Methodology for estimating ton-miles of goods movements for U.S. freight multimodal network system

Paper No. 13-4963

Original Manuscript Submitted: 08.01.2012
Revised Manuscript Submitted: 11.15.2012

Words: (5,369+ 7*250) = 7,119

by

Francisco Moraes Oliveira-Neto, Ph.D. 1
Post-doctoral Research Associate
(865) 946-1479, (865) 946-1314 (fax), oliveiranefm@ornl.gov

Shih Miao Chin, Ph.D.
(865) 946-1254 (office), (865) 946-1314 (fax), chins@ornl.gov

Ho-ling Hwang, Ph.D.
(865) 946-1224 (office), (865) 946-1314 (fax), hwanghl@ornl.gov

Bruce Peterson
(865) 946-1352 (office), (865) 946-1314 (fax), petersonbe@ornl.gov

Center for Transportation Analysis
Oak Ridge National Laboratory
2360 Cherahala Boulevard, Knoxville, TN 37932

1 Corresponding Author
ABSTRACT
Ton-miles is a commonly used measure of freight transportation output. Estimation of ton-miles in the U.S. transportation system requires freight flow data at disaggregated level (either by link flow, path flows or origin-destination flows between small geographic areas). However, the sheer magnitude of the freight data system as well as industrial confidentiality concerns in Census survey, limit the freight data which is made available to the public.

Through the years, the Center for Transportation Analysis (CTA) of the Oak Ridge National Laboratory (ORNL) has been working in the development of comprehensive national and regional freight databases and network flow models. One of the main products of this effort is the Freight Analysis Framework (FAF), a public database released by the ORNL. FAF provides to the general public a multidimensional matrix of freight flows (weight and dollar value) on the U.S. transportation system between states, major metropolitan areas, and remainder of states.

Recently, the CTA research team has developed a methodology to estimate ton-miles by mode of transportation between the 2007 FAF regions. This paper describes the data disaggregation methodology. The method relies on the estimation of disaggregation factors that are related to measures of production, attractiveness and average shipments distances by mode service. Production and attractiveness of counties are captured by the total employment payroll. Likely mileages for shipments between counties are calculated by using a geographic database, i.e. the CTA multimodal network system. Results of validation experiments demonstrate the validity of the method. Moreover, 2007 FAF ton-miles estimates are consistent with the major freight data programs for rail and water movements.

KEY WORDS: commodity flow survey, freight analysis framework, freight data disaggregation, ton-miles.
1 INTRODUCTION

Despite the daunting complexities of freight system performance measurement, freight performance measures have gaining importance and it appears that efforts to evaluate the freight system performance will likely be undertaken. A performance-based federal transportation program has been strongly endorsed by the National Surface Transportation Policy and Revenue Study Commission (1), as well as the Government Accountability Office (GAO) (2). Various Congressional proposals related to reauthorization (1-5) of federal transportation programs include new provisions requiring a setting of targets and measurement of progress. The American Association of State Highway and Transportation Officials (AASHTO) has formed a series of committees and task forces to recommend a set of national transportation performance metrics (6). Several states, including Washington, Iowa, and Minnesota, have added a handful of freight-related performance measures to their suite of performance metrics.

The most basic and commonly used measure of freight system performance is ton-miles (one ton of freight shipped over one mile). This measure is the primary physical measure of freight transportation output. It has a similar meaning as other transportation measures, vehicle-miles-travelled (VMT) or person-miles-travelled (PMT), used for passenger transportation to measure the person or vehicle throughput in urban or inter-city network systems. Ton-miles reflect how the transportation system is actually used for moving freight.

Freight flow data at disaggregated level (either network flow links or origin-destination flows between counties) is required to estimate ton-miles in the U.S. transportation system. Although considerable mode and shipper specific data is collected, the sheer magnitude of the freight data system as well as industrial confidentiality concerns in Census survey, limit the freight data which is made available to the public. Typically, a considerable portion of the freight data is not disclosed by public agencies and the portion available is restricted to high levels of geographic aggregation (e.g. at metropolitan or state level).

Through the years, the Center for Transportation Analysis (CTA) of the Oak Ridge National Laboratory (ORNL) has provided assistance to the Federal Highway Administration (FHWA) in the development of comprehensive national and regional freight databases and network flow models. One of the main products of this effort is the Freight Analysis Framework (FAF), a public database released by ORNL, which is includes U.S. domestic and international (imports and exports) freight flows (7). The main source of FAF is the Commodity Flow Survey (CFS, which is a shipper-based survey) carried out every five years as part of the U.S. Economic Census (8). Given that the CFS does not contemplate all industry sectors and that a considerable portion of the data is undisclosed, FAF complements the CFS data by estimating the missing cells and integrating other sources for sectors not covered by the survey. The final FAF product provides estimates of U.S. freight flows only by aggregated levels of detail, i.e. flows between states, major metropolitan areas and remainder of states.

National transportation freight analysis and planning usually requires data at reasonable level of detail. It is agreed by the transportation community that the minimum resolution level for such studies is county-to-county flows. Recently, the CTA research team has developed a methodology to disaggregate the FAF flows to U.S. county level of geography, and estimate ton-miles. In this paper, we describe the data disaggregation method deployed, and discuss how we validate the methodology by using data from CFS tabulations and supplemented with data on economic activity.
The remainder of the paper is organized as follows. Section 2 describes the framework for freight data developed by the CTA team. Section 3 presents the methodology for estimating freight flows and ton-miles at the county level. Section 4 presents the results, including validation of the method, and the estimated total ton-miles of the 2007 FAF database. Section 5 concludes.

2 FREIGHT DATA

The CFS is the most comprehensive survey of freight activities in the U.S. The survey is conducted every five years as a part of the Economic Census by the U.S. Census Bureau, in partnership with Bureau of Transportation Statistics (BTS) within the Research and Innovative Technology Administration (RITA) of the Department of Transportation (DOT). The last survey was conducted in 2007 and previous years of study include 1997, and 2002.

The survey includes freight activities originated in U.S., thus, it does not include imports movements. The sample survey of 2007 covered business establishments with paid employees located in the U.S., and classified according to the 2002 North American Industry Classification System (NAICS) industry codes in mining, manufacturing, wholesale trade, and select retail trade industries (e.g. electronic, stores and mail-order houses). In addition it also covered auxiliary establishments (e.g. warehouses and managing offices) of multi-establishment companies with shipping activity. Approximately 102,000 establishments were selected from a universe of about 760,000 “in-scope” establishments. “In-scope” is a term used to refer to industries that are covered by CFS sample. Conversely a “out-scope” industry is not included in the CFS sample frame. Table 1 presents summary statistics of the total tons by major sectors based on data from 2007 CFS.

<table>
<thead>
<tr>
<th>Industry</th>
<th>Tons (thousands)</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td>4,794,660</td>
<td>38.2</td>
</tr>
<tr>
<td>Mining</td>
<td>3,638,118</td>
<td>29.0</td>
</tr>
<tr>
<td>Wholesale</td>
<td>3,605,531</td>
<td>28.7</td>
</tr>
<tr>
<td>Management</td>
<td>250,262</td>
<td>2.0</td>
</tr>
<tr>
<td>Warehousing</td>
<td>187,219</td>
<td>1.5</td>
</tr>
<tr>
<td>Retail</td>
<td>55,743</td>
<td>0.4</td>
</tr>
<tr>
<td>Information</td>
<td>11,892</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12,543,425</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>


FAF integrates CFS data with a variety of supplemental sources for industry sectors and movements not covered in the CFS (out-scope to the CFS, and imports movements) to create a comprehensive picture of freight movement among states, major metropolitan areas, and remainder of states by all modes of transportation. FAF also provides forecasts on freight flows up to 30 years in the future as well as providing annual provisional updates for the current year, and truck flow assignments for the base year and outlying future year. Current FAF3 is the third database of its kind, with FAF1 providing similar freight data products for calendar year 1997, and FAF2 providing freight data products for calendar year 2002. The CTA was heavily involved in the development and execution of the methodology for FAF2 and FAF3. With data from the 2007 CFS and additional sources, for FAF3, CTA developed estimates for freight flows, by commodity type, mode, origin, and destination as well as ton-miles by mode.
In terms of the geographic dimension, the FAF3 data provides freight trading between 123 domestic zones and 8 external zones. The geographic level within the national boundaries is based on the CFS geographic strata, as shown in Figure 1, including 74 metropolitan areas, 33 remainder of states, and 16 regions identified as entire states. External zones, or foreign zones, are defined by 8 world regions: Canada, Mexico, and six other regions defined according to United Nations geographic region. Hence a domestic flow is spatially characterized by its origin and destination zones; imports are reported by foreign origin, FAF domestic zone of entry, and FAF destination zone; and exports are reported by FAF domestic origin, FAF domestic zone of exit, and foreign destination. The transportation modes considered are truck, rail, water, air, pipeline, and multiple modes. Multiple modes are combinations of truck-rail or truck-water modes. In terms of commodity classification, FAF reports freight flows using the same 43 2-digit Standard Classification of Transported Goods (SCTG) classes, as reported by the CFS.

![Figure 1 FAF geographic zones](image)

Figure 2 presents a summary of the mode composition of freight traffic based on FAF3 database. Clearly, truck carries most of the freight traffic in both tons and dollar value in the U.S. system. Although the mode composition highlights the importance of each mode for the transportation matrix, the system usage by each transportation mode should be represented in terms of ton-miles.

For FAF3, the CTA team provided estimations of mileage and ton-miles of freight flows using network system and freight flow models. The team manages detailed geographical data and information of a large multimodal and intermodal freight network, including highways, railroads, waterways, airways, and pipelines, and their associated infrastructure (e.g., intermodal terminals, transfer points, seaports and airports). A detailed description of the network system is presented by (9-10). This geographic system makes it possible to estimate freight movements between geographic zones.
FIGURE 2 Share of freight traffic by mode

3 TON-MILES ESTIMATION METHOD
The calculation of ton-miles considers the following steps: 1) estimation of total freight produced and attracted by county; 2) calculation of shipment distance between counties; and 3) distribution of the total freight produce/attracted between counties. The individual ton-miles are then calculated by simply multiplying tons and distances estimates.

The basic idea for disaggregating FAF freight flows is to estimate the relationship between freight generated and industry activity within zones, though freight modeling. As suggested by Oliveira Neto et al. (2011) we estimate aggregate generation (production and attraction) models based on data from CFS at the state level (11). The generation models also consider the association between industry producer and consumers of commodities, which can be observed in the 2007 Input-Output Accounts (12). The models are then applied to estimate freight shipped/received by county.

Likely shipment distances by mode between counties are estimated by using routing routines in the CTA network system. The routing algorithm uses impedance functions to route shipments onto network system. The impedance functions consider link functional category (e.g. major and local roads), the interactions between different companies (i.e. railroad companies) or vehicles (e.g. different vessels categories) as well as intermodal and multimodal operations in transfer points and terminals. A route for a single-mode is generated using a “shortest-path” algorithm that finds the minimum impedance between county centroids. If multiple modes or vehicles are used, the algorithm finds minimum routes connecting different sub-networks through terminal and transfer points. More details of the routing method can found at (9-10). County distance matrices are used not only to calculate ton-miles, but they are also applied to find the spatial distribution of freight generated. In this case, a gravity model is applied to generate freight flows between counties.

In sum, the disaggregation of FAF database includes:
1- *Estimate total freight (in thousand tons) shipped and received by county.* For a given FAF origin-destination (O-D) pair, production and attraction shares for tons are estimated for the set of counties within FAF zone O and D, respectively.
2- *Distribute the total freight shipped/received between counties.* The technique for freight distribution relies on matrix balancing procedures, or a doubly constraint gravity model.
3- *Calculate ton-miles between counties.* Ton-miles for a county-to-county pair is calculated by multiplying the flow in tons by the shipment distance.
In Section 3.1, the disaggregation model is formally described. The model uses disaggregation parameters to share the freight flows from FAF zone level down to county level. The technical details for estimating three disaggregation parameters (production, attraction and distribution coefficients) are presented in Section 3.2. These parameters are referred to disaggregation coefficients in the remainder of the paper.

As a part of the modeling exercise it is important to verify the validity of the models. The disaggregation methodology is validated though two procedures: a) disaggregation of 2007 CFS from state to FAF zone level, and b) disaggregation of 2007 CFS from state to county zone level. Section 4.1 presents the results of these two validation exercises. The results of FAF database disaggregation for movements by truck, rail and water are presented in Section 4.2.

3.1 Disaggregation model

This section presents a methodology for calculating disaggregated freight flows. Let \( I \) and \( J \) be sets of counties inside an origin FAF zone \( i \) and a destination FAF zone \( j \), respectively. Let \((r,s), r \in I \) and \( s \in J\), denotes a pair of origin-destination counties in which we wish to estimate the freight flow, \( x_{rspm} \), for a given aggregate FAF level of freight flow, \( x_{ijpm} \), between zones \((i,j)\). The flow \( x_{rspm} \) can be estimated according to

\[
x_{rspm} = x_{ijpm} Pr[rs/ijpm]
\]  

where \( Pr[rs/ijpm] \) is the probability of selecting origin-destination subzones \((r,s)\) for given flow between FAF zones \( i \) and \( j \), by industry \( p \) and mode \( m \).

A minimum information estimate, or a balance matrix procedure, can be applied to find the probability \( Pr[rs/ijpm] \). Such procedure results in the expression of Equation 2 below, which should be consistent with constraints 3-5.

\[
Pr[rs/ijpm] = g_{ijpm} h_{ijpm} w_{r/ijp} w_{s/ijp} z_{ijm} \tag{2}
\]

\[
\sum_{s \in J} Pr[rs/ijpm] = w_{r/ijp}, \text{ for all } r \in I \tag{3}
\]

\[
\sum_{r \in I} Pr[rs/ijpm] = w_{s/ijp}, \text{ for all } s \in J \tag{4}
\]

\[
\sum_{r \in I, s \in J} Pr[rs/ijpm] = 1 \tag{5}
\]

where \( g_{ijpm} \) and \( h_{ijpm} \) are multipliers that should be estimated to comply with constraints 3-5. \( z_{ijm} \) represents the average shipment mileage between counties \( r \) and \( s \) by mode \( m \). The \( w_{r/ijp} \) and \( w_{s/ijp} \) are production and attraction disaggregation factors, respectively. They represent our knowledge about the industry production and attraction shares by counties \( r \) and \( s \) respectively, for given FAF origin and destination \( i \) and \( j \). These two factors are estimated by

\[
w_{r/ijp} = y_{rp} \hat{\xi} / \sum_{s \in J} y_{rp} \hat{\xi} \tag{6}
\]

\[
w_{s/ijp} = y_{sp} \hat{\kappa} / \sum_{r \in I} y_{sp} \hat{\kappa} \tag{7}
\]
Variables $y_{rp}$ and $y_{sp}$ are the observed heterogeneity of the production and attractiveness of counties $r$ and $s$ by industry $p$. Parameters $\xi$, $\zeta$, and $\gamma_m$ correspond to coefficient of disaggregation for production, attraction and distribution of the total freight flow by FAF zone.

The procedure represented by Equation 1-5 is known in the literature as bi-proportional matrix balancing or iterative proportional fitting. This balancing method was apparently first described by Kruithof (1937) (13), who used the model for prediction of telephone traffic distribution. For more details see also (14). In transportation planning this method is referred to “Fratar Method” in the U.S. or “Furness Method” elsewhere (15).

3.2 Disaggregation coefficients

Regression analysis is widely used for modelling freight flows with objective of disaggregating the FAF database (16-18). Such modelling processes rely on the traditional approach for modelling transportation flows onto a study area. Therefore they estimate generation and distribution of freight flows among geographic zones. Given that disaggregated data is not available, we estimate the disaggregation coefficients using an aggregate 2007 CFS tabulation by U.S. state level, according to the following expression:

$$ x_{ipm} = a_p a_m y_{ip} \xi y_{jp} \zeta z_{ijm} \gamma_m g_i h_j $$

where $a_p$ is a multiplier representing the fixed effect of industry $p$; $a_m$ represents the scale of mode service $m$; $y_{ip}$ and $y_{jp}$ are the observed heterogeneity of the production and attractiveness of state $i$ and $j$ by industry $p$, respectively; $z_{ijm}$ represents the average shipment mileage; $g_i$ and $h_j$ are state fixed effects of the origin and destination regions, respectively. We are interested in estimating coefficient parameters $\xi$, $\zeta$, and $\gamma_m$. $\xi$ and $\zeta$, common to all industries, measures the contribution of production and attraction variables to the freight flow. $\gamma_m$ measures the contribution of the shipment distances in estimating the freight flow by mode $m$.

As discussed before disaggregated data at county level is not available. However the specification in Equation 8 considers the heterogeneity of industry activity and all kinds of determinants associated with origin and destination zones. Therefore, it is assumed that the same specification is valid for coarse levels of geographic detail, which will be useful for the data disaggregation.

Two methods are deployed to obtain the disaggregation coefficients. The first one is based on traditional two step modeling, which aims to estimate descriptive freight generation models and distribution models, separately. In the first step $\xi$ and $\zeta$ are estimated from the production and attraction equations, reading

$$ \ln x_{ip} = \lambda_i + \gamma_p + \xi \ln y_{ip} + \epsilon_{ip}, $$

$$ \ln x_{ip} = \lambda_j + \theta_p + \zeta \ln y_{jp} + \epsilon_{jp}, $$

where $\gamma_p$ and $\theta_p$ are the natural logarithm of the fixed effect parameters of industry $p$; $\lambda_i$ and $\lambda_j$ are the natural logarithm of the U.S. state fixed effects; $\epsilon_{ip}$ and $\epsilon_{jp}$ represent unobserved disturbance factors.

In the second step $\gamma_m$ is then estimated for each mode service. To this end, doubly constraint gravity models with power deterrence functions were calibrated by the maximum likelihood method.
using the iterative procedure due to Hyman (1969) (19). For a given mode \( m \), such approach corresponds to find the minimum information estimate, denoted by \( x_{ij/m} \), under macro constraints on the total originated and attracted flows and additional constraint on transportation costs. This is equivalent to find the Lagrange multipliers of the following entropy-maximization problem (more details can be found at (20-21)):

\[
\text{Maximize } E_m = -\sum_{ij}(x_{ij/m} \ln x_{ij/m} - x_{ij/m}) \tag{11}
\]

\[
\sum_i x_{ij/m} = x_{i/m}, \text{ for all } i \tag{12}
\]

\[
\sum_j x_{ij/m} = x_{j/m}, \text{ for all } j \tag{13}
\]

\[
\sum_i \sum_j x_{ij/m} \ln(z_{ij/m}) = Z_{m}, \text{ for all } i \text{ and } j \tag{14}
\]

where the objective function in Equation 11 is a monotonic function, often referred to as entropy function; Equations 12-13 are constraints representing our knowledge about the total productions and attractions by state for a given mode \( m \); Equation 14 is a constraint corresponding to our knowledge about the total expense, denoted by \( Z_{pm} \), in using the mode system \( m \) by industry sector \( p \). The average distances \( z_{ij} \) at the state level are obtained from weighted average distance in miles of the set of paths on the CTA network system between all corresponding pair of contiguous counties. The distances are weighted by the county populations. The notation \( ./m \) is used to emphasize that that the models are estimated separated for each mode.

It can be demonstrated that the Lagrange multiplier of Equation 8 is actually the absolute value of the disaggregation coefficient \( \gamma_m \). Thus the solution of problem 11-14 is given by

\[
x_{ij/pm} = g_{i/m} h_{j/m} x_{i/m} x_{j/m} z_{ij/m} \tag{15}
\]

where \( g_{i/m}, h_{j/m} \) are the exponential of the Lagrange multipliers of constraints 12-13.

It is worth noting that the result in Equation 15 is equivalent to estimate the following expression:

\[
x_{ij/pm} = x_m \ Pr[jj/m] \tag{16}
\]

where \( x_m \) is the total freight flow by mode \( m \) and \( Pr[jj/m] \) is an estimated probability of choosing the pair of states \( (i, j) \) given that the mode selected was \( m \). \( Pr[jj/m] \) corresponds to the response of a multinomial logit model, as demonstrated by Anas (1983) at (21).

In spite of the similarity between Equations 2-5 and 12-15, they are actually applied with different objectives. Equation set 2-5 is applied to calculate the disaggregation factors for given disaggregation coefficients, production and attraction heterogeneity, and shipment distances. Whereas the equation set 12-15 does the inverse. For a given freight flow intensity, the system of Equations12-15 delivers an estimate for disaggregation coefficient \( \gamma_m \). In sum Equations 2-5 seeks to find a response for a freight intensity between states given exogenous variables and coefficient parameters, whereas Equations 12-15 seeks to estimate a model parameter for a given movement pattern between states.

The second approach to estimate the parameters of Equation 8 was to combine the data into a pooled cross-section structure, from where the following equation was estimated. We call this approach the one-step estimation.
\[ \ln x_{ijpm} = \lambda_p + \lambda_m + \lambda_i + \gamma_j \ln y_{jp} + \zeta \ln y_{jp} + \gamma_m \ln z_{ijm} + \epsilon_{ijpm}, \]  

(17)

where the parameters denoted by \( \lambda \)'s correspond to the natural logarithm of the industry fixed effect and mode scale parameters, \( a_p \) and \( a_m \), as well as the state fixed effect parameters, \( g_i \) and \( h_j \).

4 RESULTS

The following data is used in the modeling process. The freight flow pattern is based on data from a 2007 CFS tabulation that contains domestic movements between U.S. states by mode of transportation (classified as truck, rail, water, air, multiple, pipeline, and unknown), and by seven major industry sectors, as shown in Table 1. As proposed by Oliveira Neto et al. (2011) (11) the economy activity by industry sector is captured from the 2007 County Business Pattern (CBP), which contains employment information for industry classified according to the NAICS industry code. More details about CBP can be found at (22). The production variables correspond to the total annual payroll by industry, while measures of attractiveness are related to those industries that use commodities produced by the originated industries in 2007 CFS. The association between producers and consumers, which is necessary for estimating measures of attractiveness, is obtained from the Input-Output Accounts (12).

Table 2 shows the disaggregation coefficients obtained with the two estimation methods described above. The robust standard errors are presented in parenthesis for the statistical models. The goodness-of-fit statistics, \( R^2 \), of the production and attraction models of Equation 9 and 10 were 0.89 and 0.94, respectively. The estimated model of Equation 17 presented \( R^2 \) of 0.58.

<table>
<thead>
<tr>
<th>Variables</th>
<th>2-step estimation</th>
<th>1-step estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{ii} )</td>
<td>0.531 (0.073)**</td>
<td>0.584 (0.023)**</td>
</tr>
<tr>
<td>( y_{ji} )</td>
<td>0.712 (0.213)*</td>
<td>0.563 (0.099)**</td>
</tr>
<tr>
<td>( z_{ij_truck} )</td>
<td>-2.941</td>
<td>-1.932 (0.022)**</td>
</tr>
<tr>
<td>( z_{ij_rail} )</td>
<td>-2.012</td>
<td>-1.268 (0.049)**</td>
</tr>
<tr>
<td>( z_{ij_water} )</td>
<td>-1.312</td>
<td>-0.601 (0.084)**</td>
</tr>
<tr>
<td>( z_{ij_air} )</td>
<td>-0.866</td>
<td>-0.316 (0.071)**</td>
</tr>
<tr>
<td>( z_{ij_multiple} )</td>
<td>-1.488</td>
<td>-0.936 (0.039)**</td>
</tr>
<tr>
<td>( z_{ij_pipeline} )</td>
<td>-4.065</td>
<td>-1.676 (0.136)*</td>
</tr>
<tr>
<td>( z_{ij_unknown} )</td>
<td>-3.486</td>
<td>-1.494 (0.048)**</td>
</tr>
</tbody>
</table>

** Significant at 5%
* Significant at 1%

4.1 Validation

Two procedures are undertaken to validate the disaggregation methodology. They are presented in the following sections.
a) 2007 CFS disaggregation from State to FAF zone level
The disaggregation method is used to share the 2007 CFS flows from state to FAF zone level. FAF zones corresponding to entire states are left out from this analysis. Two matrices were obtained by applying the two sets of disaggregation coefficients presented in Table 2. We compare the freight flow estimates with known freight flows at FAF zone level provided by the CFS.

The two graphics in Figure 3 shows scatter plots of observed \( n_{ij} \) and estimated \( x_{ij} \) flows, in thousand tons, between FAF zones \((i,j)\). Statistics for comparing the set of estimated flows and the set of observed flows are also provided in Figure 3. The first measure is the \( R^2 \) statistics which measures the percentage of observed variation that can be explained by the models. The second one, expressed by Equation 18, is the standardized root mean square (s.r.m.s), which was proposed by Pitfield (1978) as an alternative to deal with sparse matrices and also to consider the scale of the variables involved (23). However, it does not have any statistical property. It is merely a descriptive measure of goodness-of-fit.

\[
s.r.m.s = \left[ \frac{\sum_{ij} (x_{ij} - n_{ij})^2}{q} \right]^{1/2} / \sum_{ij} n_{ij} / q
\]  

(18)

where \( q \) is the number of cells in the estimated set \( \{x_{ij}\} \).

Although such validation might not be sufficient to verify the disaggregation method, a negative result here would probably invalidate the method. The results indicate that the disaggregation procedure can provide a good fit to the observed data for both methods of estimating the disaggregation coefficients. The 2-step estimation method achieved a slightly better performance.

![FIGURE 3 Goodness-of-fit estimates of the disaggregation method](image)

b) CFS Disaggregation from State to County Level
The disaggregation method is used to share the 2007 CFS flows from state to county level. The estimates of county-to-county freight flows are used to generate distribution of total tons by shipment distance range. These distributions are compared against similar distributions provided by the CFS sample.
The charts in Figure 4 show a comparison of the distributions obtained from disaggregation methodology with the 2007 CFS estimates. The 2-step estimation is more consistent with the distribution reported in 2007 CFS. Similar results are also observed when comparing distributions for movements by rail, water, and multiple modes.

FIGURE 4  Tonnage distribution by distance on the U.S. highway system

4.3 FAF3 ton-miles

The disaggregation coefficients obtained by the 2-step estimation are used to estimate ton-miles of the FAF3 database considering both domestic and foreign trade. The following steps summarize how FAF3 database is disaggregated:

1- FAF database is organized by industry sector and mode; thus flows classified by 2-digit SCTG are grouped to the corresponding industry groups, at 3-digit NAICS.

2- For each FAF origin-destination flow (classified by NAICS and mode), total freight shipped/received by county within the given FAF origin/destination zones are estimated.

3- A matrix balancing procedure (see Equations 2-5) is then applied to distribute total freight shipped and received between counties.

Table 3 shows comparisons of the ton-miles estimates based on FAF3 and other data sources over the major U.S. freight transportation systems: highway, railway, and waterway. In addition to the 2007 CFS, other sources of information railway and waterway flows include: a) 2007 U.S. freight railroad statistics from Association of American Railroads (AAR) (24), and b) 2007 waterborne commerce of the U.S. Army Corp of Engineers (USACE) Navigation Data Center (25). Note that the 2007 U.S. freight railroad statistics reports all freight activities for all North American railroad companies using the U.S. system. The 2007 waterborne commerce data was based on a complete census of all cargo via the U.S. waterway system (except military cargo carried on military ships). Besides CFS, there is no other current source for comparison of highway movements, however.

As expected all FAF3 estimates are larger than the corresponding CFS estimates due to the inclusion of freight activities from out-scope shipments and import trade movements in the FAF3 database. Both FAF3 estimates of ton-miles for rail and water are consistent with estimates reported
by modal data programs, AAR and USACE, respectively; with less than 10% difference in waterway movements and about 5% for railway movements. The discrepancy observed for railway movements is likely due to a couple of reasons. The FAF3 database does not account for transshipments from Canada (e.g. Canada-Mexico). Furthermore, AAR report includes some movements of empty containers and the weight of containers for mixed freight, which are not considered in the FAF3 database. With these two considerations in mind, the gap between FAF3 and the AAR could be considerably reduced.

<table>
<thead>
<tr>
<th>U.S. Network Sub-system</th>
<th>Data source / Modes</th>
<th>Ton-miles (billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highway</td>
<td>FAF3 (Truck single mode only)</td>
<td>2,473</td>
</tr>
<tr>
<td></td>
<td>2007 CFS (Truck single mode only)</td>
<td>1,342</td>
</tr>
<tr>
<td>Railway</td>
<td>FAF3 (Rail single mode plus rail portion of multiple modes)</td>
<td>1,726</td>
</tr>
<tr>
<td></td>
<td>2007 CFS (Rail single mode and portion of multiple modes which includes rail)</td>
<td>1,530</td>
</tr>
<tr>
<td></td>
<td>2007 AAR report (all rail activities)</td>
<td>1,820</td>
</tr>
<tr>
<td>Waterway</td>
<td>FAF3 (water and the water portion of multiple modes)</td>
<td>554</td>
</tr>
<tr>
<td></td>
<td>2007 CFS (water and the portion of multiple modes which includes water)</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>2007 USACE waterborne commerce (all water activities)</td>
<td>506</td>
</tr>
</tbody>
</table>

5 CONCLUDING REMARKS

In this paper a methodology is proposed to disaggregate the FAF3 database down to the county level of geographic detail. The rationality for that is to carry out national transportation freight analysis and planning at reasonable level of detail. One of the goals of this disaggregation modeling effort is to estimate ton-miles of the U.S. goods movements by truck, rail, water, air and multiple modes. Freight models are estimated and validated with sample data from the 2007 CFS at different levels of aggregation. The validation results suggest that the traditional modeling techniques provide consistent results for macro analysis in transportation planning.

The framework for national freight data (i.e. FAF database) as well as the analytical and geographic tools (CTA’s network system and the freight flow models) presented in this paper can be used for estimating the generation and spatial distribution of goods in U.S. at the county level of geography. The estimation of freight ton-miles is an import application of such tools to gauge the freight system usage and provide insightful information for national political decisions. Measure of freight activity such as ton-miles estimation, are useful inputs for impact analysis at macro level such as:

- **Energy use and environmental impacts of the freight system**: The freight movement causes adverse impacts on the environment, mainly in the form of greenhouse gases. The models presented herein can help evaluate alternatives (e.g. modal shifting, vehicle technologies, etc.) for alleviating some of these negative effects;
Effects of external events on the freight movements: Certain natural and manmade events can disrupt and impede freight movement. Natural disasters such as flooding and hurricane (e.g. hurricane Katrina, Irene) can damage and cause disruptions to the system. Man-made disasters can be another source of threat to the transportation system. The hazardous natures of certain commodities can also pose risk to the freight network system (e.g. chemical spills). Freight demand and network supply models can help identify the system vulnerability and the economic impact of such threats on the transportation system.

To further improve the disaggregation method, it is recommended an investigation of methods capable of incorporating other information onto network system. With respect to highway activities, information of truck flows on highway links, weight stations, and from remote systems can be incorporated to further enhance the methodology proposed in this paper. County-to-county flows can be adjusted to meet additional constraints such as total link flows through the network system.

REFERENCES