GENERATING COUNTY-LEVEL FREIGHT DATA USING FREIGHT ANALYSIS FRAMEWORK (FAF2.2) FOR REGIONAL TRUCK EMISSIONS ESTIMATION

BY

Minyan Ruan

PhD candidate, Department of Civil and Materials Engineering,

University of Illinois at Chicago, 842 W. Taylor Street (M/C 246),

Chicago, IL, 60607, U. S. A.

Email address: mruan2@uic.edu (M. Ruan)

Jie (Jane) Lin[1]

Assistant Professor, Corresponding author.

Department of Civil and Materials Engineering, University of Illinois at Chicago,

842 W. Taylor Street (M/C 246), Chicago, IL, 60607, U. S. A.,

Tel: 1-312-996-3068, fax: 1-312-996-2426,

Email addresses: janelin@uic.edu (J. Lin)

Number of Words: 5,791

Number of Tables: 5

Number of Figures: 6

[1] Corresponding Author

ABSTRACT

Current publicly available national commodity flow database (e.g., Freight Analysis Framework, Commodity Flow Survey, etc.) is restricted to the state or metropolitan region as a whole, hence is of little use to regional freight transportation planning, which requires detailed data at a sub-regional level (e.g., county, traffic analysis zone (TAZ)). High-resolution proprietary databases (e.g., TRANSEARCH database), typically at the county level, provide limited information on their methodologies and validations. Some local agencies have conducted regional surveys to obtain freight data in areas as small as the TAZ or ZIP code level, but high cost and uncertainties associated with them are hindrance to the availability and usefulness of the data. Therefore, the objective of this paper is to explore different synthesis methodologies to generate high geographic resolution freight outbound shipment data. Three methods are investigated, (i.e., proportional weighting, direct regression and optimal disaggregation model. Proportional weighting is the most popular and straightforward synthesis method used in existing freight studies, followed by direct regression method. Optimal disaggregation models are developed in economic research and applied in freight data in this paper. A case study of machinery outbound shipment values in State of Wisconsin suggests that proportional weighting and optimal disaggregation are promising tools. Limitations of the methods and future research are also identified at the end of the paper. The outcome of this research will be a set of applicable algorithms and national county level commodity flow database that can be used for regional truck emissions estimation.

INTRODUCTION

Freight activities in the United States continue to grow at a rapid pace. This increase, along with the continuing growth in passenger vehicle miles, place unprecedented strain on the nation's already choking highway system. Compared to passenger travel demand modeling, however, freight modeling in long-range regional transportation planning is rather primitive.

One of the biggest difficulties in freight modeling is the issue associated with freight data. As the National Cooperative Highway Research Program (NCHRP) Synthesis 298 *(1)* points out, "The increased importance of truck activity both in transportation engineering and planning has created a need for truck-oriented analytical tools. A particular planning need is for trip generation data that can be used to estimate truck traffic patterns, beginning with the ability to estimate truck trips generated by a variety of common land uses. However, the current state of the practice in truck trip generation data falls short of the needs of today's transportation engineers and transportation planners." Much of the freight movement and commodity shipment information is proprietary and guarded with strict confidentiality. Aggregate data, usually at the state level, is publicly available but of limited use for regional and local forecasting. Commercially available data products, e.g., TRANSEARCH by Global Insight, Inc and IMPLAN by Minnesota IMPLAN Group, are prohibitedly expensive for academic research. More importantly, the raw data and the data creation methodologies are a black box to the public, a least desirable situation to academic researchers. Currently, the best public data sources are the state level Commodity Flow Survey (CFS) and the

Federal Highway Administration's Freight Analysis Framework (FAF), an enhanced version of CFS available at the FAF region level. FAF regions are geographic units composed of 114 domestic regions, 17 import and export gateways and 7 international regions *(2)*. Domestic regions are defined based on metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs), and states or balances of states. In addition, a few regions have performed regional freight surveys, such as the Tampa Bay goods Movement Survey, Atlanta Area Commercial Vehicle Survey, etc. Overall, the limited availability of good freight data at sub-state levels imposes a big challenge in freight modeling.

This paper is aimed at exploring existing and potential synthesis methodologies in disaggregating freight data available in large geographic areas into small geographic areas. Section 2 briefly introduces the existing nationwide freight data, and section 3 present the existing synthesis studies on freight data. In section 4, three synthesis methods are explored, followed by the case study of machinery outbound shipments in Wisconsin in Section 5. Section 6 concludes the study and suggests the future work.

BRIEF OVERVIEW OF EXISTING FREIGHT DATA

Commodity Flow Survey 2002 (CFS 2002)

The CFS 2002 *(3)* is part of the Economic Census performed by the U.S. Census Bureau and U.S. Department of Commerce (DOC), partnered with the Bureau of Transportation Statistics (BTS) and the U.S. Department of Transportation (DOT). The survey contains a total of 50,000 establishment samples of commodity shipment information (weight, value, commodity category, and transportation mode) originated and destined in 50 states and Washington DC. The transportation modes included are truck (both private and for-hire trucks), railroad, waterway (shallow draft, the Great Lakes and deep draft), air (including truck-to-air), pipeline, multi-mode (including parcel, USPS or courier, truck-water, water-rail, truck-rail, and other multiple modes) and other unspecified/unknown modes.

The CFS 2002 covers business establishments in manufacture, mining, wholesale trade, selected retail trade industries (i.e., electronic stores, mail-order houses), and auxiliary establishments (i.e., warehouses and management offices), all of which were coded based on the 1997 North American Industry Classification System (NAICS). Certain industrial sectors were "out-of-scope" of the CFS 2002, including services, transportation, construction, and most retail industries. In addition, establishments like farms, fisheries, foreign- and most government-owned ones were excluded from the survey, although they belong to the "in-the-scope" industries. The shipment information from these establishments to processing centers, which are comprised primarily of local and short-haul movements, was therefore excluded. For example, the shipments of agriculture products from processing facilities to distribution centers were counted in the CFS but the shipments from farms to processing centers or terminal elevators were not reported. Therefore, the survey covers only 54% of the total freight shipments in tons, 67% in ton-miles and 63% in dollar values.

Public accessible CFS 2002 data provides the shipment distance and weight information (in tons, dollar values and ton-miles) in the state or metropolitan level by mode of transportation. More detailed state level origin-destination (O-D) matrices by 43 commodity classes and 7 modes/mode combinations are available upon request. It is worth mentioning that those O-D matrices contain a number of empty cells. Besides the reason discussed above, an empty cell could also be due to small sample size (and therefore producing unreliable estimates) or a true zero value. Special attention must be paid to handling the empty cells when using the CFS. Supplemental data sources are needed to fill in the missing information.

TRANSEARCH Database

TRANSEARCH may be considered the best available nationwide county level commodity flow data source. The 2001 TRANSEARCH Database is derived from various data sources (up to 100 sources) on commodity volumes (in the 4-digit STCC) and modal values in 172 business economic area (BEAs), 50 states and 3,145 U.S. counties. Seven transportation modes are included, i.e., truckload, less-than-truckload, private motor carrier, rail carload, rail intermodal, waterborne and air. The most crucial input data source is the Motor Carrier Data Exchange program, which includes the participation of truckload carriers and distributors in providing shipment information on size, origin-destination, and annual flows, etc. TRANSEARCH provides the O-D matrices in the state level that were then disaggregated into the BEA and the county level. The two-step disaggregation procedure involved separate allocation of commodities based on employment shares at origins and employment and population shares at destinations.

Mani and Prozzi *(4)* pointed out the main limitations of TRANSEARCH. Foremost, there are inevitable biases in regions and types of commodities due to limited number of participated carriers. Second, detailed shipment information was available only in four modes, which are trucks, rail, domestic air and domestic waterborne. Shipment information was also excluded for international air and petroleum, unprocessed agriculture and mining. Last but not the least, the validation information of the TRANSEARCH data was not available owing to the proprietary ownership.

Freight Analysis Framework (FAF)

The Federal Highway Administration (FHWA)'s Freight Analysis Framework (FAF) has been developed into Version 2.2, $FAF^{2.2}$, which is based on year 2002 and forecasts into years 2010 through 2035 in a 5-year interval the commodity flows origins and destinations, as well as the truck counts on the national highway network.

FAF is built on the CFS within-scope and out-of-scope sectors, and non-CFS data sources, with CFS as the core database. Non-CFS data sources include the counts missing in CFS but in other freight resources like Waterborne commerce or Waybill, etc. FAF contains four datasets of commodity movements, each in terms of commodity value and weight, in domestic activities, trans-border activities with Canada and Mexico, seaport international activities and air international activities. The domestic dataset contains all freight movements between domestic FAF regions—to be defined later in this paragraph, while all other datasets are assigned the international movements by mode with the origin or destination in a domestic FAF region. The international datasets contain information of origin/destination FAF regions, port of entry or exit, mode of the domestic leg, and commodity type in 2-digit SCTG. Specifically, the trans-border dataset captures the commodity movements between the U.S. and Canada and Mexico via land; the sea/air dataset describes the international movements via water/airplane. FAF regions definition adopts the 114 domestic geographic regions from CFS, which includes 64 metropolitan areas, and 50 other completed states or the rest of the states. In addition, 17 crucial international gateways not included in the metropolitan areas are selected from over 400 gateways around the country. The foreign market is divided into 7 regions, Canada, Mexico, South America, Asia, Europe, Middle East and the rest of the world. Collectively, there are a total of 138 FAF regions inclusive in the FAF O-D tables.

EXISTING FREIGHT DATA SYNTHESIS STUDIES

Florida Study

A Florida state study *(5)* disaggregated the FAF regional commodity O-Ds to the county level O-Ds. Regression models were fitted for the productions and attractions respectively for each commodity group from the freight data in all the 114 domestic FAF regions, relating the production or attraction annual tonnage with the 3-digit North American Industry Classification System (NAICS) employment rates, population and total employment within the FAF region. Goodness of fit results range from 0.09 for the Coal attraction to 0.94 for Lumber production. With the regression models, Florida county level productions and attractions were obtained and the county-to-region ratios were calculated. The county level O-D matrix was then augmented by multiplying the FAF region O-Ds by the county-to-region ratios. The results in one FAF region were compared with the TRANSEARCH data in terms of county percentage trends within the FAF region and found a fair degree of similarity. The authors defended their results by noting the discrepancies between FAF2.2 and TRANSEARCH. First of all, the commodity classification systems, STCC and SCTG, are different. Secondly, TRANSEARCH contains unlinked base commodity trips while the FAF considers linked trips. In short, this study implemented a direct regression disaggregation method enabling the inclusion of multiple covariates, but some commodities produced poor goodness of fit, which suggests the need for more robust disaggregation methods.

New Jersey Study

Another statewide *(6)* FAF data disaggregation project was conducted in New Jersey, which contains three FAF regions. Using the matrix-disaggregate function in TRANSCAD, the authors explored the proportional weighting method – to be explained in next section – by employing six separate sets of proportional parameters, including employment, population, income, Truck Vehicle Miles Traveled (TVMT), number of trucks going out and in, etc. Again, the results were compared with TRANSEARCH. The authors claimed the results for all the models were acceptable, especially for the model using trucks going out and coming in, and the other using employment adjusted and Income (25-54 yrs) adjusted TVMT. The study is valuable in attempting different proportional surrogate variables to disaggregating commodity flows from FAF regions to counties, but finer disaggregation methods need to be explored further.

SYNTHESIS METHODS

Proportional Weighting

The most widely used allocation method is proportional weighting, which allocates values from a source unit to target units proportional to some surrogate variable. The source unit typically contains the target units geographically. The relationship of the surrogate variable with the variable of interest is crucial to the errors of the results. The steps used to perform proportional weighting are as follows:

Step 1: Identify the appropriate surrogate variable.

Step 2: Calculate the values of the surrogate variable in the counties and sum them up to the FAF regions.

Step 3: Calculate the percent proportions of each child county to the parent FAF region by the surrogate variable.

Step 4: Allocate the FAF region commodity outbound totals proportionally to counties based on the percentage shares as a result of Step 3.

The selection of the surrogate variable is critical to the quality of disaggregation results. Past studies like the Stammer and Pratt's *(7)* found employment by industry was a reasonable surrogate variable. A Wisconsin study *(8)* allocated commodity flows to counties based on employment share by producing economic sector. Another study in Los Angeles metropolitan area *(9)* allocated production and attraction commodities based on employment share by producing and consuming sector. In this research, total employment of industries related to the commodity of interest is used as the surrogate variable in proportional weighting.

Direct Regression

The direct regression method, similar to that in Chang and Liu's study (*10)*, involves two steps: first, an outbound shipment regression model is developed for the FAF regions, and then the model is applied to estimating the county level outbound shipments. The underlying assumption is that the same relationship between the outbound shipments and explanatory variables applies to both an FAF region and its inclusive counties. The linear relationship used in direct regression is expressed as follows:

$$\mathbf{Y_F} = \mathbf{X_F}\boldsymbol{\beta} + \mathbf{u_F} \tag{1}$$

$$\mathbf{Y_c} = \mathbf{X_c}\boldsymbol{\beta} + \mathbf{u_c} \tag{2}$$

where $\mathbf{Y_F}$ is a vector of the FAF regional total outbound shipment values; $\mathbf{Y_c}$ is a vector of the county total outbound shipment values; $\mathbf{X_F}$ is a vector of the FAF regional explanatory variables; $\mathbf{X_c}$ is a vector of the county explanatory variables. Coefficient vector $\boldsymbol{\beta}$ is estimated in (1).

Optimal Disaggregation Model

The direct regression method may produce unbalanced estimations between the upper and lower levels. That is, there is no guarantee the sum of the lower level (i.e., counties) estimates returns the value of the upper level (i.e., the FAF region containing those counties). To address this problem, we adopt an optimal disaggregation model proposed by Chow and Lin (*11*). It is essentially a best linear unbiased regression models, one at the county level and the other at the FAF regional level, linked by a disaggregation matrix that converts the FAF regional value to the inclusive county values. By introducing the disaggregation matrix it is guaranteed that the county level estimates will add up to the regional total. The best linear unbiased regression is estimated by minimizing the trace of the covariance matrix of residuals.

In mathematical language, the optimum disaggregation model is formulated as follows. Suppose

$$\mathbf{Y}_F = \mathbf{X_F}\boldsymbol{\beta} + \mathbf{u_F} \tag{3}$$

where $\mathbf{Y_F}$ is a vector of FAF regional values of a commodity type; $\mathbf{X_F}$ is a matrix of independent variables; $\mathbf{u_F}$ is the error vector of the FAF region model. Let $\mathbf{Y_c}$ be the vector of the county total values of the same commodity type and $\mathbf{u_c}$ be the error vector of the county model. Introduce also the disaggregate matrix $\mathbf{D}$ between an FAF region and inclusive counties. Thus,

$$\hat{\mathbf{Y}}_\mathbf{c} = \mathbf{DY_F} = \mathbf{D(X_F\boldsymbol{\beta} + u_F)} \tag{4}$$

$$\text{and,}\quad \mathrm{E}(\hat{\mathbf{Y}}_\mathbf{c} - \mathbf{Y_c}) = \mathrm{E}[\mathbf{D(X_F}\boldsymbol{\beta} + \mathbf{u_F}) - (\mathbf{X_c}\boldsymbol{\beta} + \mathbf{u_c})] = (\mathbf{DX_F} - \mathbf{X_c})\boldsymbol{\beta} = \mathbf{0} \tag{5}$$

From (4) and (5), we have $\mathbf{DX_F} - \mathbf{X_c} = \mathbf{0}$ and $\hat{\mathbf{Y}}_\mathbf{c} - \mathbf{Y_c} = \mathbf{Du_F} - \mathbf{u_c}$. So,

$$\begin{aligned} \mathrm{cov}(\hat{\mathbf{Y}}_\mathbf{c} - \mathbf{Y}_c) &= \mathrm{E}[(\mathbf{Du_F} - \mathbf{u_c})(\mathbf{Du_F} - \mathbf{u_c})'] \\ &= \mathbf{DV_FD}' - \mathbf{DV_{Fc}} - \mathbf{V_{cF}D}' + \mathbf{V_c} \end{aligned} \tag{6}$$

where $\mathbf{V_F}$ denotes $\mathrm{E}\,\mathbf{u_F u_F}'$ and $\mathbf{V_{Fc}}$ denotes $\mathrm{E}\,\mathbf{u_F u_c}'$. For the best linear unbiased estimation, set $\mathrm{cov}(\hat{\mathbf{Y}}_\mathbf{c} - \mathbf{Y_c}) = \mathbf{0}$, and hence the trace of (6), $tr(\mathbf{DV_FD}' - \mathbf{DV_{Fc}} - \mathbf{V_{cF}D}' + \mathbf{V_c})$, is to be minimized. Solving for the minimal trace, we finally obtain the following:

$$\begin{cases} \mathbf{D} = \mathbf{X_c}(\mathbf{X_F^{'}}\mathbf{V_F^{-1}}\mathbf{X_F})^{-1}\mathbf{X_F^{'}}\mathbf{V_F^{-1}} + \mathbf{V_{cF}}\mathbf{V_F^{-1}}(\mathbf{I} - \mathbf{X_F}(\mathbf{X_F^{'}}\mathbf{V_F^{-1}}\mathbf{X_F})^{-1}\mathbf{X_F^{'}}\mathbf{V_F^{-1}}) \\ \hat{\mathbf{Y}}_\mathbf{c} = \mathbf{D}\hat{\mathbf{Y}}_\mathbf{F} = \mathbf{X_c}\hat{\boldsymbol{\beta}} + (\mathbf{V_{cF}}\mathbf{V_F^{-1}})\hat{\mathbf{u}}_\mathbf{F} \\ \hat{\boldsymbol{\beta}} = (\mathbf{X_F^{'}}\mathbf{V_F^{-1}}\mathbf{X_F})^{-1}\mathbf{X_F^{'}}\mathbf{V_F^{-1}}\mathbf{Y_F} \end{cases} \quad (7)$$

and the FAF region model error vector is

$$\begin{aligned} \hat{\boldsymbol{\mu}}_\mathbf{F} &= \mathbf{Y_F} - \mathbf{X_F}\hat{\boldsymbol{\beta}} \\ &= (\mathbf{I} - \mathbf{X_F}(\mathbf{X_F^{'}}\mathbf{V_F^{-1}}\mathbf{X_F})^{-1}\mathbf{X_F^{'}}\mathbf{V_F^{-1}})\mathbf{Y_F} \end{aligned} \quad (8)$$

All above parameters of interest are dependent on the variance and covariance matrices, $\mathbf{V_F}$ and $\mathbf{V_{Fc}}$, which must be estimated by assuming some variance-covariance structure to the error terms. This will be elaborated further in the results section.

CASE STUDY IN WISCONSIN

A pilot study is performed to investigate the three aforementioned disaggregation methods, i.e., proportional weighting, direct regression and the optimal disaggregation model, in allocating Wisconsin' Machinery outbound shipments into counties.

Wisconsin FAF Data

The State of Wisconsin has two FAF regions and they are, as shown in Figure 1,

- Zone 112: Milwaukee-Racine-Waukesha, which contains 5 counties
- Zone 113: The rest of Wisconsin, which contains 67 counties

Table 1 lists the top 10 commodities that have the highest total outbound shipping value in Wisconsin. As can be seen, the commodity with the highest outbound shipment value in Wisconsin is machinery. Collectively, the sum of these 10 commodities accounts for 64% of the total outbound shipments values of 43 commodity types in Wisconsin. Pilot study will be performed for the top 1 commodity-Machinery.

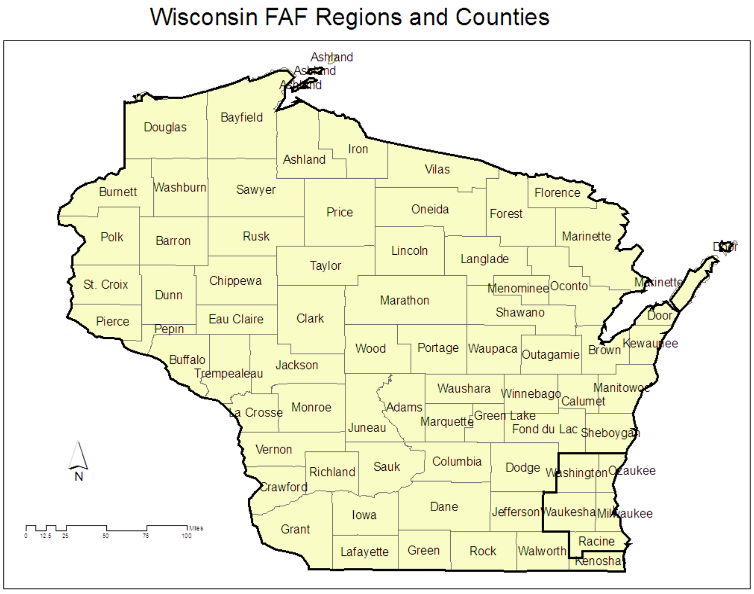

FIGURE 1 Wisconsin FAF Regions And Counties

TABLE 1 Commodity Outbound Shipment Values In The State of Wisconsin

SCTG	Commodity	Value (Million Dollars)
34	Machinery	34,706.46
42	Mixed freight	30,897.32
7	Other foodstuffs	26,245.02
36	Motorized vehicles	22,925.23
35	Electronics	14,257.00
40	Misc. mfg. prods.	12,294.91
33	Articles-base metal	10,533.90
30	Textiles/leather	9,397.40
28	Paper articles	9,143.73
43	Unknown	8,746.17

Surrogate Variables In Proportional Weighting

As suggested in the literature, total employment of an industry is a good indicator of the total commodity outbound shipments in an area. Hence, the total employment in machinery related industries is chosen as the surrogate variable proportional weighting disaggregation. According to the NAICS to SCTG mapping provided by FHWA *(12)*, the industries associated with machinery production are listed in Table 2.

TABLE 2 Machinery Related Industries

NAICS code	Description
314999	All other miscellaneous textile product mills
316110	Leather and Hide Tanning and Finishing
322291	Sanitary Paper Product Manufacturing
323122	Prepress Services
332212	Hand and Edge Tool Manufacturing
332410	Power Boiler and Heat Exchanger Manufacturing
332439	Other Metal Container Manufacturing
332710	Machine Shops
332991	Ball and Roller Bearing Manufacturing
332997	Industrial Pattern Manufacturing
333120	Construction Machinery Manufacturing
333131	Mining Machinery and Equipment Manufacturing
333132	Oil and Gas Field Machinery and Equipment Manufacturing
333210	Sawmill and Woodworking Machinery Manufacturing
333291	Paper Industry Machinery Manufacturing
333292	Textile Machinery Manufacturing
333293	Printing Machinery and Equipment Manufacturing
333294	Food Product Machinery Manufacturing
333298	All Other Industrial Machinery Manufacturing
333311	Automatic Vending Machine Manufacturing
333411	Air Purification Equipment Manufacturing
333414	Heating Equipment (except Warm Air Furnaces) Manufacturing
333415	Air-Conditioning and Warm Air Heating Equipment and Commercial and Industrial Refrigeration Equipment Manufacturing
333511	Industrial Mold Manufacturing
333512	Machine Tool (Metal Cutting Types) Manufacturing
333513	Machine Tool (Metal Forming Types) Manufacturing
333516	Rolling Mill Machinery and Equipment Manufacturing
333611	Turbine and Turbine Generator Set Units Manufacturing
333613	Mechanical Power Transmission Equipment Manufacturing

333618	Other Engine Equipment Manufacturing
333912	Air and Gas Compressor Manufacturing
333913	Measuring and Dispensing Pump Manufacturing
333921	Elevator and Moving Stairway Manufacturing
333922	Conveyor and Conveying Equipment Manufacturing
333923	Overhead Traveling Crane, Hoist, and Monorail System Manufacturing
333993	Packaging Machinery Manufacturing
333994	Industrial Process Furnace and Oven Manufacturing
333995	Fluid Power Cylinder and Actuator Manufacturing
333996	Fluid Power Pump and Motor Manufacturing
333997	Scale and Balance (except Laboratory) Manufacturing
334516	Analytical Laboratory Instrument Manufacturing
335222	Household Refrigerator and Home Freezer Manufacturing
335224	Household Laundry Equipment Manufacturing
336211	Motor Vehicle Body Manufacturing
336412	Aircraft Engine and Engine Parts Manufacturing
336415	Guided Missile and Space Vehicle Propulsion Unit and Propulsion Unit Parts Manufacturing
337129	Wood Television, Radio, and Sewing Machine Cabinet Manufacturing
339993	Fastener, Button, Needle, and Pin Manufacturing

Possible Covariates In Direct Regression and Optimal Disaggregration Regression

In this pilot study, possible covariates for building the regression models are designated to be employment counts (EMP) in related industries and number of inter-modal facilities (see Table 3), including those listed in Table 2. The main advantage for using these variables, in addition to their wide success in previous studies reviewed above, is the detailed and completed data that can be obtained from CBP 2002.

Table 3 has included industries directly related to four transportation modes (i.e., air, railroad, water, truck) that carry a majority of machinery commodity shipments. The numbers of workers in freight transportation related industries (e.g., scheduled freight air transportation, line-haul railroads, etc.) are indicators of the scale of the freight industry. Besides, eleven freight transportation supporting activities (services) are considered, including rail transportation support activities, port & harbor operations, freight transportation arrangement, etc. Two other industries, general warehousing and storage and other warehousing and storage, are included to account for commodity shipments originated at warehouses. Lastly, the number of inter-modal facilities suggests the scale of inter-modal freight shipments in the area. Inter-modal facility counts data are from the National Atlas.

TABLE 3 List Of Possible Covariates (In Addition To Table 2)

Category	Description
Air Transportation	EMP in Scheduled freight air transportation
	EMP in Nonscheduled chartered freight air trans
Railroad Transportation	EMP in Line-Haul Railroads
	EMP in Short Line Railroads
Water Transportation	EMP in Deep sea freight transportation
	EMP in Coastal & Great Lakes freight transportation
	EMP in Inland water freight transportation
Truck transportation	EMP in General freight trucking, local
	EMP in General freight trucking, long-distance, TL
	EMP in General freight trucking, long-distance, LTL
	EMP in Specialized freight (exc used) trucking, local
	EMP in Specialized freight (exc used) trucking, Ldist
Transportation support activities	EMP in Air traffic control
	EMP in Other airport operations
	EMP in Other air transportation support activities
	EMP in Rail transportation support activities
	EMP in Port & harbor operations
	EMP in Marine cargo handling
	EMP in Navigational services to shipping
	EMP in Other water transportation support activities
	EMP in Other road transportation support activities
	EMP in Freight transportation arrangement
	EMP in All other transportation support activities
Warehousing & storage	EMP in General warehousing & storage
	EMP in Other warehousing & storage
Inter-modal Facility Counts	Number of inter-modal facilities

Model Results

Proportional Weighting (PW)

Total machinery outbound shipment values in FAF region 112 and 113 are 11553.05 million dollars and 23153.41 million dollars. Total employments in all t48 industries listed in Table 2 are calculated for each county. After that, the total employments in FAF region 112 and 113 are obtained by adding the inclusive counties' total employments respectively. Employment shares of each county within the FAF region are then calculated by dividing the county total by that FAF region total. Finally, the county outbound shipment values are allocated based on the employment shares of counties.

According to the proportional weighting results, outbound shipment values of the 72 counties within Wisconsin range from 0 to 4756.80 million dollars, with mean value 482 million dollars. Figure 2 below shows the value distribution of machinery among Wisconsin counties from the proportional weighting disaggregation results of FAF region data. Color coding in the map represents the normalized outbound shipment values of the counties to the state total.

The normalized TRANSEARCH estimates are color coded in the same scale in Figure 3. To compare the result with that in TRANSEARCH, the commodity coding must first match up because TRANSEARCH uses the STCC coding system. According to the mapping provided in Quick Response Freight Manual II *(13)*, machinery commodities coded as SCTG 34 are essentially equivalent to STCC 35. Consequently, only the commodities coded as "35" in TRANSEARCH are used for comparison.

In comparison, the PW estimates show a consistent spatial distribution pattern to that of the TRANSEARCH estimates. Machinery outbound shipments in Wisconsin are highly concentrated in the Milwaukee-Racine-Waukesha Metropolitan area and nearby counties located in the southeast part, which is home to several major transportation equipment and machinery manufacturers in Wisconsin. Companies in these categories include the Kohler Company, Rockwell Automation, Johnson Controls, Briggs & Stratton, Miller Electric, Milwaukee Electric Tool Company, Bucyrus International. Counties in the north and southwest appear to have the least shares of the machinery outbound shipments. In the west and northwest areas, the percentages of machinery outbound shipments of the proportional results are generally a little higher than those of the TRANSEARCH database. One reason for this discrepancy may be that the labor productivity in these counties is lower than the other counties.

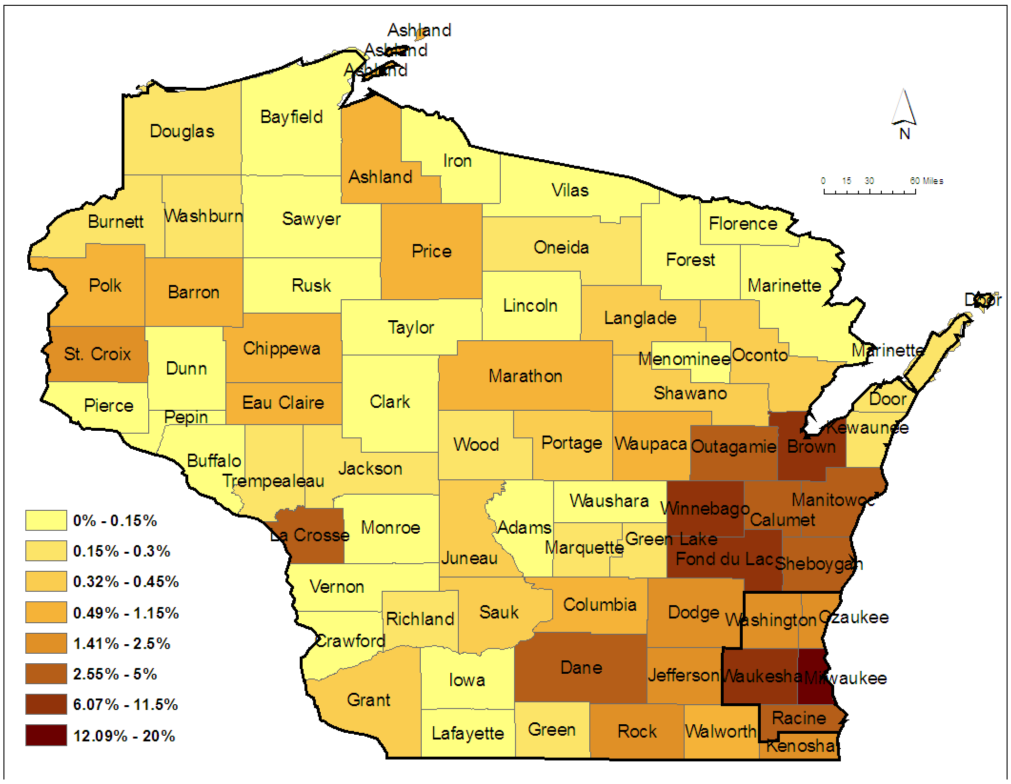

FIGURE 2 Proportional Weighting Results: County Machinery Outbound Shipment Values Percentages Within Wisconsin

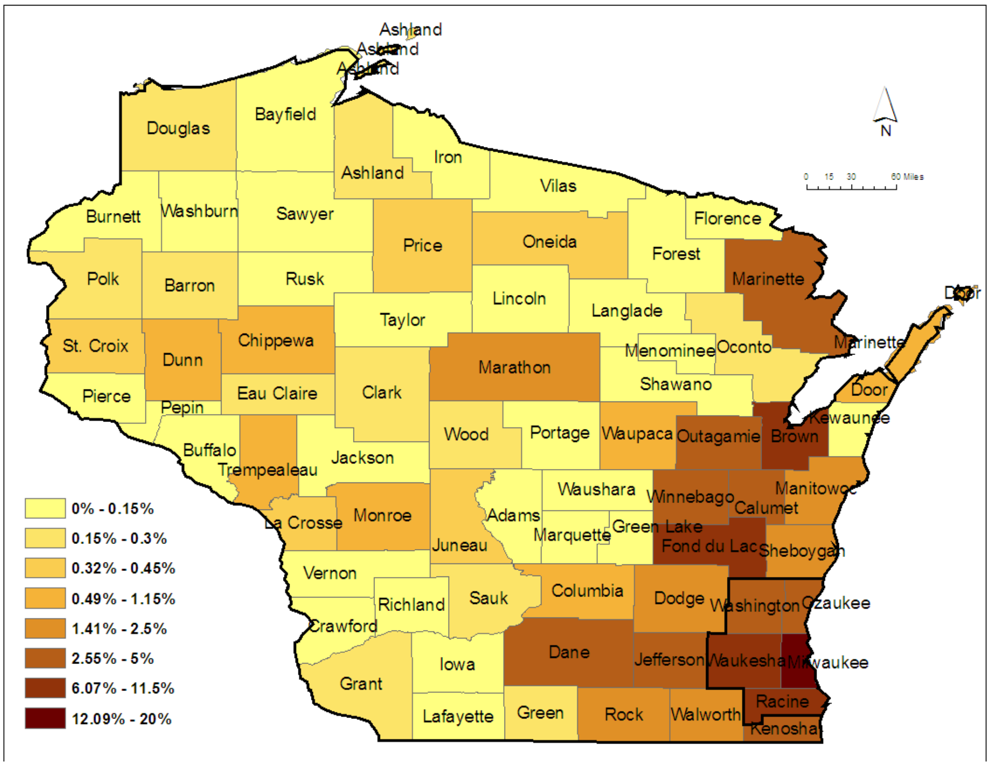

FIGURE 3 TRANSEARCH Results: County Machinery Outbound Shipment Weights Percentages Within Wisconsin

Direct Regression (DR)

Model parameters were estimated using the stepwise linear regression fitting. Square transformations of some of the independent variables were found necessary after the initial model residuals were examined. After many rounds of trial and error, the final model specification is shown in Table 4. Only the covariates with significant coefficients are presented. In this model, the intercept has been forced to zero because it is assumed that a region with zero employment in any related industries does not produce or ship machinery products. The model has an adjusted R-square value of 0.834, which means almost 85% of the variation in the data is accounted for in the final model.

All model coefficients' signs are intuitively correct. The largest coefficient value in freight transportation arrangement employment implies that the value per employment produces is the highest among all the machinery related industries. Number of inter-modal facilities is shown to be highly significantly and positively related to the outbound shipment values, suggesting that the presence of inter-modal facilities in the same FAF region will increase the total outbound shipment values. Likewise, employment in air purification equipment manufacturing, air and gas compressor manufacturing are found to be positively related to outbound machinery shipments values.

TABLE 4 Direct Regression Model Specifications

Model	Coefficient	Std. Error	T-stat	Sig.
Employment of Freight Transportation Arrangement	2.607	0.477	5.465	0.000
Employment of Air Purification Equipment Manufacturing (Square)	0.058	0.008	7.403	0.000
Employment of Air and Gas Compressor Manufacturing (Square)	0.006	0.002	2.722	0.000
Facility counts (hundred)	2.784	0.380	7.332	0.000
Dependent Variable: Mdol_total				

By applying the final model in Table 4, the county level outbound shipment total values are calculated for the State of Wisconsin. The normalized estimated results are shown in Figure 4. Compared to Figure 3, the southeast counties in both plots are shown accounting for the largest portions. Overall, the direct regression results are shown quite a different pattern than the TRANSEARCH.

There are several reasons to explain such large discrepancies observed in the direct regression results. Firstly and most importantly, directly applying the regional model to calculating for the county values implies that the same model specification and the values of the coefficients at the FAF regional level are transferable to the county level. This is undoubtedly a very strong assumption. For example, the number of employment in freight transportation arrangement in a county does not necessary translate into high outbound shipment values in that county, because the service may be provided in a nearby county within the same FAF region. In addition, employments in some small industries not included in the final model may in fact have a significant effect on counties. Consider the following scenario. One county produces one type of machinery product that is not produced in any other county of the same FAF region and the total outbound shipment value makes up a comparatively small portion of the total outbound shipment value of all machinery products. In such a case, that particular industry producing the product would mostly likely fall off the regional model such that county's outbound shipment value would be underestimated. Lastly, the sizes of the inter-modal facilities are not available. So the capacity of the inter-modal facilities, which is more critical than the count, is not available. This is an important limitation to the regression model.

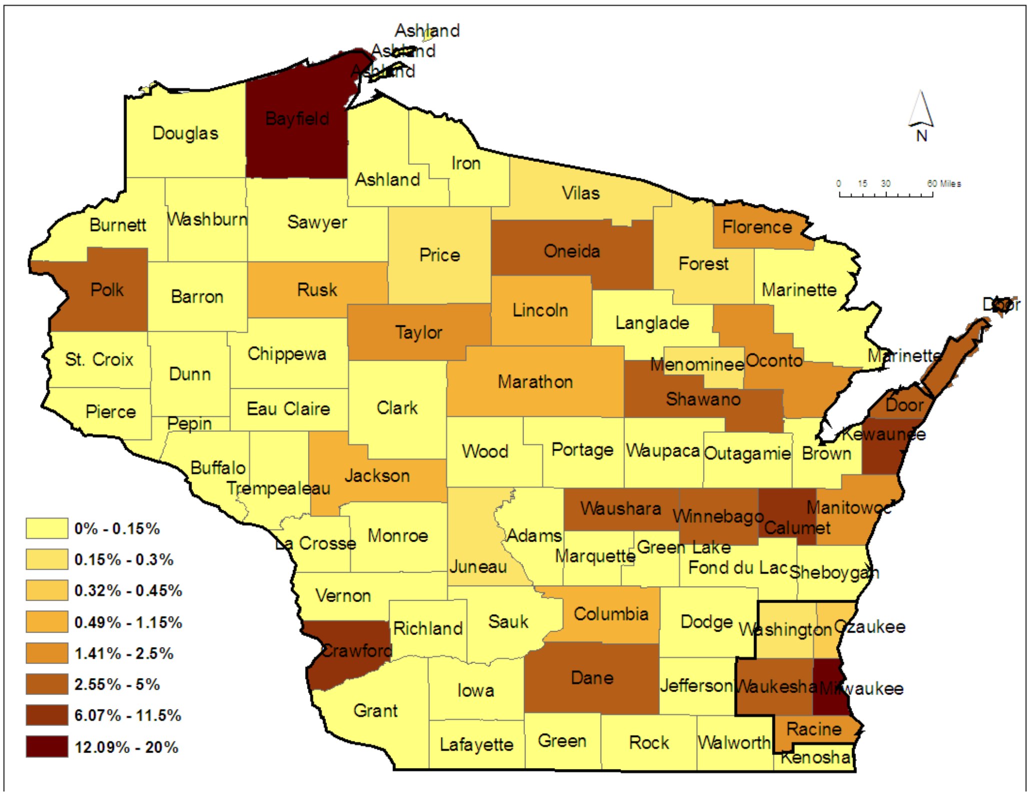

FIGURE 4 Direct Regression Results: County Machinery Outbound Shipment Values Percentages Within Wisconsin

Optimal Disaggregation Model-Uncorrelated Residual With Variance σ^2 *(ODM)*

As stated in the synthesis methods section, calculating the estimator $\hat{Y}_c$ requires the assumption of some structure to the residuals of the county regression. Supposed that the direct regression model specifications in Table 4 in the FAF regional level, the optimal disaggregation model is estimated by assuming that county level residuals are uncorrelated, each with variance σ^2. Since $\mathbf{V}_{\mathbf{cF}} = \mathrm{E}\,\mathbf{u}_{\mathbf{c}}\mathbf{u}_{\mathbf{F}}' = \mathbf{D}'\boldsymbol{\sigma}^2$ and $\mathbf{V}_{\mathbf{F}} = \mathbf{D}\mathbf{D}'\boldsymbol{\sigma}^2$, $\mathbf{V}_{\mathbf{cF}}\mathbf{V}_{\mathbf{F}}^{-1}$ in equation (7) will be reduced to $\frac{1}{n_i}\mathbf{D}'\hat{\mathbf{u}}_{\mathbf{F}}$, where n_i is the number of counties that the i th FAF region has, e.g., 5 for FAF region 112 and 67 for FAF region 113. This structure implies that the FAF region residuals are equally distributed into the counties within the region. Estimation results are shown below in Figure 5. As can be seen, the machinery outbound shipment values percentages in some counties (i.e., Washington, Ozaukee, Dunn Douglas, Kewaunee, etc.) are more consistent with the TRANSEARCH database than those in the DR model, suggesting that ODM enhanced the DR model in some extent. However, the estimations in some other counties, like Lafayette, Burnett and Dane, are worse.

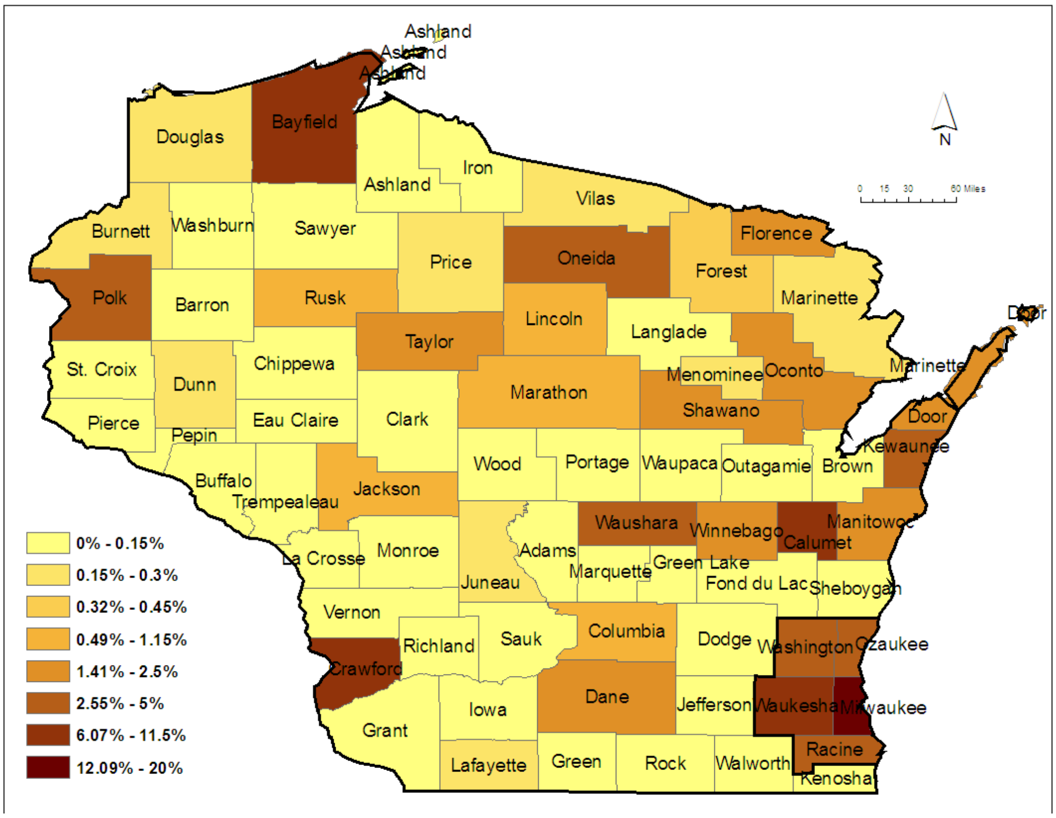

FIGURE 5 Optimal Disaggregation Model Results: County Machinery Outbound Shipment Values Percentages Within Wisconsin

Optimal Disaggregation Model -Uncorrelated Residual With Proportional Variance (ODM_PV)

This study explored another model with residuals that are uncorrelated but proportional variance. It is assumed that the residual variance is proportional to employment share within the FAF region. The model results are shown in Figure 6. Noticeably the ODM_PV model has improved the results over the ODM model, such as those of Lafayette and Menominee, while worsen those in Douglas, Richland and Marinette.

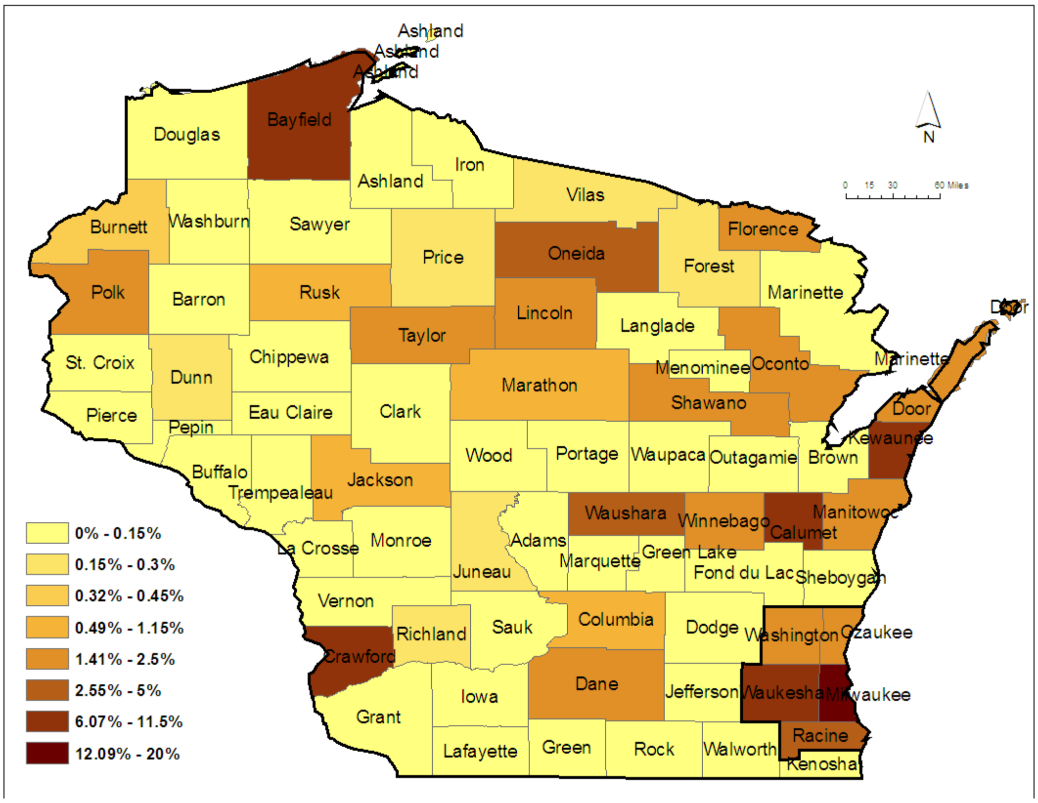

FIGURE 6 Optimal Disaggregation Model With Proportional Variance Results: County Machinery Outbound Shipment Values Percentages Within Wisconsin

Finally, Table 5 summarizes the correlations of the models' estimation results. Proportional weighting ranks the highest (0.91), while direct regression poorly related to the TRANSEARCH (0.48). Table 5 and the results plots above suggested that the optimal disaggregation models improve the direct regression results but still not outperformed by the proportional weighting method.

TABLE 5: Correlation Of Synthesis Results

	TRANSEARCH	PW	DR	ODM	ODM_PV
TRANSEARCH	1.00				
PW	0.91	1.00			
DR	0.48	0.97	1.00		
ODM	0.63	0.98	1.00	1.00	
ODM_PV	0.65	0.97	1.00	1.00	1.00

CONCLUSION AND FUTURE WORK

This paper explores three synthesis methods in generating high geographic resolution freight data, which is of great value in the regional freight modeling. A case study performed in machinery outbound shipment values in Wisconsin shows that proportional weighting method provides relatively good quality results when compared with TRANSEARCH database. Direct regression gives reluctant results, mainly due to the strong assumption that the FAF region specifications can be transferable to the county level. The optimal disaggregation models, assured the total sum equal the FAF region total, noticeably improves the direct regression estimates. It is worth pointing out that the above conclusions are based on the comparisons with the TRANSEARCH data. More robust validation should be to compare the estimated results with true regional samples (e.g., from a regional survey).

There is ongoing investigation of these and other synthesis methods in other commodity types and states. After that, shipments' values will be converted to trips by mode, based on which the traffic assignment will be performed to obtain the county level VMT. Finally, the county level freight VMT will be used as input to the MOVES model for estimating the national county-level freight emissions inventories.

ACKNOWLEDGEMENT

This study is funded by the National Center for Freight and Infrastructure Research and Education (CFIRE) at University of Wisconsin, Madison. We thank our colleague Dr. Kazuya Kawamura of Urban Transportation Center at the University of Illinois at Chicago for his support and insight to the subject matter.

REFERENCE

1. National Cooperative Highway Research Program Synthesis 298 - *Truck Trip Generation Data: a synthesis of highway practice.* Transportation Research Board, National Research Council. Washington D.C., 2001.

2. Freight Analysis Framework (FAF) *Version 2.2*, *User Guide*. Federal Highway Administration (FHWA), U.S. Department of Transportation (USDOT). http://ops.fhwa.dot.gov/freight/freight_analysis/faf/faf2userguide/index.htm. Access on May 7, 2008

3. Commodity Flow Survey (CFS), www.bts.gov, Bureau of Transportation Statistics, 2008

4. Mani, A., Prozzi, J.: *State-of-the-Practice in Freight Data: A Review of Available Freight Data in the U.S.* Project Report for Development of Sources and Methods for Securing Truck Travel Data in Texas (0-4713-P2), 2004.

5. Viswanathan, K., Beagan, D., Mysore, V. and Srinivasan, N.: Disaggregating FAF^2 Data For Florida: Methodology and Results. *Transportation Research Board Annual Meeting*

CD-ROM, National Research Council, Washington, D.C., 2008.

6. Rowinski, J., Opie, K. and Spasovic, L.N.: Development of a Method to Disaggregate the 2002 FAF^2 Data Down to the County Level For New Jersey. *Transportation Research Board Annual Meeting CD-ROM*, National Research Council, Washington, D.C., 2007.

7. Stammer, R.E. and Pratt, G..: *Statewide Modeling Practices and Prototype Statewide Model Development for Tennessee*. Final Report Prepared for Tennessee Department of Transportation In Cooperation with U.S. Department of Transportation, Federal Highway Administration, April 2002.

8. Sorratini, J.A. and Smith, R.L.Jr.: Development of a Statewide Truck Trip Forecasting Model Based on Commodity Flows and Input-Output Coefficients. *Journal of the Transportation Research Board, Transportation Research Record* 1707, pp. 49–55, 2000

9. Fischer, M., Ang-Olson, J. and La, A.: External Urban Truck Trips Based on Commodity Flows: A Model. *Journal of the Transportation Research Board, Transportation Research Record* 1707, pp. 73–80, 2000.

10. Chang, C. and Liu, T. Monthly Estimates of Certain National Product Components, 1946-49, *The Review of Economics and Statistics*, XXXIII, 219-227, 1951.

11. Chow, C., Lin, A. Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by related Series. *The Review of Economics and Statistics*, Vol 53, No 4, 372-375, Nov 1971.

12. Methodology for the Freight Analysis Framework-2: Forecasts of Inter-regional Commodity Flows. Federal Highway Administration (FHWA), U.S. Department of Transportation (USDOT). http://ops.fhwa.dot.gov/freight/freight_analysis/faf/faf2_reports/reports8/s5_naics.htm#toc5a. Access on May 7, 2008

13. Federal Highway Administration: *Quick Response Freight Manual II.* FHWA Publication Number: FHWA-HOP-08_010 (EDL 14396), 2007.