

DISAGGREGATING FAF² DATA FOR CALIFORNIA

Sarah E. Aly

Department of Civil and Environmental Engineering, Institute of Transportation Studies

Amelia Regan

Department of Computer Science, Institute of Transportation Studies

University of California, Irvine, CA 92697-2175, USA

Abstract

The Freight Analysis Framework Version 2 (FAF²) data estimates commodity flows and related freight transportation activity among states, sub-state regions, and major international gateways (FHWA 2006). The FAF² commodity flow origin-destination database includes tons and value of commodity movements among regions by mode of transportation and type of commodity. The FAF² data for the State of California is represented through a Commodity Origin-Destination Database comprising aggregated data for five major FAF² zones. These zones are: Los Angeles, San Diego, Sacramento, San Jose, and “the remainder of California” and they encompass a total of 58 counties. Disaggregating such data from a 5 zone level to a 58 county level can significantly enhance the understanding of freight flow volume and distribution in California which can be used to improve state freight flow planning.

A variety of methods for the disaggregation of the FAF² data have been developed. Some techniques use socioeconomic factors such as employment and population as a basis for their disaggregation procedures, while others use a Truck Vehicle Miles Traveled (TVMT) approach. This paper presents an analysis of a revised Truck Vehicle Miles Traveled (TVMT) based method (Rowinski et al 2007) to disaggregate the FAF² data pertaining to California for the year 2002. This work is part of a larger project aimed towards developing a freight analysis framework for the State of California. The main factor used for this disaggregation is the ratio of TVMT within the county level to the TVMT within the respective FAF² zone level. The analysis was carried out using this factor for both the Origins and Destinations as illustrated in a numerical example. The results of this work provide a unique insight to freight flow volume and distribution within the State of California. The methodology has proven to be an efficient approach to disaggregating FAF² data and can be easily applied to future projections.

California FAF² Data Overview

FAF² data estimates commodity flows and related freight transportation activity among states, sub-state regions, and major international gateways. The FAF² commodity flow origin-destination database includes tons and value of commodity movements among regions by mode of transportation and type of commodity. The FAF² Commodity Origin-Destination Database for the year 2002, specifically, was developed entirely from public data sources that include, but are

not limited to, the 2002 Commodity Flow Survey (CFS) as well as Foreign Waterborne Cargo data developed by the U.S. Army Corps of Engineers.

The Database is comprised of 114 regions as defined and used in the 2002 CFS plus 17 additional international gateways and 7 international regions. Commodities are defined at the 2-digit SCTG (Standard Classification of Transported Goods) level. The complete FAF² 2002 U.S. Commodity Flows Matrix consists of 138 origin and destination regions, broken down by 43 commodity classes and by 7 major mode/mode combinations.

The FAF² data for California is represented through a Commodity Origin-Destination Database representing five major FAF² zones (see Figure 1). These zones are: Los Angeles, San Diego, Sacramento, San Jose, and “the remainder of California” and they encompass a total of 58 counties (see Table 1 for a list of the counties related to each CA FAF² zone).



Figure 1: California FAF² Zones

<i>County</i>	<i>FAF² Zone</i>	<i>County</i>	<i>FAF² Zone</i>
<i>Los Angeles</i>	<i>CA Los Angles</i>	<i>Glenn</i>	<i>CA Remainder</i>
<i>Orange</i>	<i>CA Los Angles</i>	<i>Humboldt</i>	<i>CA Remainder</i>
<i>Riverside</i>	<i>CA Los Angles</i>	<i>Imperial</i>	<i>CA Remainder</i>
<i>San Bernardino</i>	<i>CA Los Angles</i>	<i>Inyo</i>	<i>CA Remainder</i>
<i>Ventura</i>	<i>CA Los Angles</i>	<i>Kern</i>	<i>CA Remainder</i>
<i>San Diego</i>	<i>CA San Diego</i>	<i>Kings</i>	<i>CA Remainder</i>
<i>El Dorado</i>	<i>CA Sacramento</i>	<i>Lake</i>	<i>CA Remainder</i>
<i>Nevada</i>	<i>CA Sacramento</i>	<i>Lassen</i>	<i>CA Remainder</i>
<i>Placer</i>	<i>CA Sacramento</i>	<i>Madera</i>	<i>CA Remainder</i>
<i>Sacramento</i>	<i>CA Sacramento</i>	<i>Mariposa</i>	<i>CA Remainder</i>
<i>Yolo</i>	<i>CA Sacramento</i>	<i>Mendocino</i>	<i>CA Remainder</i>
<i>Alameda</i>	<i>CA San Jose</i>	<i>Merced</i>	<i>CA Remainder</i>
<i>Contra Cosa</i>	<i>CA San Jose</i>	<i>Modoc</i>	<i>CA Remainder</i>
<i>Marin</i>	<i>CA San Jose</i>	<i>Mono</i>	<i>CA Remainder</i>
<i>Napa</i>	<i>CA San Jose</i>	<i>Monterey</i>	<i>CA Remainder</i>
<i>San Benito</i>	<i>CA San Jose</i>	<i>Plumas</i>	<i>CA Remainder</i>
<i>San Francisco</i>	<i>CA San Jose</i>	<i>San Joaquin</i>	<i>CA Remainder</i>
<i>San Mateo</i>	<i>CA San Jose</i>	<i>San Luis Obispo</i>	<i>CA Remainder</i>
<i>Santa Clara</i>	<i>CA San Jose</i>	<i>Santa Barbara</i>	<i>CA Remainder</i>
<i>Santa Cruz</i>	<i>CA San Jose</i>	<i>Shasta</i>	<i>CA Remainder</i>
<i>Solano</i>	<i>CA San Jose</i>	<i>Sierra</i>	<i>CA Remainder</i>
<i>Sonoma</i>	<i>CA San Jose</i>	<i>Siskiyou</i>	<i>CA Remainder</i>
<i>Alpine</i>	<i>CA Remainder</i>	<i>Stanislaus</i>	<i>CA Remainder</i>
<i>Amador</i>	<i>CA Remainder</i>	<i>Sutter</i>	<i>CA Remainder</i>
<i>Butte</i>	<i>CA Remainder</i>	<i>Tehama</i>	<i>CA Remainder</i>
<i>Calaveras</i>	<i>CA Remainder</i>	<i>Trinity</i>	<i>CA Remainder</i>
<i>Colusa</i>	<i>CA Remainder</i>	<i>Tulare</i>	<i>CA Remainder</i>
<i>Del Norte</i>	<i>CA Remainder</i>	<i>Tuolumne</i>	<i>CA Remainder</i>
<i>Fresno</i>	<i>CA Remainder</i>	<i>Yuba</i>	<i>CA Remainder</i>

Table 1: California Counties Comprising FAF² Zones

Freight Flow in California

Tables 2 and 3 present information on the distribution of commodity flows by mode for California for the years 2002 and 2035 respectively (OPS 2008). As shown in Table 2, trucks are responsible for a large percentage of the tonnage of shipments, followed by rail and water. This distribution is expected to remain fairly equal for the year 2035 as displayed in Table 3. Since commodity flows are projected to more than double by 2035, truck traffic is also expected to grow throughout the state. Much of this growth will occur in urban areas and on the Interstate highway system.

It vitally important to account this future freight flow increase in future development plans for the California highway and intermodal transportation systems. It is especially imperative to have a good understanding of the distribution of this traffic in order to properly plan for future expansions and enhancements. At the moment, the FAF² data only gives information on commodity flows between five major FAF² zones representing the entire State of California.

These FAF² zones comprise a total of 58 counties. Therefore, the level of detail provided by FAF² data would be much enhanced through the disaggregation of this data from a zonal level to a county level. This is especially true in the case of USA Zone 12: CA Remainder as it consists of a total of 36 counties scattered all around the state (see Figure 1). Note that it is not the case that the FHWA collects disaggregate data and then makes it available to agencies in only aggregate forms. The limitations of the commodity flow survey, which include small sample sizes and a need to carefully protect the privacy and proprietary data of participating establishments, means that for all practical purposes disaggregated data do not exist.

Mode	Within State		From State		To State	
	Number	%	Number	%	Number	%
Truck	971.0	86	92.8	73	144.2	44
Rail	9.8	<1	11.7	9	63.6	19
Water	7.8	<1	1.2	1	23.6	7
Air, air & truck	0.1	<1	0.4	<1	0.6	<1
Truck & rail	1.3	<1	4.0	3	4.8	1
Other intermodal	1.5	<1	5.0	4	2.8	<1
Pipeline & unknown	140.8	12	12.4	10	87.2	27
Total	1,132.2	100	127.4	100	326.7	100

Table 2: Shipments by Weight for 2002 (Millions of Tons)

Mode	Within State		From State		To State	
	Number	%	Number	%	Number	%
Truck	2,179.4	88	366.0	77	518.4	63
Rail	15.8	<1	35.4	7	82.1	10
Water	11.9	<1	2.2	<1	42.0	5
Air, air & truck	0.8	<1	2.6	<1	1.2	<1
Truck & rail	2.5	<1	14.3	3	6.6	<1
Other intermodal	4.4	<1	29.5	6	6.2	<1
Pipeline & unknown	251.6	10	26.7	6	168.5	20
Total	2466.3	100	476.9	100	825.1	100

Table 3: Shipments by Weight for 2035 (Millions of Tons)

Disaggregation Method Review

A literature review indicates the existence of many different methods for disaggregation. Most methods tend to use socioeconomic factors such as employment and population as a basis for their disaggregation procedures. Krishnan Viswanathan et al. (2008) uses the relationship between employment by industry and the commodities those industries produce and consume in order to disaggregate FAF² data for Florida.

Battelle (2006) outlines a Truck Vehicle Miles Traveled (TVMT) approach to disaggregate commodity flow data. The method uses the proportion of zone TVMT present within a county in that zone to disaggregate the FAF² zone based commodity flows down to the county level. The

disaggregation is done in the software package TransCAD, using the matrix-disaggregate procedure. The disaggregated commodity flows at the county level are then distributed using TransCAD's Tri-Proportional Trip Distribution Gravity Model. Finally, Battelle assigns the resulting matrix over the FAF² network for the purpose of calibrating the approach.

Lastly, in the second year of a three year project for the State of New Jersey aimed at developing methods to disaggregate the large commodity origin-destination data covered in FAF² into small geographic areas (county level), truck flows exiting and entering the county borders were one of many factors used for disaggregation (Opie et al, 2008).

Methodology

Based on the literature reviewed and data availability, the method chosen to disaggregate the California FAF² data was the proportional approach described by Rowinski et al (2007) as D4. The D4 method uses an approach similar to the previously discussed Battelle methodology with the difference being that commodity flow data is used for validation rather than trip-based data. The main factor used for this disaggregation is the ratio of Truck Vehicle Miles Traveled (TVMT) at the county level to the TVMT within the larger FAF² zone level. The TVMT data for the year 2002 was found in a report developed by the California Department of Transportation (2005).

The process carried out to obtain commodity flows in Ktons on a county level is as follows. First, the commodity flow data for California is extracted from the FAF² domestic data for the year 2002. This includes all records with either an origin or a destination consisting of one of the 5 FAF² zones that make up California. This data was then divided into 20 matrices based on which zone it pertained to. For example, the commodity flow data originating from, destined to, or traveling within US Zone 8: CA Los Angeles was split into three matrices: CA Los Angeles From, CA Los Angeles To, and CA Los Angeles Within. Also, a fourth matrix was created to represent commodity flows within California that originated in CA Los Angeles. This system was implemented for each of the 5 zones thus resulting in a total of 20 matrices. After the matrices were created, the respective data for 2002 for commodity flows to, from, or within the zone was multiplied by the disaggregation factors for each county (see Table 4) comprising the five zones. The equation used is as follows:

$$F_{c(n)} = F_{(n)} \times \left(\frac{TVMT_{c(n)}}{TVMT_{(n)}} \right) \quad (1)$$

where

$F_{c(n)}$ = Freight tons generated by county $c(n)$.

$F_{(n)}$ = Freight tons generated by FAF² zone (n) .

$TVMT_{c(n)}$ = Total truck vehicle miles travelled in county $c(n)$.

$TVMT_{(n)}$ = Total truck vehicle miles travelled in FAF² zone (n) .

$$\left(\frac{TVMT_{c(n)}}{TVMT_{(n)}}\right) = \text{Disaggregation factor (DF) for county } c(n).$$

It is important to note that this procedure was applied twice when performing the disaggregation within California, either within or between any of the 5 FAF² zones in order to obtain the appropriate county level flows. Example 1 below demonstrates the need for such a procedure.

Example 1

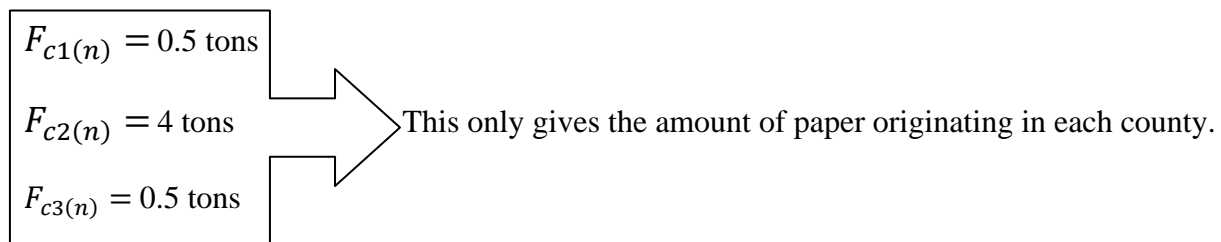
There are five tons of paper originating and ending up in zone (n) which consists of three counties - c1, c2, and c3.

$$F_{(n)} = 5 \text{ tons}$$

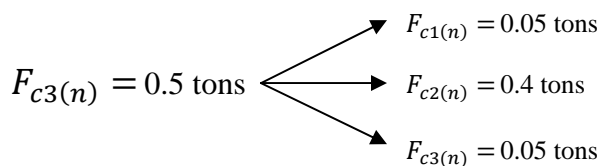
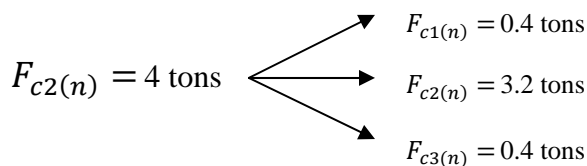
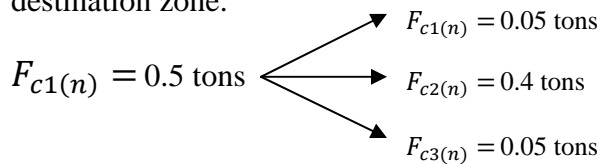
$$DF_{c1(n)} = 0.1$$

$$DF_{c2(n)} = 0.8$$

$$DF_{c3(n)} = 0.1$$



Next, it's important to disaggregate the amount of paper destined to each county. In this case, we multiply each value by the same disaggregation factor since the origin and destination zones are the same. In the case where they vary, however, we would have to multiply the freight tons value obtained from the initial step by the disaggregation factors pertaining to the counties in the destination zone.



Results

The disaggregation factors determined through the procedure of dividing the total TVMT of a county by that of the entire FAF² zone was applied for the year 2002 in this study. This calculation was performed based on data provided by the California Department of Transportation. The factors related to each FAF² zones are presented in the table below.

<i>County</i>	<i>DF</i>	<i>FAF² Zone</i>	<i>County</i>	<i>DF</i>	<i>FAF² Zone</i>
<i>Los Angeles</i>	0.3908	<i>CA Los Angles</i>	<i>Glenn</i>	0.0122	<i>CA Remainder</i>
<i>Orange</i>	0.1196	<i>CA Los Angles</i>	<i>Humboldt</i>	0.0133	<i>CA Remainder</i>
<i>Riverside</i>	0.1981	<i>CA Los Angles</i>	<i>Imperial</i>	0.0356	<i>CA Remainder</i>
<i>San Bernardino</i>	0.2608	<i>CA Los Angles</i>	<i>Inyo</i>	0.0072	<i>CA Remainder</i>
<i>Ventura</i>	0.0308	<i>CA Los Angles</i>	<i>Kern</i>	0.2129	<i>CA Remainder</i>
<i>San Diego</i>	1	<i>CA San Diego</i>	<i>Kings</i>	0.0276	<i>CA Remainder</i>
<i>El Dorado</i>	0.0509	<i>CA Sacramento</i>	<i>Lake</i>	0.0050	<i>CA Remainder</i>
<i>Nevada</i>	0.0904	<i>CA Sacramento</i>	<i>Lassen</i>	0.0080	<i>CA Remainder</i>
<i>Placer</i>	0.1971	<i>CA Sacramento</i>	<i>Madera</i>	0.0338	<i>CA Remainder</i>
<i>Sacramento</i>	0.4460	<i>CA Sacramento</i>	<i>Mariposa</i>	0.0010	<i>CA Remainder</i>
<i>Yolo</i>	0.2156	<i>CA Sacramento</i>	<i>Mendocino</i>	0.0092	<i>CA Remainder</i>
<i>Alameda</i>	0.2933	<i>CA San Jose</i>	<i>Merced</i>	0.0596	<i>CA Remainder</i>
<i>Contra Cosa</i>	0.1137	<i>CA San Jose</i>	<i>Modoc</i>	0.0026	<i>CA Remainder</i>
<i>Marin</i>	0.0319	<i>CA San Jose</i>	<i>Mono</i>	0.0039	<i>CA Remainder</i>
<i>Napa</i>	0.0210	<i>CA San Jose</i>	<i>Monterey</i>	0.0432	<i>CA Remainder</i>
<i>San Benito</i>	0.0278	<i>CA San Jose</i>	<i>Plumas</i>	0.0029	<i>CA Remainder</i>
<i>San Francisco</i>	0.0206	<i>CA San Jose</i>	<i>San Joaquin</i>	0.1246	<i>CA Remainder</i>
<i>San Mateo</i>	0.0934	<i>CA San Jose</i>	<i>San Luis Obispo</i>	0.0294	<i>CA Remainder</i>
<i>Santa Clara</i>	0.1960	<i>CA San Jose</i>	<i>Santa Barbara</i>	0.0328	<i>CA Remainder</i>
<i>Santa Cruz</i>	0.0248	<i>CA San Jose</i>	<i>Shasta</i>	0.0318	<i>CA Remainder</i>
<i>Solano</i>	0.1122	<i>CA San Jose</i>	<i>Sierra</i>	0.0021	<i>CA Remainder</i>
<i>Sonoma</i>	0.0653	<i>CA San Jose</i>	<i>Siskiyou</i>	0.0249	<i>CA Remainder</i>
<i>Alpine</i>	0.0006	<i>CA Remainder</i>	<i>Stanislaus</i>	0.0448	<i>CA Remainder</i>
<i>Amador</i>	0.0037	<i>CA Remainder</i>	<i>Sutter</i>	0.0068	<i>CA Remainder</i>
<i>Butte</i>	0.0093	<i>CA Remainder</i>	<i>Tehama</i>	0.0204	<i>CA Remainder</i>
<i>Calaveras</i>	0.0027	<i>CA Remainder</i>	<i>Trinity</i>	0.0021	<i>CA Remainder</i>
<i>Colusa</i>	0.0176	<i>CA Remainder</i>	<i>Tulare</i>	0.0601	<i>CA Remainder</i>
<i>Del Norte</i>	0.0027	<i>CA Remainder</i>	<i>Tuolumne</i>	0.0034	<i>CA Remainder</i>
<i>Fresno</i>	0.0963	<i>CA Remainder</i>	<i>Yuba</i>	0.0063	<i>CA Remainder</i>

Table 4: 2002 Disaggregation Factors per County and Corresponding FAF² Zones

Next, after determining the disaggregation factors for the counties in California, the data was imported into Excel and the necessary formulas were carried out to translate the base origin-destination matrix for freight flow in California to a county level. A snapshot of the results describing commodity flow between Zone 7: Arkansas and Zone 8: CA Los Angeles is depicted in Table 5.

Origin	Commodity	Mode	2002	Destination (Zone 8- CA Los Angeles)				
				Los Angeles	Orange	Riverside	San Bernardino	Ventura
AK	Alcoholic bev.	Other Intermodal	0.2	0.07816	0.02392	0.0396	0.05216	0.00616
AK	Furniture	Truck	0.55	0.21494	0.06578	0.1089	0.14344	0.01694
AK	Meat/seafood	Air & Truck	0.28	0.109424	0.033488	0.05544	0.073024	0.00862
AK	Meat/seafood	Other Intermodal	0.34	0.132872	0.040664	0.06732	0.088672	0.01047
AK	Mixed freight	Truck	0.8	0.31264	0.09568	0.1584	0.20864	0.02464
AK	Motorized veh.	Truck	6.00E-03	0.0023448	0.0007176	0.001188	0.0015648	0.00018
AK	Plastics/rubber	Pipeline & Unknown	0.49	0.191492	0.058604	0.09702	0.127792	0.01509
AK	Printed prods.	Truck	1.66	0.648728	0.198536	0.32868	0.432928	0.05113

Table 5: Commodity Flow from Zone 7: Arkansas to Zone 8- CA Los Angeles in KTONs.

Analysis of Results

An important factor of this study is to determine the validity of using such a model to disaggregate FAF² data for California. This is because this data is of immense importance and must be a good representation in order to be useful for future planning purposes. It is, however, quite difficult to measure the goodness of fit of this, or any, disaggregated commodity flow data as it is, essentially, an estimate of data that is not readily available.

A literature review proposes the use of Transearch data as a basis for comparison (Guiliano et al 2007). In this instance, the data would be validated by measuring it against data that is considered to be a fairly good estimate of reality. However, since the methods used to develop the Transearch data are not publicly available, the validation of this data will really only result in a rough estimation of its goodness of fit.

Another major issue to consider is that a discrepancy between the Transearch data, when aggregated to the FAF² zone structure, and the FAF² data (Rowinski et al 2007). Since the method used in this paper disaggregates the data from the zonal level to a county level as opposed to from an aggregated state level to a county level, an overall difference between disaggregated FAF² and Transearch is inevitable. Therefore, our future research is partly concerned with developing a reliable unbiased estimator that will allow us to validate selected freight flows and therefore provide confidence in this overall method.

Summary

This paper applied a previously developed method to disaggregate FAF² commodity data down to the California county level. This was done using disaggregation factors developed using the ratios of total TVMT within the county to the total TVMT within the FAF² zone level. The results obtained using this procedure appear to be fairly good and satisfied the objective of disaggregating the FAF² data for California down to a county level. However, the development

of a new method for validation is extremely important as it will give greater value to the results of this study. Future work will be focused on the development of such a method as well as testing and validating other related disaggregation methods. In addition, we are concerned with the development of data analysis techniques that will allow us to combine disparate freight flow information sources to improve the quality of the underlying data used in this and other studies.

Acknowledgement

This research has been supported by the California Department of Transportation and the California Air Resource Board. The authors gratefully acknowledge this support. Opinions expressed are those of the authors alone. We welcome input into this study and will make our data available to interested researchers or public agencies upon request.

References

1. Battelle Institute. *Development of County-to-County Freight O-D Matrix from 2002 FAF2 Freight Data for Subsequent Network Assignment*. Columbus, Ohio, 2006.
2. California Department of Transportation. *Truck Miles of Travel, California State Highway System*. July 2005.
3. Guiliano, Genevieve, Peter Gordon, Qisheng Pan, JiYoung Park, LanLan Wang. 2007. Estimating Freight Flows for Metropolitan Area Highway Networks Using Secondary Data Sources. *Networks and Spatial Economics* (July 20), <http://www.springerlink.com/content/871647u167624630/?p=8b1ea37b59524109b9fecccc609b0f07&pi=91>.
4. Federal Highway Administration (FHWA). *Freight Analysis Framework (FAF) Version 2.2, User Guide-Commodity Origin Destination Database: 2002-2035*. November 22, 2006.
5. Office of Freight Management and Operations (OPS). *Freight Shipments To, From, and Within California*. Federal Highway Administration. http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/state_info/faf2/ca.htm
6. Opie Keir, Jacob Rowinski, Lazar N. Spasovic. 2008. Commodity Specific Disaggregation of the 2002 FAF² data to the County Level for New Jersey. TRB, January 11-15, in Washington DC, USA.
7. Rowinski, Jacob, Keir Opie, and Lazar N. Spasovic. 2007. Development of a Method to Disaggregate the 2002 FAF² Data Down to the County Level for New Jersey. TRB, January 13-17, in Washington DC, USA.
8. Viswanathan, Krishnan, Daniel Beagan, Vidya Mysore, and Nanda Srinivasan. 2008. Disaggregating Freight Analysis Framework Version 2 Data for Florida-Methodology and Results. *Transportation Research Record: Journal of the Transportation Research Board* 2049: 167–75.