directly map our five flows to the CFS data. We therefore use a series of calculations to generate the mode proportions for each sector/flow. Modal information not available in CFS was drawn from WISERTrade.

Intraregional flows

For the estimation of intraregional freight trip ends, two expressions are used to estimate attractions and productions of commodities at each traffic analysis zone (TAZ). The approach requires a regional input-output transactions table and small area employment data (introduced in Cho et al., 1999). The regional input-output model, with foreign shipments removed, provides the basis for estimating zone-to-zone shipments, once the regional coefficients are combined with small-area jobs-by sector data. In this case, we use traffic analysis zones (TAZs) which are approximately the size of census tracts.

Specifically, equation (1) calculates the total commodity *i* required to support production in zone *z*:

[14] Details on these calculations and others below are available upon request to the authors.

$$D_i^z = \sum_j a_{ij} X_j^z \quad \textbf{(1)}$$

where, X_j^z is the total regional output of commodity *j* in zone *z*, given base year employment in sector *j* and zone *z*,

a_{ij} is the *i*, *j*th element of **A**, the matrix of value demand coefficients for the (open) input-output model; this represents the flow from *i* to *j* per unit output of *j*,

D_i^z is the total flow attracted from sector *i* in response to the demand in zone *z*, expressed in dollars.

D_i^z represents the shipments of commodity *i* to zone *z* from all other zones to accommodate local demand, excluding household final demand and non-local final demand (imports).

Similarly, Equation (2) calculates the total supply of output *j* furnished by zone *z*,

$$O_j^z = \sum_i b_{ij} X_i^z \quad \textbf{(2)}$$

where, X_i^z is the total regional output of commodity *i* in zone *z*, given base year employment in sector *i* and zone *z*,

b_{ij} is the *i*, *j*th element of **B**, the matrix of value supply coefficients for the (open) input-output model. This is the flow from *i* to *j* per unit output of *i*.

O_j^z is the total flow produced from zone *z* to satisfy the demands by sector *j*, in dollars.

O_j^z represents the shipments of commodity *j* to all other zones to accommodate local demand,, again excluding household and non-local final demand (exports).

The total output of commodity *j* in zone *z* is calculated as zone *z*'s proportion of all employment associated with commodity *j*:

$$X_j^z = \frac{E_j^z}{\sum_z E_j^z} \left(\sum_i x_{ij} + \alpha H_j + R_j \right) \qquad \textbf{(3)}$$

where E_j^z is the employment by sector *j* at zone *z*,

H_j is the regional household consumption of commodity *j*,

R_j is the final demand not associated with households,

α is an estimated parameter between 0 to 1 representing the portion of household consumption (e.g final demand) contributing to freight movements. It is usually set to zero or a very small number.

Employment data by sector by zone are from a data file provided by the Southern California Association of Governments (SCAG, 2000). The input-output transactions table provides the dollar values of inter-sector commodity flows that serve household consumption and the parts of local final demand not associated with households. We can carry out these calculations for any number of commodity sectors. In this application, we aggregate to a smaller number of SCTG sectors because of data constraints imposed by the estimation procedures for international and interregional trip ends as described above. These calculations are the basis for the dollar values of TAZ-level intraregional O-D matrices of commodity flows. We assume that the mode for these flows is truck.

Generating the Origin-Destination Matrix and Traffic Assignment

The O-D matrix combines the intraregional and interregional flows. We first distribute intraregional trips by converting the tonnage flows to heavy duty truck (HDT) equivalent units and using a conventional gravity model formulation. The distance decay coefficient values for freight trips are calibrated to minimize the difference between the "observed" and "estimated" freight trip productions (Cho et. al. 1999, Gordon and Pan 2001). Coefficients are calculated for nine freight sectors. These are applied to the HTD trips associated with commodity flows for the given sector.

We then distribute inter-regional commodity flows to a limited number of zones or entry/exit points: the two major seaports (Long Beach and Los Angeles); the five major airports involved in freight shipping; three major rail yards and twelve highway entry-exit points. A variety of data sources were used to distribute the inter-regional flows to these entry/exit points. The five airports' trade information was either provided by airports themselves or by RAND; seaports' trade data are drawn from WCUS 2001; and the inter-regional truck and rail flow information is taken from the ITMS 1996. The inter-regional freight flows were then distributed to these points according to their regional shares.

To generate the O-D matrix of these inter-regional flows, we used an algorithm developed for this project. Once the intra-regional freight trips are distributed, inter-regional freight trips can be distributed based on the attracted trips at internal TAZs. The following formula can be used to distribute the internal-external freight trips:

$$F^{Eo,d} = \sum_i Inb_i^{Eo} \frac{A_i^d}{\sum_d A_i^d} \qquad \textbf{(4)}$$

$$F^{o,Ed} = \sum_i Outb_i^{Ed} \frac{P_i^o}{\sum_o P_i^o} \qquad \textbf{(5)}$$

where, $F^{Eo,d}$ are freight trips from regional entry-exit point Eo to internal TAZ *d*,

$F^{o,Ed}$ are freight trip from an internal TAZ *o* to a regional entry-exit point Ed ,

Inb_i^{Eo} are inbound commodity *i* at regional entry-exit point Eo ,

$Outb_i^{Ed}$ are outbound commodity *i* at regional entry-exit point Ed ,

A_i^d and P_i^o are the attraction of commodity *i* at zone *d* and the production of commodity *i* at zone *o*, respectively.

The result is a 3203 by 3203 (3191 SCAG TAZs plus 12 external zones as highway entry/exit points) HDT O-D trip table. Once the origin-destination matrix of freight flows was estimated, a traffic assignment model was used to assign these flows to region's highway network. Freight flows (in passenger-car equivalents) were added to SCAG's estimated passenger flows on all links. Traffic assignment models the trip-maker's choice of path between all available zonal pairs. Equilibrium-based travel demand models are usually adopted for the purpose of traffic assignment. For a congested network condition, strict network assignment models are appropriate to predict the equilibrium flows. Based on the theory of User-Optimal-Strict (UO-S) On Network Assignment (NA), Sheffi (1985) provided a traffic assignment model that assumes perfect rationality among travelers, no temporal fluctuations and no modal or link interactions. Sheffi's method was implemented to assign the passenger and truck trip volumes to the highway network of the Los Angeles five-county region.

The optimization function used was:

$$MIN_{f_a} \sum_a \int_0^{f_a} C_a(x)dx \qquad \textbf{(6)}$$

subject to $f_a = \sum_{r \in R} \delta_{ar} h_r \quad \forall a \in A$

$\sum_{r \in R} h_r = D_{ij} \quad \forall i \in I, j \in J$

$h_r \geq 0 \quad \forall r \in R$

where f_a = total flow on arc *a*

C_a = average travel cost on arc *a*

δ_{ar} = arc-path incidence variable; equal to one if arc *a* belongs to path *r*

h_r = flow on path *r*

r = network path, and *R* is the set of all paths available between place *i* and *j*.

Applying this algorithm to the minimization of the UO-S model requires a solution of all feasible values to be generated at each iteration step. At convergence, total travel time on the network is minimized, assigning all trips to the shortest travel-time path of the origin destination pairs.

In summary, we developed a methodology for estimating the detailed commodity flows based on secondary sources. We also used plausibly reliable secondary sources, as available, and derived the estimates in a more reasonable way than the competing approaches discussed previously.

EMPIRICAL RESULTS

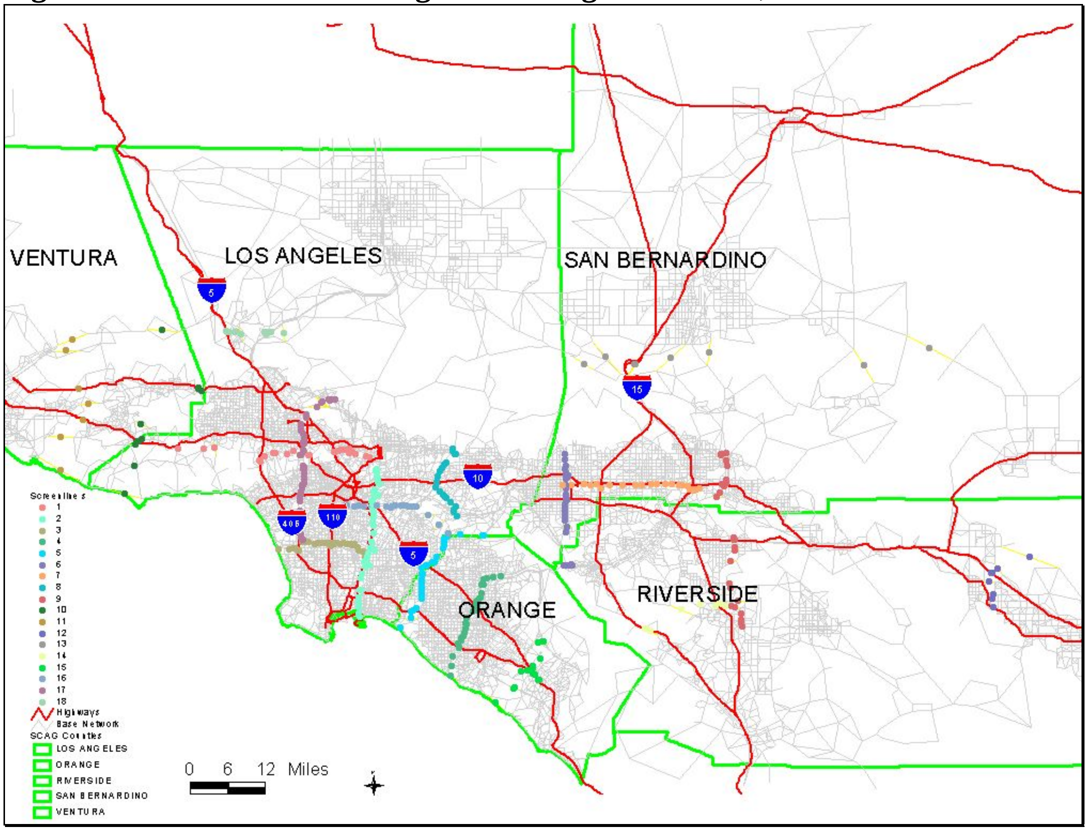
The modeling process described above makes it possible to estimate freight commodity flows on the regional highway network. Rail and air network flows are straightforward; the more challenging problem is to estimate truck flows on the highway network. We have run the model and estimated truck flows for the five-county region, and we now briefly discuss our results.

The assignment of HDTs to the highway network takes as given the equilibrium assignment of passenger car trips (e.g. there is already congestion, and the passenger car assignment does not change). The traffic assignment submodel performs a 3-hour AM peak assignment for passenger car equivalents (PCEs), hence the HDT O-D trip table must be converted to PCEs and factored down to the peak hour. The HDT traffic assignment was performed using the Sheffi (1985) UO-S approach described earlier.

In order to assess the plausibility of the traffic assignment results, it is desirable to compare estimated HDT volumes with actual volumes. We were able to use actual count data collected by CalTrans and SCAG as part of the 2003 Heavy Duty Truck Model study[15]. Eighteen regional screenlines were established for the study, and 24-hour average daily traffic counts were conducted for all freeways, state highways and major arterials crossing each screenline. Figure 3 gives the location of the 18 screenlines.

[15] SCAG defines HDT as vehicles with gross vehicle weight greater than 8500 pounds.

Figure 3 SCAG/LAMTA Freight Modeling Screenlines, 2003



Source: SCAG/LAMTA, 2004, Regional Screenline Traffic Count Program Final Report, Prepared by Meyer, Mohaddes Associates, Inc.

We compare our HDT estimates with the SCAG 2003 screenline counts by factoring up our network assignment model results to 24 hours and using two different PCE factors. The first is a fixed factor based on truck volume counts conducted by the California Department of Transportation. It is obtained as a weighted proportion of trucks by number of axles in the Los Angeles metropolitan area. The second is a set of PCEs unique to each screenline and generated by SCAG. In each case we compare our model PCE estimates with the "actual", the PCEs from the daily 24 hour screenline truck counts, and the SCAG model estimates with the "actual" in PCE. The results are summarized in Table 1.

Tables 2 and 3 give the screenline results for each comparison, and Figures 4 and 5 show the regression results. The weighted average differences are 20 percent and 17 percent respectively. The range of differences is slightly greater for comparison 1 than 2; using the SCAG screenline specific PCEs (all greater than our fixed factor) generates mostly negative differences.

Table 4 and Figure 6 provide a comparison of SCAG 2000 model estimates with the "actual" ground counts, both in PCEs. From the details of comparisons in Table 2 to 4 and the summary table 1, it is clear that our model results are better than the SCAG 2000 model estimates in multiple comparison categories, including average percent difference, minimum difference, average weighted percent difference, and weighted percent mean square difference. In maximum difference and R-square, our results in comparison 1 are slightly worse than SCAG estimate but our comparison 2 are slightly better than the comparison of SCAG modeling results.

Table 1: Screenline comparison results

	Comparison 1 (author fixed PCE)	Comparison 2 (SCAG screenline PCE)	Comparison 3 (SCAG Modeling Results in PCE)
Average % difference	36.5	-5.3	76.47
Min % difference	0.8	7.5	-29.67
Max % difference	206.8	134.0	196.84
Average weighted % difference	20.0	17.0	69.58
Weighted % mean sq. error	17.8	10.0	25.79
Regression R^2	0.80	0.75	0.76

We are encouraged by these results. Our model estimates explain a large proportion of the variation in screenline PCEs, and our screenline estimates in most cases are reasonably close to the actual truck counts. We have made no effort to calibrate or adjust our results, and we have relied extensively on secondary data sources, many of which were not intended for this application. We have incorporated many simplifying assumptions in order to bring together disparate data.

DISCUSSION

We have presented a method for using widely available data sources to estimate link-specific freight flows on trucks for the highway network of a major metropolitan area, in this case the Los Angeles five-county area. This ambitious undertaking is made in light of a variety of heroic assumptions about the veracity of the data used, as well as the accuracy of our data reconciliation efforts. Our approach facilitates analysis and comparison of different scenarios, for example the impacts of expanded international trade, of increased highway or facilities congestion, the contribution of trucking to highway congestion, the relationship between employment location and commodity flows, etc. Although our research focuses initially on the Los Angeles metropolitan area, our approach was developed with the intent of transferability across metropolitan areas. The computational framework will be the same, and most of our data sources have a national scope. We would only need to add selected sources specific to a particular area. Our main data sources are regularly updated, and our automated data integration and computation will make updating and scenario generation easy and efficient. In summary, we hope that our work provides a tool that will allow regional planners and policymakers to make better and more informed decisions.

Acknowledgement

This research was supported by the National Science Foundation under grant EIA 0138998, by the METRANS Transportation Center through grants supported by the USDOT and the California State Department of Transportation. All errors and omissions are the responsibility of the authors.

Table 2 Comparison 1: 2000/2001 Model Estimates and 2003 Actual HDT, 24 hr, constant PCE Truck Ratio

Screenline	2003 AAWTT*1	PCE(ADT)*2	SCPM 2005	Difference	%Difference
1	58,389	131,612	174,448	42,836	32.55%
2	131,453	296,302	326,348	30,047	10.14%
3	67,137	151,330	184,397	33,067	21.85%
4	74,204	167,260	176,556	9,297	5.56%
5	88,761	200,072	214,266	14,195	7.09%
6	96,592	217,723	299,405	81,682	37.52%
7	68,288	153,925	78,258	-75,666	-49.16%
8	84,845	191,245	192,772	1,528	0.80%
9	31,038	69,961	92,330	22,369	31.97%
10	18,515	41,734	42,517	783	1.88%
11	9,072	20,449	32,860	12,411	60.70%
12	19,388	43,702	63,809	20,107	46.01%
13	18,495	41,689	127,898	86,209	206.79%
14	12,696	28,617	60,449	31,832	111.23%
15	26,763	60,325	34,873	-25,452	-42.19%
16	69,698	157,103	204,217	47,114	29.99%
17	100,378	226,257	247,084	20,827	9.20%
18	32,782	73,892	174,460	100,568	136.10%

Note: 1. 2003 Annual Average Weekday Truck Traffic Counts (AAWTT) was estimated by SCAG/LAMTA (2004)
2. HDT ADT (Average Daily Traffic) was converted into PCE/day using the ratio 2.254 calculated by the Author
3. The average % difference weighted by 2003 Actual HDT (PCE) is 19.96% and the weighted mean square error for this weighted average difference is 17.79%.

Figure 4 Comparison 1 Regression Plot

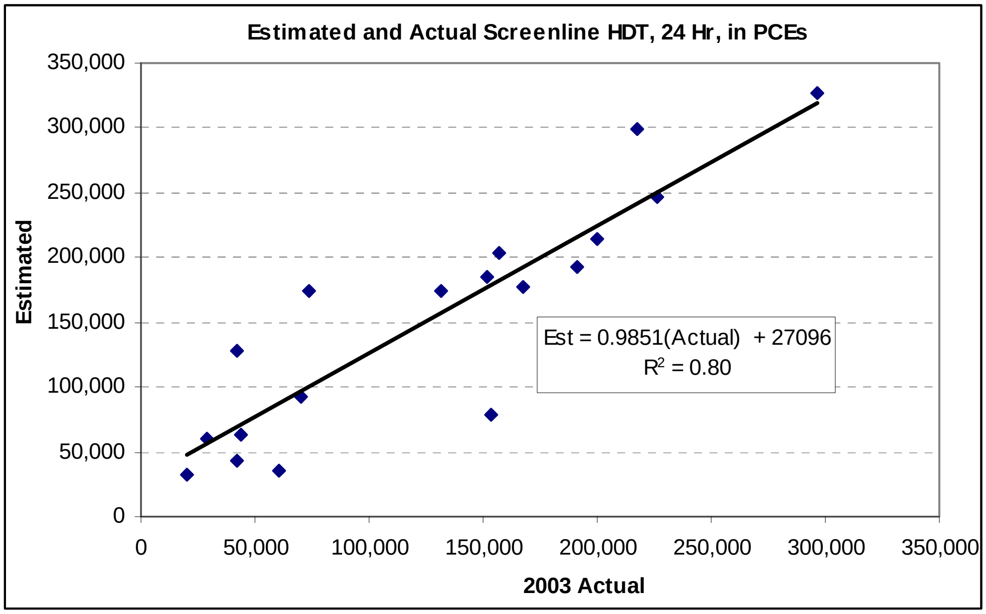


Table 3 Comparison 2: 2000/2001 Model Estimates and 2003 Actual HDT, 24 hr, SCAG screenline PCEs

Screenline	2003 AAWTT*1	PCE*2	PCE(ADT)*3	SCPM2005	Difference	%Difference
1	58,389	3.93	229,663	174,448	-55,215	-24.04%
2	131,453	2.87	376,832	326,348	-50,484	-13.40%
3	67,137	3.22	215,957	184,397	-31,560	-14.61%
4	74,204	3.12	231,269	176,556	-54,713	-23.66%
5	88,761	3.22	285,515	214,266	-71,248	-24.95%
6	96,592	2.95	284,946	299,405	14,458	5.07%
7	68,288	2.87	195,759	78,258	-117,501	-60.02%
8	84,845	3.22	272,918	192,772	-80,146	-29.37%
9	31,038	3.22	99,839	92,330	-7,509	-7.52%
10	18,515	3.93	72,826	42,517	-30,309	-41.62%
11	9,072	3.93	35,683	32,860	-2,823	-7.91%
12	19,388	2.95	57,195	63,809	6,614	11.56%
13	18,495	2.95	54,560	127,898	73,338	134.42%
14	12,696	3.22	40,839	60,449	19,610	48.02%
15	26,763	3.93	105,268	34,873	-70,395	-66.87%
16	69,698	3.22	224,195	204,217	-19,979	-8.91%
17	100,378	3.93	394,820	247,084	-147,736	-37.42%
18	32,782	3.22	105,449	174,460	69,012	65.45%

Note: 1. 2003 Annual Average Weekday Truck Traffic Counts (AAWTT) was estimated by SCAG/LAMTA (2004)
2. The PCE ratios were estimated basing on Table 18 of the SCAG Heavy Duty Truck Model and VMT Estimation Final Report (SCAG 1999)
3. HDT ADT (Average Daily Traffic) was converted into PCE/day using the estimated PCE value
4. The average % difference weighted by 2003 Actual HDT (PCE) is -16.95% and the weighted mean square error for this weighted average difference is 9.97%.

Figure 5 Comparison 2 Regression Plot

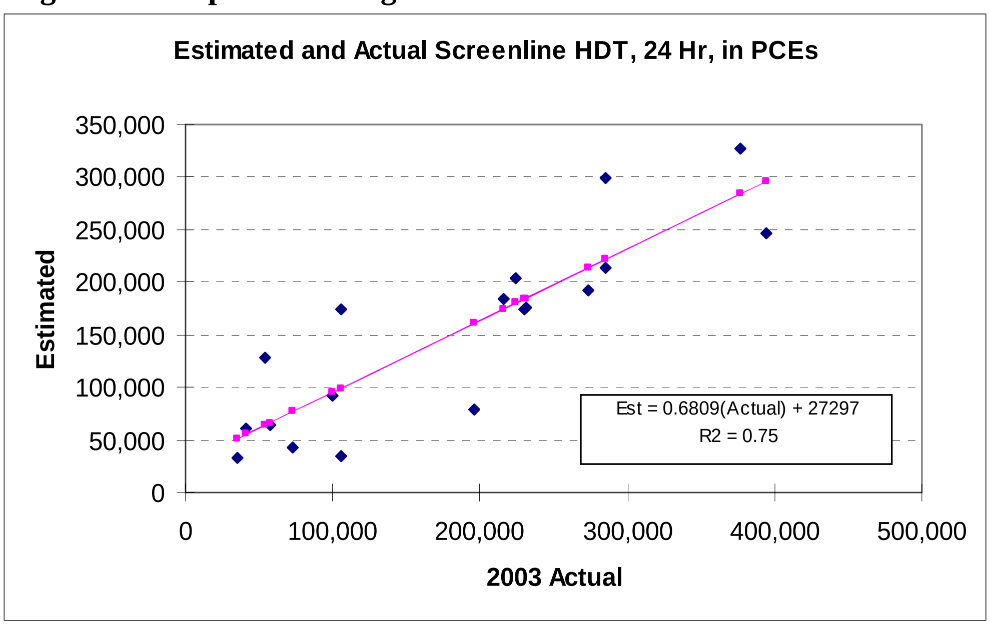
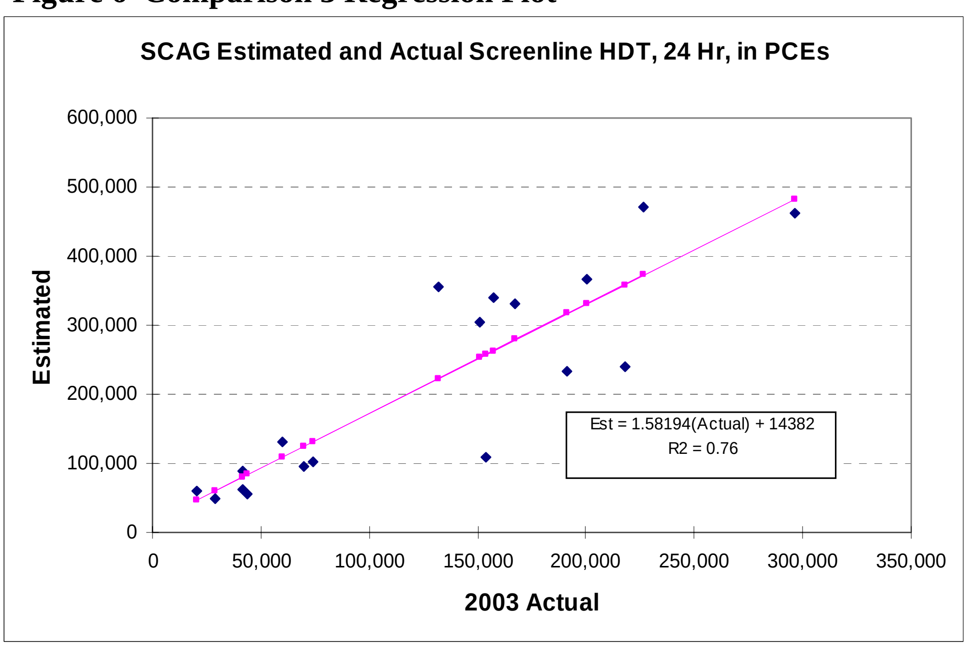


Table 4 Comparison 3: 2000 SCAG Model Estimates and 2003 Actual HDT, 24 hr, SCAG screenline PCEs

Screenline	2003 AAWTT*1	PCE*2	Actual PCE(ADT)*3	2000 SCAG Model Estimate*4	SCAG Estimate in PCE*5	Difference	%Difference
1	58,389	3.93	229,663	90,276	355,086	223,474	169.80%
2	131,453	2.87	376,832	161,422	462,743	166,441	56.17%
3	67,137	3.22	215,957	94,598	304,290	152,960	101.08%
4	74,204	3.12	231,269	105,942	330,186	162,926	97.41%
5	88,761	3.22	285,515	114,196	367,330	167,259	83.60%
6	96,592	2.95	284,946	81,540	240,543	22,820	10.48%
7	68,288	2.87	195,759	37,764	108,257	-45,668	-29.67%
8	84,845	3.22	272,918	72,724	233,929	42,684	22.32%
9	31,038	3.22	99,839	29,822	95,927	25,966	37.12%
10	18,515	3.93	72,826	22,338	87,863	46,129	110.53%
11	9,072	3.93	35,683	15,432	60,699	40,250	196.84%
12	19,388	2.95	57,195	18,891	55,728	12,027	27.52%
13	18,495	2.95	54,560	20,716	61,112	19,424	46.59%
14	12,696	3.22	40,839	15,031	48,350	19,732	68.95%
15	26,763	3.93	105,268	33,059	130,032	69,707	115.55%
16	69,698	3.22	224,195	105,832	340,426	183,323	116.69%
17	100,378	3.93	394,820	119,754	471,032	244,775	108.18%
18	32,782	3.22	105,449	31,529	101,418	27,526	37.25%

Note: 1. 2003 Annual Average Weekday Truck Traffic Counts (AAWTT) was estimated by SCAG/LAMTA (2004)
2. The PCE ratios were estimated basing on Table 18 of the SCAG Heavy Duty Truck Model and VMT Estimation Final Report (SCAG 1999)
3. HDT ADT (Average Daily Traffic) was converted into PCE/day using the estimated PCE value
4. SCAG 2000 model estimates
5. SCAG HDT ADT (Average Daily Traffic) was converted into PCE/day using the estimated PCE value.
6. The average % difference weighted by 2003 Actual HDT (PCE) is 69.58% and the weighted mean square error for this weighted average difference is 25.79%.

Figure 6 Comparison 3 Regression Plot



REFERENCES

Ambite, J. L., M. Weathers (2005), Automatic Composition of Aggregation Workflows for Transportation Modeling, dg.o, pp. 41-49.

Ambite, J. L., D. Kapoor, M. Jinwala (2006) Data Processing Workflows in the Social Sciences: Representation and Automatic Generation, dg.o.

Cho, Sungbin et. al. (1999) *Integrating Transportation Network and Regional Economic Models to Estimate the Costs of a Large Earthquake: NSF Draft Report*. Los Angeles, California.

De Jong, G., H. Gunn, W. E. Walker, J. Widell (2002) Study on Ideas on a New National Freight Model System for Sweden, http://www.rand.org/pubs/monograph_reports/2005/MR1663.pdf.

Giuliano, G. (1998) Information Technology, Work Patterns and Intra-metropolitan Location: A Case Study, *Urban Studies*, 35, , pp. 1077–1095.

Gordon, P.; Y-CH Liao, and H. W. Richardson (1998) Household commuting: implications of the behavior of two-worker households for land-use/transportation models, in *Network infrastructure and the urban environment: advances in spatial systems modeling*, edited by Lars Lundqvist, Lars-Göran Mattsson, Tschangho John Kim, New York: Springer.

Gordon, P. and Q-Sh. Pan (2001) *Assembling and Processing Freight Shipment Data: Developing a GIS-Based Origin-Destination Matrix for Southern California Freight Flows*, National Center for Metropolitan Transportation Research (www.metrans.org).

Graham, S. and S. Marvin (1996) *Telecommunications and the city: electronic spaces, urban places*, New York : Routledge.

Hedges, C. A. (1971) Demand forecasting and Development of a Framework for Evaluation of Urban Commodity flow: Statement of the Problem, in *Special Report 120: Urban Commodity Flow*, pp. 145-148, Highway Research Board, Washington DC.

Holguin-Veras, J., G. F. List, A. H. Meyburg, K. Ozbay, H. Teng, S. Yahalom (2001) An Assessment Of Methodological Alternatives for a Regional Freight Model in the NYMTC Region, http://www.utrc2.org/research/assets/6/regionalfreight1.html.

List, G. F. and M. A. Turnquist (1994) *Estimating truck travel patterns in Urban areas,* Transportation Research Record 1430, pp.1-9.

Peacock, K. L., M. J. Demetsky (2003) Research Needs For Developing A Commodity-Driven Modeling Approach, http://www.virginiadot.org/vtrc/main/online_reports/pdf/04-cr13.pdf.

Pendyala, R. (2002) *Urban Freight Modeling Including Intermodal Connectors for Florida*, Tallahassee, FL: Florida Department of Transportation.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, New Jersey: Prentice Hall.

Southern California Association of Governments (SCAG) (1999) *SCAG Heavy Duty Truck Model and SCAG VMT Estimates*, Los Angeles, California.

Starcrest Consulting (2004) *2002 Baseline Emissions Inventory,* http://www.polb.com/civica/filebank/blobdload.asp?BlobID=2159

US Census Bureau. (1997) *Commodity Flow Survey,* Washington DC: US Department of Transportation and US Department of Commerce.

US Department of Transportation (USDOT) (1996) *Quick Response Freight Manual –Final Report*, Washington, D.C.

US Department of Transportation (USDOT) (2005) *National Transportation Statistics 2005*, http://www.bts.gov/publications/national_transportation_statistics/2005/index.html.

Willumsen, L. G. (1978), *OD Matrices from network data: a comparison of alternative methods for their estimation*, Proceedings of the PTRC Summer annual meeting: 1978 Seminar in transportation models, PTRC educational research services ltd. London.

Willumsen L. G. (1984) *Estimating time dependent trip matrices from traffic counts*, Ninth International symposium on transportation and traffic theory, VNU Science press, Utretch, The Netherlands.

Wilson, A. G. (1970) *Entropy in Urban and Regional Modeling*, London: Pion.

Xu, J. H., and K. L. Hancock (2004) Enterprise-Wide Freight Simulation in an Integrated Logistics and Transportation System, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 4.