A goal-programming approach to allocate Freight Analysis Framework mode flow data

Daniel Rodriguez-Roman
Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California
Irvine, CA, USA, 92697
d.rodriguez@uci.edu
(Corresponding author)

Neda Masoud
Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California
Irvine, CA, USA, 92697
nmasoud@uci.edu

Kyungsoo Jeong
Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California
Irvine, CA, USA, 92697
kyungsoj@uci.edu

Stephen G. Ritchie
Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California
Irvine, CA, USA, 92697
sritchie@uci.edu

Word Count: 5,982
Tables and Figures: 3 Tables + 3 Figures = 1,500
Total Word Count: 7,477

ABSTRACT
Several methods have been proposed to disaggregate Freight Analysis Framework (FAF) commodity flows to zonal structures of greater geographical detail. This disaggregation is usually performed on the basis of explanatory variables related to the supply and demand of goods. In this paper a complementary procedure is presented to determine the mode splits of disaggregated FAF flows. A goal-programming approach is proposed to allocate FAF mode flow data on the basis of mode-related variables. The formulated goal-programming problem minimizes the deviation between the mode flow decision variables and target mode flow values, subject to given FAF mode flow information. The use of mode split models is proposed to define the problem's target values. In a sample application of the procedure a method to estimate aggregate mode split models with FAF data is discussed. These mode split models could be used by transportation organizations that do not have access to freight mode choice models to define the goal-programming problem's target mode flow values. Additionally, an optimization problem is formulated to account for FAF mode flow data in the disaggregation of total commodity flows. Lastly, validation procedures for FAF disaggregation and mode allocation results are discussed, and an example of a validation approach is presented.

Keywords: Freight Analysis Framework, disaggregation procedures, goal-programming problems, aggregate mode split models

INTRODUCTION

Commodity-based freight transportation models attempt to capture the relationship between the supply and demand of goods by different economic sectors and the resulting generation of freight movements. Given its focus on the driving forces behind the demand for freight transportation, this modeling approach permits the analysis of sophisticated scenarios and policy questions of interest to metropolitan and state transportation organizations. Unfortunately, many transportation agencies do not have access to the commodity flow data necessary to develop these types of models, either because the data have not been collected, its purchase from third parties is prohibitively expensive, or the available data do not exist at the necessary level of geographic detail. In response to this data availability problem, methods have been developed to derive the commodity flows of interest from public data sources. The FHWA's Freight Analysis Framework regional database (FAF) is one of such public data sources available to transportation analysts in the US *(1)*.

The latest version of the Freight Analysis Framework, FAF^3, contains commodity flow data – in term of annual tons, ton-miles, and dollar value – on 43 different types of commodity groups. The FAF^3 database includes commodity flow information for 123 domestic and eight international origin-destination (OD) regions, the latter being used to account for the export and import of commodities. The database also provides commodity flow information by mode of transportation categories, namely, truck, rail, water, air (including air-truck), multiple modes and mail, pipeline, other and unknown, and the "no domestic mode" category used for import flows *(1)*. This comprehensive database has the potential of addressing some of the data gaps facing many transportation agencies, and thus several methods have been developed to disaggregate FAF commodity flows to zonal structures with greater geographical detail.

The methodology presented in this paper can be regarded as a second-stage procedure following the distribution of FAF flows to sub-FAF zonal levels. As will be discussed in the next section, the disaggregation of FAF flows is generally performed by computing or estimating expansion factors that are functions of variables related to the supply and demand of goods. In this paper it is proposed that the disaggregated commodity flows can then be split into mode-specific flows by using variables related to freight mode choice. A goal programming-based procedure is discussed to determine the mode splits of disaggregated FAF commodity flows. The objective of the goal-programming problem is to minimize the difference between sub-FAF mode flows and associated target mode flow values subject to sets of constraints that ensure consistency with the FAF mode flow data. Mode split models and analyst assumptions can be used to define target flows for each mode.

The next section reviews previous efforts to disaggregate the FAF database. The third section presents the general outline of the procedure to determine the disaggregated flows mode splits. This is followed by a case study application of the proposed methodology in which a method to estimate aggregate mode split models using FAF^3 data is discussed. The fourth section integrates previous disaggregates approaches with the goal-programming model. The last section discusses possible future research and applications of the proposed models.

LITERATURE REVIEW

One of FAF's data products comprise sets of truck flows assigned to selected highway links. These estimated truck flows are the result of a procedure that disaggregates the FAF level OD flows to corresponding freight activity centers at the county or sub-county level. These activity centers include major manufacturing plants, truck-rail intermodal terminals, seaports, airports, and other truck-generating urban clusters. For each center a measure related to the production or consumption of freight is computed, and its division by the corresponding FAF total activity constitutes the share of the total flow allocated to the center. In FAF's second version (FAF^2), the freight activity measure for domestic flows was a function of the number of industrial establishments associated with each center (as reported in the 2005 County Business Patterns) and related annual vehicle-miles travelled (VMT, obtained from the 2002 Highway Performance Monitoring System). A similar procedure was followed to disaggregate international truck freight flows, but the disaggregation was performed considering truck flow data at international crossings *(2)*.

Viswanathan *et al. (3)* note that employing VMT in the computation of disaggregation factors has several drawbacks, including the fact that truck VMT measures capture non-freight related truck traffic (e.g., utility and service truck VMT), truck traffic that is merely passing by the activity center but not originating from or arriving at the center, or both. Therefore, Viswanathan et al. proposed creating commodity-specific disaggregation factors using the freight flows estimates of freight production and attraction models, akin to generation models in four-step commodity-based freight models. The commodity-specific production and attraction models were specified as functions of different types of employment and, for selected commodities, population. In contrast to the previous methodology, this methodology considers total OD flows, rather than the truck OD flows. Viswanathan et al. estimated the models using FAF domestic data as the dependent variable. The equations were then applied at the county and FAF zonal levels to estimate freight production and attraction. These estimates were used to compute proportional production and attraction weighting factors that map the total FAF OD flow to their corresponding sub-FAF OD zones. Opie *et al. (4)* also applied proportional weighting methods to disaggregate FAF data to county level zones, but in their study the production and attraction factors were computed by directly dividing the county level proxy variable (e.g., employment, population, VMT) by the related FAF level proxy variable. In contrast, in the methodology proposed by Ruan and Lin *(5)* the disaggregation factors that map FAF OD flows to sub-FAF OD flows were jointly estimated with the freight production and attraction equations.

As would be expected, a common theme in the reviewed studies is that the zones with higher values of the identified proxy production or consumption variables are allocated greater portions of the FAF OD flows. In the studies by Viswanathan *et al. (3)* and Ruan and Lin *(5)* this relationship was formalized with the estimation of freight production and consumption equations. In the next section, a procedure that splits the disaggregated FAF commodity flows into mode flows based on the given FAF mode data is presented. And, analogous to the use of freight production and attraction models used in the previously discussed disaggregation procedures, the proposed methodology allocates the modal flows according to mode split model estimates.

MODE FLOW ALLOCATION PROCEDURE

Consider the FAF level commodity flow F_{IJc} for origin I, destination J, and commodity group c, and its disaggregated flows f_{ijc}, where $i \in I$ and $j \in J$. The objective is to determine reasonable mode flow splits for each f_{ijc} that are in agreement with the FAF modal information. Let m_{ijck} represent the flow by mode k between zones i and j, and M_{IJck} represent the given FAF level mode flow between zones I and J. For an OD, a mode flow allocation is in agreement with the FAF level information if it satisfies the following constraints:

$$\sum_i \sum_j \alpha_{ijck} m_{ijck} = M_{IJck} \qquad \forall\, k \qquad (1)$$

$$\sum_k m_{ijck} = f_{ijc} \qquad \forall\, i, j \qquad (2)$$

$$m_{ijck} \geq 0 \qquad \forall\, i, j, k \qquad (3)$$

The first constraint ensures that, for each mode k, the sum of all sub-FAF m_{ijck} flows equals the parent M_{IJck} flow. α_{ijck} is a coefficient that indicates if there is access to mode k between zones i to j, an important consideration for the rail, water and pipeline modes. The second constraint guarantees that the proposed mode flows m_{ijck} add to the total commodity flow f_{ijck}. Given this set of constraints, a reasonable allocation of the FAF mode flows can be determined by solving the following goal-programming problem:

$$\text{Minimize } X = \sum_i \sum_j \sum_k \left| m_{ijck} - \widehat{m}_{ijck} \right| \tag{4}$$

$$\textit{subject to constraints } (1) - (3)$$

The values $\widehat{m}_{ijck}$ are target (or goal) mode split values. Therefore, the solution to the goal programing problem are the mode flows m_{ijck} that result in the least absolute deviation from the target values. Defining the target mode splits for each OD ij is a subjective task that depends, in part, on the analyst's judgment. One approach to objectively define the target $\widehat{m}_{ijck}$ values is to multiply the f_{ijc} flows by mode shares estimated via a mode split model. However, given that it is unlikely for a mode split model to consider all FAF modes (e.g., consider the other and unknown mode category), complementary assumptions might be required for ODs with modes out of the mode split model scope.

Note that the objective function can be linearized by introducing variables that measure the positive (PD_{ijck}) or negative (ND_{ijck}) deviations between each m_{ijck} and $\widehat{m}_{ijck}$. The problem is then reformulated as:

$$\text{Minimize } Y = \sum_i \sum_j \sum_k \left(PD_{ijck} + ND_{ijck} \right) \tag{5}$$

$$\textit{subject to constraints } (1) - (3) \textit{ and}$$

$$PD_{ijck} - ND_{ijck} = m_{ijck} - \widehat{m}_{ijck} \qquad \forall\, i, j, k \tag{6}$$

$$PD_{ijck} \geq 0, ND_{ijck} \geq 0 \qquad \forall\, i, j, k \tag{7}$$

The presented objective function could easily be substituted by other formulations, such as a least squares objective function. However, considering that the goal-programming model must be solved for each FAF OD being disaggregated and for each commodity group of interest, a linear objective function is presented since it is one of the least computationally demanding formulations of the problem. Besides the selection of objective function, it should also be noted that another possible modification is formulating the problem in terms of mode shares (instead of mode flows).

An alternative methodology to the one presented in this section is to perform the FAF flow disaggregation for each mode's commodity flow OD matrix separately, instead of disaggregating the total commodity flow OD matrix and then allocating the flows to mode groups (the approach assumed in this study). The latter approach can be considered as conceptually more appealing as it extends the argument that data disaggregation procedures should be based on variables related to what is being disaggregated. In the case of disaggregating procedures for commodity flows, the variables of interest are those related to the demand and supply of goods. Analogously, variables associated with freight mode choice, such as modal service attributes and decision-maker characteristics, should be used for the mode allocation of disaggregated flows. Moreover, although disaggregating each mode's commodity flow OD matrix separately has the benefit of circumventing the need for a commodity flow mode allocation procedure, the mode-specific disaggregation factors generated in this approach may undercut the rationale for using production and consumption variables to compute FAF disaggregation factors. This is because these mode-specific factors would generally rely on the same zonal demand and supply variables (e.g., population, employment, farmland acres). Since each mode transports a different amount of goods, mode-specific disaggregation factors could imply that for each mode the explanatory variables result in different rates of consumption and production, which might not be reasonable. This observation is particularly meaningful when production and attraction models are estimated to develop disaggregation factors *(e.g., see 3, 4)*. As an example, consider an analyst interested in modeling the zonal production and attraction (or consumption) of agricultural shipments using farmland acreage as a production variable and population as an attraction variable. A zone's population has a single demand level for agricultural products, not distinct demand levels by mode, as the consumption decisions of each person do not account for which modes were used to transport the products. Similarly, the farmland production levels at each zone do not vary by mode, but, in part, as a function of the aggregate demand. Therefore, if the analyst is interested in computing disaggregation factors that reflect the supply and demand dynamics associated with agricultural product markets, he/she should use total production and consumption of commodity flows instead of introducing data segmentations that, although convenient, may mask the causal connections between the flows and the selected explanatory variables.

The next section presents an application of the proposed procedure. In addition, a method to estimate mode split models using FAF data is discussed, which might be useful for transportation agencies that have not developed or do not have access to freight mode split model.

SAMPLE APPLICATION OF THE PROPOSED PROCEDURE

The domestic commodity flow of manufactured metal products between the Los Angeles FAF zone and the Houston FAF zone was selected for the application of the allocation procedure. The definition of the manufactured metal products commodity group is taken from Ranaiefar *et al. (6).* This commodity group is composed of FAF^3's base metal, articles-based metals, and machinery groups. The FAF^3 database for domestic movements reports 324.6 ktons flow of manufactured metal products from the Los Angeles FAF zone to the Houston FAF zone, of which 73.4 percent was transported by truck, 18.4 percent was transported by rail, 7.5 percent was transported by multiple modes, and 0.7 percent was transported by air-truck. The disaggregated zonal structure for California's FAF zones presented by Ranaiefar *et al.* was used in this study. Figure 1 shows the disaggregated Los Angeles zone; the Houston FAF zone is not disaggregated as it is regarded as an external zone. The rail stations shown in Figure 1 represent those stations that are known to handle manufactured metal products flows according to the 2007 Surface Transportation Board Carload Waybill Sample (CW) *(7).* Additionally, Figure 1 presents the truck-rail intermodal facilities in the area.

Disaggregating commodity flows

The methodology proposed by Viswanathan *et al. (3)* was implemented to disaggregate the 324.6 ktons flow. This methodology was previously used by Cambridge Systematics to disaggregate California's FAF^2 domestic flows *(8)*. The freight production model estimated by Ranaiefar *et al. (6)* with FAF^3 data was used for the disaggregation. This model is a function of the number of establishments in the fabricated metal product manufacturing industry (industry 332 in the North American Industry Classification System) and the manufacturing sector's gross domestic product. Given that the destination zone is not disaggregated, the equation used to obtain the sub-FAF level flows using the proportional weighting method is:

$$f_{ijc} = F_{IJc} \times \frac{p_{ic}}{p_{Ic}} \tag{11}$$

For the OD under consideration, F_{IJc} is the FAF level flow of 324.6 ktons. p_{Ic} and p_{ic} are the total generation of manufactured metal products by the Los Angeles zone and the related sub-FAF origins, respectively, as predicted by the production models. Table 1 presents the computed f_{ijc} flows for the example under consideration.

Allocating mode flows

As mentioned in the previous section, the targets of the goal-programming problem can be determined using mode flow estimates from mode split models. However, the same data availability challenges encountered with commodity flow data are encountered with freight mode choice data. Therefore, the estimation of an aggregate mode split model using FAF[3] data is discussed in this subsection. This procedure might be useful for transportation agencies that do not have access to a suitable mode choice model. Alternatively, agencies in this situation could also consider borrowing the parameters from another organization's freight mode split model, and attempt to update the parameters based on available mode data (e.g., mode shares from FAF).

The multivariate fractional regression (MFR) model (or fractional split model) was selected to estimate an aggregate mode split model for the manufactured metal products commodity group. The MFR model structure has been used in transportation studies to examine commodity flow distribution *(9)*, time-use allocation *(10)*, highways' VMT mix *(11)*, and to estimate binomial freight mode split models *(12)*. Model parameters in the MFR model are estimated using a quasi-likelihood estimation approach. The MFR model is used to estimate the expected values for fractional dependent variables. It is especially useful when a non-negligible number of observed fractions take boundary values of zero or one.

In the current application the MFR model was estimated using California-related mode shares as the dependent variables. These mode shares were computed using FAF mode flows that had an origin or destination in California. Similar to the approach taken by Viswanathan *et al. (3)* and Ranaiefar *et al. (6)*, the model was estimated using FAF level data, but applied at the sub-FAF geographical scale. Only three modes were included in the model: truck, rail, and multiple modes and mail. The multiple modes and mail category was incorporated in the model because of the importance of the truck-rail mode in mode shift analyses. Unfortunately, as its name indicates, in addition to truck-rail the multiple modes and mail category includes modes such as truck-water, rail-water, and, most importantly, mail and parcel. Therefore, assumptions are required to remove the non-truck-rail component of FAF's multiple modes and category. For simplicity, in this study the 2007 Commodity Flow Survey (CFS) *(13)* was used to obtain commodity and OD specific factors to subtract the mail and parcel flows from FAF's multiple modes and mail category, and the remaining flow was modeled as if it was truck-rail flow. The CFS information is presented for state level ODs, so the factors are state OD specific. For example, the CFS data for the manufactured metal products flows between California and Texas shows that mail and parcel represents 2 percent of the total flow. Given the stated assumptions, this implies that 6.6 ktons of the Los Angeles-Houston flow is transported by the mail and parcel mode, while the remaining 17.6 ktons in the multiple modes category is modeled on the basis of truck-rail modal attributes.

A multinomial logit formulation was assumed for the MFR model as this functional form is often used in aggregate freight mode split modeling *(14)*. Therefore, the target mode flows $\widehat{m}_{ijck}$ were computed as follows:

$$\widehat{m}_{ijck} = f_{ijc} \times \frac{e^{\gamma_{ijck}}}{\sum_n e^{\gamma_{ijcn}}} \tag{12}$$

γ_{ijck} was specified as:

$$\gamma_{ijck} = \beta_{0k} + \beta_{cost} \log\left(cost_{ijck}\right) + \beta_{time} \log\left(time_{ijck}\right) \tag{13}$$

β_{0k} is the mode k specific constant, β_{cost} is the coefficient of the cost attribute, and β_{time} is the coefficient of the transit time attribute. This specification attempts to capture the effects of cost and transit time on freight mode shares, a standard practice in aggregate mode split models. Demo-economic variables *(12)*, measures of commodity values and the magnitude of the mode flows themselves are examples of other explanatory variables that could be considered in the specification of γ_{ijck}.

The cost variable ($cost_{ijck}$) was computed by multiplying mode k's shortest path distance between OD ij by that mode's shipping freight rate ($/ton-mile). The equation for the rail and truck-rail modes was estimated using the California-related carload data in the CW. This equation is a function of distance, a dummy variable for intermodal shipments, location dummies (i.e., dummies that indicate from which state the shipment originated or arrived), and an interaction term between distance and the intermodal dummy.

The truck freight rate equation was estimated using data from a commercial freight rates database *(15)*, and is a function of distance, average fuel price in the region of origin *(16)*, and, again, location dummies. An additional freight rates model was estimated for short-haul truck movements using as explanatory variables fuel, dummies for origin zones, and lag rate variables of the second order. This short-haul truck rate equation was used to quantify the drayage cost associated with truck-rail cost; the drayage cost was added to the truck-rail rail-haul cost. The three freight rate equations were estimated using stepwise linear regression. Additional assumptions related to truck-rail terminal costs were made based on information presented by Resor and Blaze *(17)*. The rate functions are not included in this paper due to space limitations, but are available upon request. Distanced-based functions were used to compute transit times for the truck, rail, and truck-rail modes. These functions are also available upon request.

Table 2 presents the estimated parameters for the aggregate mode split mode. Truck-rail is selected as the reference so its alternative specific constant is set to zero. Table 1 shows the mode split model target values for the truck, rail and truck-rail modes, as well at the uniformly distributed target flows for the remaining modes. The targets for the remaining mode flows (the air and the subtracted mail and parcel mode flows) are the total f_{ijc} flows multiplied by their corresponding aggregate mode share. Table 1 also presents two solutions to the optimization problem: the solution for the least absolute deviations objective function (L_1) and the solution for the least squares objective function (L_2). By design, both solutions add to the FAF^3 mode flows, so the analyst must judge which objective function result is more reasonable, given available data and constraints.

Ortuzar and Willumsen note that aggregate mode split model results "may turn out to be very approximate" *(14)*. Not surprisingly, the same could be observed with models estimated using the FAF^3 mode flow data, particularly given the size of FAF zones and the related levels of data aggregation. In the presented example, the last row in Table 1 shows that the target mode flows underestimate the FAF rail flow for the selected OD pair while overestimating the truck and truck-rail flows. In general, a reason for the relative inaccuracy of aggregate mode split models is that zonal level average cost and time measures fail to capture the influence of several important mode choice determinants, such as perceptions of mode travel time reliability, mode accessibility, and shipment characteristics. The development of additional proxy variables to account for these factors is a possible way to improve model accuracy.

ENSURING CONSISTENCY BETWEEN DISAGGREGATED FLOWS AND FAF MODE DATA

For some ODs the existence of feasible solutions to the goal-programming problems will depend on whether the FAF mode information was taken into account when disaggregating the total commodity flows. If the flows are disaggregated solely on the basis of the information provided by surrogate freight generation variables, the resulting flows may conflict with the FAF mode data. As an example, consider the FAF^3 animal feed commodity flow from California's "remainder zone" to Hawaii's "remainder zone". Activities associated with the production of animal feed (e.g., the harvest of hay) occur throughout all counties of California's "remainder zone", so a disaggregation based on this commodity's freight production variables would generate flows originating from all the counties in this zone. However, the FAF modal information for this OD indicates that its animal feed flow is transported exclusively by the water mode (which does not include truck-water flows). Thus, the goal-programming problem for this OD would be infeasible as not all counties in California's "remainder zone" have seaports, and consequently constraint set (1) could not be satisfied. Similar incongruities could arise for other ODs with rail, water or pipeline flows, as the facilities required to access these modes are not as ubiquitous as roadways. Therefore, a procedure is required to ensure that the total commodity flow disaggregation results, f_{ijc}, do not conflict with the FAF mode data. Define R_{IJ} as the set of all modes with non-zero flow from I to J that have mode availability restrictions (i.e., modes that are not available in all ij pairs), and r as a mode in that set. Furthermore, let α_{ijcR} refer to an indicator variable that assumes the value of one when a mode in R is available in OD ij. Then, a set of disaggregated flows that are consistent with the FAF mode data can be determined by solving the following problem:

$$\text{Minimize } Z = \sum_i \sum_j \left| f_{ijc} - \hat{f}_{ijc} \right| \tag{14}$$

$$subject\ to$$

$$\sum_i \sum_j \alpha_{ijck} f_{ijc} \geq M_{IJck} \qquad \forall\ k \tag{15}$$

$$\sum_i \sum_j \alpha_{ijcR} f_{ijc} \geq \sum_r M_{IJcr} \tag{16}$$

$$f_{ijc} \geq 0 \qquad \forall\ i,j \tag{17}$$

The target values $\hat{f}_{ijc}$ can be obtained via one of the disaggregation procedures reviewed. The α_{ijck} indicators in constraint set (15) have the same interpretation as in constraint set (1). Constraints (15) ensure that, for each mode k, the sum of the allocated flows to zones with access to mode k is greater than or equal to the corresponding FAF mode flow. However, since it is possible for modes in R to be present in the same zones, the constraints (15) might allow for solutions in which the combined flow for all r modes cannot be satisfied. Thus constraint (16) ensures that the combined flow of zones with modes in set R_{IJ} is greater than or equal to the related FAF mode flow. If the $\hat{f}_{ijc}$ targets satisfy constraints (15) and (16) no adjustment occurs (i.e., f_{ijc} equals $\hat{f}_{ijc}$ for all ij pairs). This is the case of the example presented in section 4. Conversely, in situations like the one presented in the animal feed example the modal information would completely override the disaggregation suggested by the production and attraction variables. Note that objective function (14) can be linearized, and it could also be substituted by formulations such as least squares deviations. Figure 2 presents a schematic of the integrated procedure suggested by the two optimization problems presented.

POSSIBLE VALIDATION PROCEDURES FOR DISAGGREGATION AND MODE FLOW ALLOCATION RESULTS

Testing the validity of disaggregated commodity flows and the related mode allocations at the sub-FAF level is primarily hindered by the same problem that motivates the use of FAF disaggregation and mode allocation procedures, namely, the dearth of commodity flow data at county and sub-county levels. FAF disaggregation studies *(e.g., 3, 4)* have attempted to validate their model results with county level data from Transearch, a proprietary database *(21)*. These comparisons have shown non-trivial discrepancies between FAF and Transearch, in part because of definitional and conceptual differences between the two databases *(3)*. An alternative approach to Transearch-based validations is to compare modeled traffic assignments resulting from the estimated sub-FAF mode flows with observed transportation network data. Ideally these network data would be available for all the modes and commodity types, but in reality it is unlikely that a transportation agency has this level of information. Nevertheless, typically transportation agencies do collect truck count data at major roadways, and these data could allow for the creation of cordon, screenline, or cutline counts that can be used for a partial validation of the commodity flow disaggregation and truck allocation results without requiring significant data collection efforts. This validation approach requires of multiple assumptions, such as percentage of empty trucks, percentage of service trucks, tonnage-to-truck conversion factors, and seasonality factors. Available literature, including FAF^3 documentation *(22)*, provides reference values that can be used by transportation agencies to make informed assumptions on unknown factors. Note that, unless the transportation agency can identify the types of commodities being moved by the observed trucks, the truck count validation approach would have to be performed on the basis of aggregated truck counts, not commodity-specific truck counts.

In this study neither Transearch nor extensive truck count data were available to perform validation tests. However, for illustrative purposes Southern California truck assignments flows, estimated in part by procedures presented in this study, were compared to cutline counts (i.e., corridor level counts) (Figure 3). The comparisons between cutline and modeled counts have been suggested as an aggregate level validation and reasonableness check for travel models *(23)*. The cutlines used in this study were created with 2007 truck count data from eight weight-in-motion (WIM) stations located at major highways in the region. FAF^3-based county and sub-county flows for all commodities were borrowed from preliminary results of the California Statewide Freight Forecasting Model's (CSFFM) commodity generation and distribution modules *(6)*. These disaggregated commodity flows were allocated to truck, rail, truck-rail, and "remainder mode" categories using the least absolute deviation goal-programming model. The goal-programming problem targets were defined by FAF-based MFR mode split models and complementary assumptions similar to the ones presented in the sample application section. Factors based on the 2002 Vehicle Inventory and Use Survey were used to convert truck tonnage flows to truck vehicle flows, and to add empty trucks (*24)*. Additionally, truck vehicle flows were converted from annual to daily flows utilizing the CSFFM truck seasonality model. The truck segments of truck-rail movements were associated to transshipment facilities using the CSFFM transshipment model. Finally, the truck flows were assigned to a model of California's truck road network using a stochastic truck assignment model. The WIM stations data labeled each truck with the corresponding vehicle classification. Consequently, in this study the comparisons between the modeled and observed daily truck volumes could focus on the truck classes more likely to be carrying commodities in California (FHWA vehicle classes 8, 9, and 10). Table 3 reports the WIM and mode counts for each cutline, as well as the percent errors. Significant errors were observed for cutlines 2a, 4a, and 4b. There could be multiple reasons for these errors, including model prediction errors (e.g., commodity flow distribution, mode allocation), incorrect assumptions (e.g., percent of empty trucks), and even data problems in FAF^3. Given the limited extend of the cutlines (which is the result of limited availability of reliable truck counts), in this study it was not possible to identify with confidence what are the reasons for the discrepancies between the WIM and model counts. But this example demonstrates a relatively standard approach that can be used by transportation agencies to check if their FAF commodity flow disaggregation and mode allocation results are consistent with network observations.

CLOSING REMARKS

A goal-programming approach was presented for the mode allocation of disaggregated FAF commodity flows. The proposed method is intended as a second-step procedure following the distribution of FAF flows to sub-FAF zonal levels. A framework was suggested to connect previous total commodity flow disaggregation procedures with the optimization problems proposed in this paper. Note that the structure of the integrated procedure presented in Figure 2 is similar to the structure of the four-step commodity-based freight forecasting model, where commodity flow generation and mode split models are usually the first and third steps, respectively. This integrated framework could be expanded by introducing optimization problems to recalibrate the commodity flow disaggregation or mode allocation outputs based on traffic assignment results obtained from the mode flow allocations (the fourth step in the four-step model) and observed transportation network data. The formulation of these optimization problems could be the subject of future research. Additional research is also needed on the development of the FAF-based mode split models. For example, clustering procedures could be utilized to segment the commodity flow data (e.g., based on magnitude of the flow), and then separate mode split models could be estimated with each data group. Also, the utility of more complex model functional forms (e.g., nested logit) should be explored.

For agencies that do not have access to freight mode choice models, a method was proposed to estimate aggregate mode split models using FAF data. In this paper these models were presented as a way to define target values for the goal-programming models. However, the mode split models could also be used by these agencies to perform aggregate level policy analysis, as the mode split model coefficients reflect to some degree the sensitivity of mode shares to changes in modal attributes, such as transit costs. A possible modeling approach is to develop incremental logit models by borrowing the parameters of the estimated mode split models and using sub-FAF mode shares (or even the FAF mode shares) as the base shares. However, caution should be taken in the estimation, interpretation, and use of FAF-based aggregate mode split models since segments of the FAF data itself are the result of several intermediate models, especially for the CFS out-of-scope commodity flows *(1, 13)*.

ACKNOWLEDGEMENTS

The research reported in this paper was supported by the California Department of Transportation. The authors gratefully acknowledge comments provided by Fatemeh Ranaiefar on an earlier version of this paper. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. Southworth, F., Davidson, D., Hwang, H., Peterson, B.E., and S. Chin. *The Freight Analysis Framework, Version 3: Overview of the FAF3 National Freight Flow*. FHWA, U.S. Department of Transportation, 2010. http://faf.ornl.gov/fafweb/Data/FAF3ODDoc611.pdf. Accessed July 28, 2013.
2. Alams, M., Fekpe, E., and M. Majed. *FAF^2 Truck O-D Data Disaggregation.* FHWA, U.S. Department of Transportation, 2007. http://www.ops.fhwa.dot.gov/freight/freight_analysis /faf/faf2_reports/reports7/c4_data.htm. Accessed July 28, 2013.
3. Viswanathan, K., Beagan, D., Mysore, V., and N. Srinivasan. Disaggregating Freight Analysis Framework Version 2 Data for Florida: Methodology and Results. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2049,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 167-175.
4. Opie, K., Rowinski, J., and L.N., Spasovic. Commodity-Specific Disaggregation of 2002 Freight Analysis Framework Data to County Level in New Jersey. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2121*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 128-134.
5. Ruan, M., and J. Lin. Generating County-level freight data using Freight Analysis Framework (FAF2.20) for Regional Truck Emission Estimation. Presented at Transport Chicago, Chicago, Il., 2009.
6. Ranaiefar, F., Chow, J.Y.J, Rodriguez-Roman, D., Camargo, P.V., and S.G. Ritchie. Geographic Scalability and Supply Chain Elasticity of a Structural Commodity Generation Model Using Public Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2378,* Transportation Research Board of the National Academies, Washington, D.C., 2013.
7. Railinc. *Reference Guide for the 2007 Surface Transportation Board Carload Waybill Sample*. Surface Transportation Board, U.S Department of Transportation, 2008.
8. Cambridge Systematics. *California Commodity Origin-Destination Database Disaggregation*. U.S. Department of Transportation, 2010.
9. Sivakumar, A., and C.R., Bhat. Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1790*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 80-88.
10. Ye, X., and R.M. Pendyala. A Model of Daily Time Use Allocation Using Fractional Logit Methodology. In H.S. Mahmassani, Ed. Transportation and Traffic Theory: Flow, Dynamics, and Human Interaction. Pegamon, Elsevier Science Ltd., Oxford, 2005.

11. Bhat, C.R., and H.S. Nair. VMT-Mix Modeling for Mobile Source Emissions Forecasting: Formulation and Empirical Application. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1738*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 39-48.
12. Sivakumar, A., and C.R. Bhat. *Freight Modal Split Modeling: Conceptual Framework, Model Structure, and Data Sources.* FHWA/TX-01/0-1833-4. Texas Department of Transportation, 2000.
13. Bureau of Transportation Statistic. *2007 Commodity Flow Survey*. RITA, U.S. Department of Transportation, 2010.
14. Ortuzar, J.D., and L.G. Willumsen. Modelling Transport. John Wiley & Sons, West Sussex, 2001.
15. TMW Systems. Freight Rates. http://www.truckloadrate.com/market_truck_rates.htm. Accessed July 28, 2013.
16. U.S. Department of Energy. Weekly Retail Gasoline and Diesel Prices. http://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_w.htm. Accessed July 28, 2013.
17. Resor, R.R., and J.T. Blaze. Short-Haul Rail Intermodal: Can It Compete with Trucks? In *Transportation Research Record: Journal of the Transportation Research Board, No. 1873*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 45-62.
18. Cambridge Systematics, Global Insight, Cohe, H., Horowitz, A., and R. Pendyala. *NCHRP Report 606: Forecasting Statewide Freight Toolkits*. Transportation Research Board of the National Academies, Washington, D.C., 2008.
19. BNSF. Carload Transit Schedules. http://www.bnsf.com/bnsf.was6/goaltp/cntrl. Accessed Sept. 24, 2012.
20. BNSF. Intermodal Schedules. http://www.bnsf.com/customers/prices-and-tools/intermodal-schedules/. Accessed March 7, 2013.
21. HIS Global Insight. *Transearch*. http://www.ihs.com/products/global-insight/industry-analysis/commerce-transport/database.aspx. Accessed November 10, 2013.
22. Battelle. *FAF^3 Freight Traffic Analysis*. Final Draft Report. Oak Ridge National Laboratory, 2011.
23. Cambridge Systematics, Inc. *Travel Model Validation and Reasonableness Checking Manual*. FHWA, US Department of Transportation, 2010.
24. Bureau of the Census. *Vehicle Inventory and Use Survey*. US Department of Commerce, 2004.

List of Table Titles

List of Figure Captions

TABLE 1 Results for Disaggregation and Mode Flow Allocation Procedures

Zones	f_{ijc}	Target Mode Flows				L_1 Solution				L_2 Solution			
		T	R	TR	RM	T	R	TR	RM	T	R	TR	RM
603700	2.4	2.1	0.0	0.2	0.1	2.1	0.0	0.2	0.1	1.8	0.0	0.1	0.5
603701	1.9	1.7	0.0	0.2	0.1	1.7	0.0	0.2	0.1	1.4	0.0	0.1	0.5
603702	17.0	12.8	2.2	1.5	0.5	12.8	3.7	0.0	0.5	11.8	4.5	0.6	0.1
603703	2.7	2.4	0.0	0.3	0.1	2.4	0.0	0.3	0.1	2.0	0.0	0.2	0.5
603704	22.8	19.9	0.0	2.3	0.6	19.9	0.0	2.3	0.6	19.6	0.0	2.2	1.0
603705	30.0	22.8	3.8	2.6	0.8	5.1	24.2	0.0	0.8	21.7	6.1	1.7	0.5
603706	16.2	12.3	2.1	1.4	0.4	12.3	3.5	0.0	0.4	11.2	4.4	0.5	0.1
603707	28.6	21.7	3.6	2.5	0.8	21.7	6.1	0.0	0.8	20.6	5.9	1.6	0.4
603708	13.2	10.0	1.7	1.1	0.4	10.0	2.8	0.0	0.4	8.9	4.0	0.3	0.0
603709	0.7	0.6	0.0	0.1	0.0	0.6	0.0	0.1	0.0	0.3	0.0	0.0	0.4
603710	8.9	6.7	1.1	0.8	0.2	6.7	1.9	0.0	0.2	5.6	3.3	0.0	0.0
603711	33.6	25.6	4.2	2.9	0.9	25.6	4.5	2.7	0.9	24.5	6.5	2.0	0.6
605900	29.3	25.6	0.0	2.9	0.8	25.6	0.0	2.9	0.8	25.3	0.0	2.8	1.2
605901	3.7	3.2	0.0	0.4	0.1	3.2	0.0	0.4	0.1	2.9	0.0	0.3	0.5
605902	29.1	22.1	3.7	2.5	0.8	22.1	3.7	2.5	0.8	21.0	6.0	1.7	0.4
605903	15.6	11.9	2.0	1.3	0.4	11.9	3.3	0.0	0.4	10.8	4.3	0.5	0.1
606500	8.7	7.6	0.0	0.9	0.2	7.6	0.0	0.9	0.2	7.3	0.0	0.7	0.7
606501	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
606502	1.7	1.5	0.0	0.2	0.0	1.5	0.0	0.2	0.0	1.2	0.0	0.1	0.5
606503	13.0	9.9	1.6	1.1	0.4	9.9	1.6	1.1	0.4	8.8	3.9	0.2	0.0
607100	18.2	13.9	2.3	1.6	0.5	13.9	2.3	1.6	0.5	12.8	4.6	0.7	0.1
607101	2.3	1.8	0.3	0.2	0.1	1.8	0.5	0.0	0.1	0.2	2.1	0.0	0.0
607102	13.9	10.6	1.7	1.2	0.4	10.6	1.7	1.2	0.4	9.5	4.0	0.3	0.0
611100	11.1	9.6	0.0	1.1	0.3	9.6	0.0	1.1	0.3	9.3	0.0	1.0	0.7
Total	324.6	256.1	30.3	29.3	8.8	238.4	59.8	17.6	8.8	238.4	59.8	17.6	8.8

Notes: T=Truck, R=Rail, TR=Truck-Rail, RM=Remaining Modes. Zones with rail accessibility have bold identifications. All flows in kton units.

TABLE 2 Estimated Parameters for the MFR Model

Name	Value	Robust t-stat
$\beta_{0,truck}$	1.980	6.60
$\beta_{0,rail}$	0.517	0.79
$\beta_{0,truck-rail}$	0	-
β_{cost}	-0.413	-1.85
β_{time}	-0.947	-3.52

Sample size	:	253
Null log-likelihood	:	-277.9
Final log-likelihood	:	-125.3
$\bar{\rho}^2$	:	0.535

TABLE 3 Comparison between modeled and observed truck counts

Cutlines	Cutline Volume Direction Description	WIM Counts	Model Counts	Percent Error
1	Leaving through east of Los Angeles County	10574	10966	3.7%
2a	Leaving San Bernardino County toward Arizona	5949	3841	-35%
2b	Entering San Bernardino County from Arizona	6190	6862	10%
3	Leaving Orange County toward east direction	9878	9094	-7%
4a	Leaving Riverside County toward east and south directions	7348	10657	45%
4b	Entering Riverside County from east and south directions	5900	8172	38%

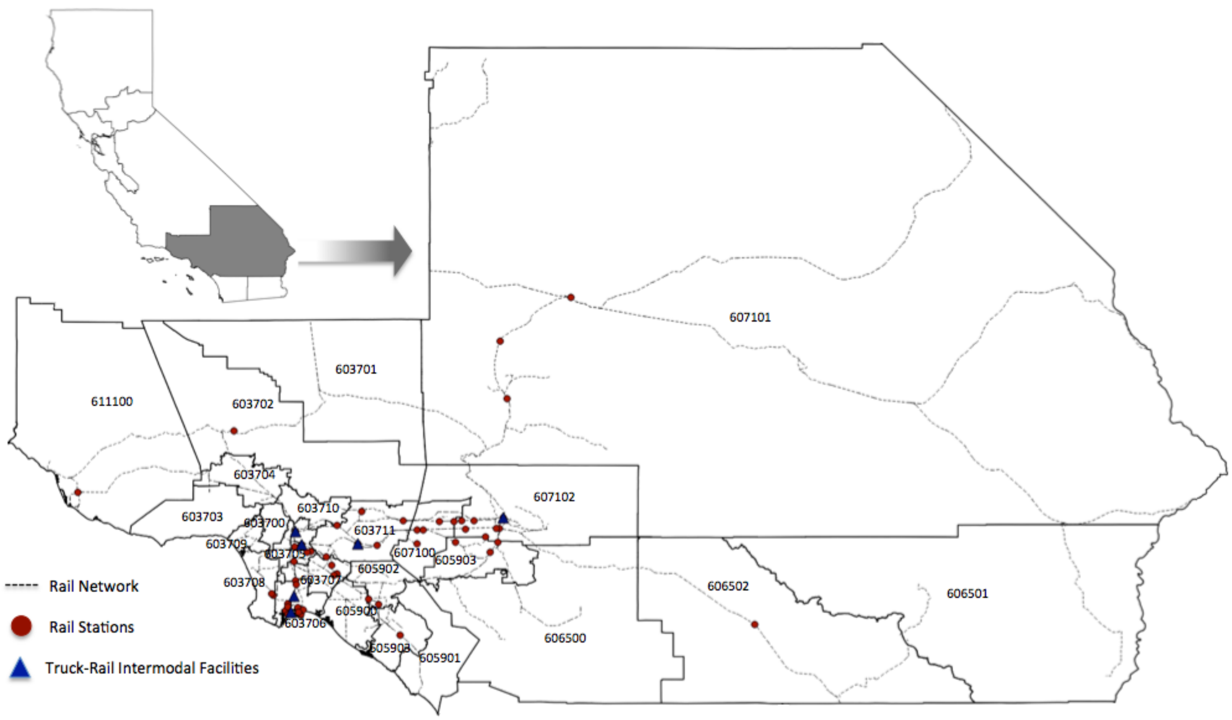

FIGURE 1 Disaggregated zonal structure for the Los Angeles FAF^3 zone.

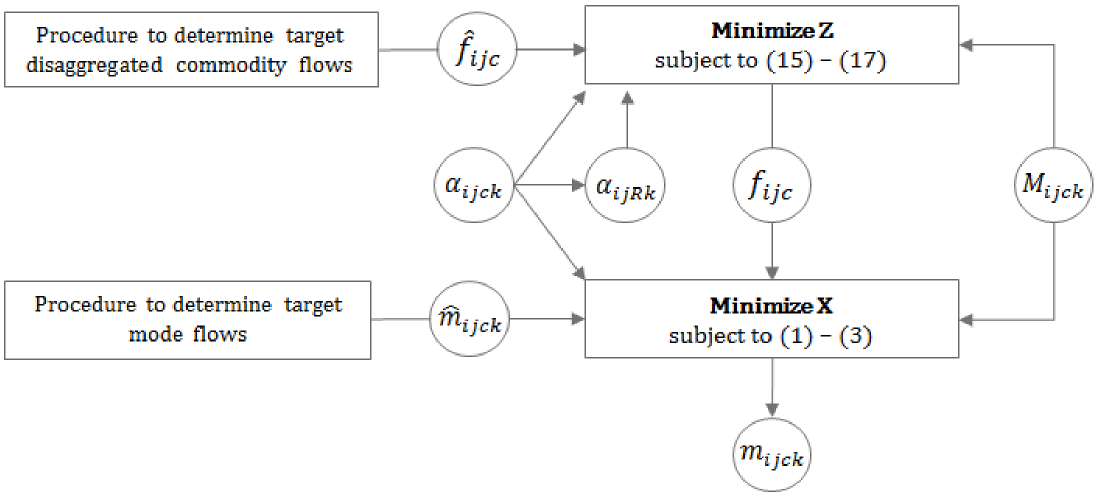

FIGURE 2 Framework to disaggregate commodity flows and allocate mode flows for a FAF OD.

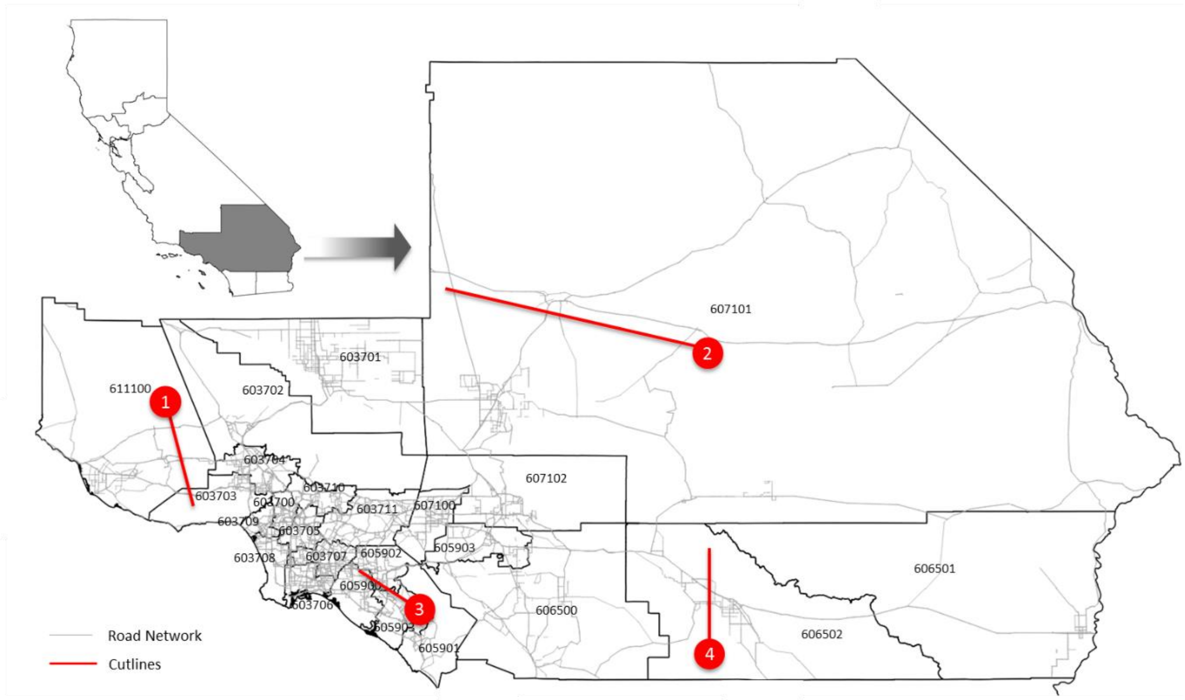

FIGURE 3 Approximate location of the corridor cutlines.