Allocating Disaggregated Freight Analysis Framework Truck–Rail Data

Roshan Kumar

ECJ 6.202 Department of Civil Engineering, University of Texas, Austin TX 78712

Krishnan Viswanathan*

Cambridge Systematics Inc., 2457 Care Drive, Suite 101, Tallahassee FL 32308

Daniel F. Beagan

Cambridge Systematics Inc., 100 CambridgePark Drive, Suite 400, Cambridge MA 02140

Vidya Mysore

Florida DOT, 605 Suwannee Street, MS 19, Tallahassee FL 32399

S. Travis Waller

ECJ 6.2 Department of Civil Engineering, University of Texas, Austin TX 78712

Abstract

This paper discusses a theoretical framework to assign disaggregated Freight Analysis Framework (FAF2) data by truck–rail between origin destination counties.

1. Introduction

The United States is entering the early stages of a capacity crisis. Freight transportation capacity is expanding too slowly to keep up with demand, and the freight productivity improvements gained though investment in the Interstate highway system and economic deregulation of the freight transportation industry in the 1980s are showing diminishing returns.

The effects of growing demand and limited capacity are felt as congestion, upward pressure on freight transportation prices, and less reliable trip times as freight carriers struggle to meet delivery windows. Higher transportation prices and lower reliability can mean increased supply costs for manufacturers, higher import prices, and a need for businesses to hold more expensive inventory to prevent stock outs. The effect on individual shipments and transactions is usually modest, but over time the costs can add up to a higher cost of doing business for firms, a higher cost of living for consumers, and a less productive and competitive economy.

To further exacerbate the situation, the transportation network has not increased at a rate commensurate with growth in travel and commerce. In the highway sector, for example, vehicle-miles traveled (VMT) increased by 80 percent while lane-miles of public roads increased by only 2 percent between 1980 and 2000. Growth in truck-miles traveled was even more dramatic, exceeding the growth in passenger VMT

*Corresponding author

Email addresses: roshan@mail.utexas.edu (Roshan Kumar), kviswanathan@camsys.com (Krishnan Viswanathan), dbeagan@camsys.com (Daniel F. Beagan), vidya.mysore@dot.state.fl.us (Vidya Mysore), stw@mail.utexas.edu (S. Travis Waller)

over the last few years . Clearly, more traffic is moving over essentially the same highway infrastructure. Other surface transportation networks are witnessing a similar overburdening of their systems as well.

In order to provide data and forecast to support policy analysis, the Federal Highway Administration (FHWA) has prepared the Freight Analysis Framework version 2 (FAF2). FAF2 provides a wealth of information to understand and evaluate the impact of freight on the transportation network. However, since the FAF2 data focuses on national policy and planning issues, and has large geographies, its usefulness to state DOTs or Metropolitan Planning Organizations (MPOs) is limited. Therefore, to make FAF2 more applicable to state DOTs and MPOs it is necessary to disaggregate the FAF2 regional origin destination (O-D) database to smaller geographies.

In this paper first we briefly describe the FAF2 database and then discuss the methodology that was developed to disaggregate the FAF2 data from FAF2 regions to US Counties. The next section talks about theoretical framework to apply the resulting disaggregated FAF2 flows for the truck–rail mode and illustrates it with an example along with the data needs to implement the framework.

2. The FAF2 Database

The FAF2 estimates commodity flows and related freight transportation activity among states, sub-state regions, and major international gateways. It also forecasts future flows among regions and relates those flows to the transportation network. FAF2 includes an origin-destination database of commodity flows among regions, and a network database in which flows are converted to truck payloads and related to specific routes.

The FAF2 commodity origin-destination database includes tons and value of commodity movements among regions by mode of transportation and type of commodity and contains projected commodity flow data ranging from 2010 to 2035 in five-year intervals as well as the 2002 base case data. In addition, the FAF2 excludes all foreign-to-foreign shipments via the United States.

The FAF2 2002 base year database is built entirely from public data sources. Key sources include the 2002 Commodity Flow Survey (CFS), developed by the Census Bureau, U.S. Department of Commerce, and the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation and Foreign Waterborne Cargo data, developed by the U.S. Army Corps of Engineers. FAF2 statistics do not match those in mode-specific publications primarily due to different definitions that are used to avoid double counting. FAF2 statistics should not be compared with the original FAF data because different methods and coverage are employed. Future projected data covering years from 2010 to 2035 with a five-year interval are based on IHS Global Insight's proprietary economic and freight modeling packages.

FAF2 has seven modes and 43 commodities. The 43 commodities are classified as per the Standard Classification for Transported Goods (SCTG) code at the 2-digit level. Details about the 43 commodities can be found in the FHWA FAF user guide (FHWA, 2006). The seven modes include:

- *Truck*: Includes private and for-hire truck. Private trucks are operated by a temporary or permanent employee of an establishment or the buyer/receiver of the shipment. For-hire trucks carry freight for a fee collected from the shipper, recipient of the shipment, or an arranger of the transportation.
- *Rail*: Any common carrier or private railroad.
- *Water*: Includes shallow draft, deep draft, and Great Lakes shipments. FAF2 uses definitions by the U.S. Army Corps of Engineers. Shallow draft includes barges, ships, or ferries operating primarily on rivers and canals; in harbors; the Saint Lawrence Seaway; the Intra-coastal Waterway; the Inside Passage to Alaska; major bays and inlets; or in the ocean close to the shoreline. Deep draft includes barges, ships, or ferries operating primarily in the open ocean.

- *Air (includes truck-air)*: Includes shipments by air or a combination of truck and air. Commercial or private aircraft and all air service for shipments that typically weigh more than 100 pounds. Includes air freight and air express.

- *Truck-Rail Intermodal*: Includes shipments by a combination of truck and rail.

- *Other Multiple Modes*: Includes shipments typically weighing less than 100 pounds by Parcel, U.S. Postal Service, or Courier, as well as shipments of all sizes by truck-water, water-rail, and other intermodal combinations.

- *Pipeline and Unknown*: Pipeline is included with unknown because region-to-region flows by pipeline are subject to large uncertainty.

3. FAF2 Data Disaggregation Methodology

A methodology was proposed which utilizes the relationship between employment by industry and the commodities which those industries produce and consume. For certain commodities, farm acreage, livestock and electricity generation information is used instead of or with North American Industrial Classification System (NAICS) employment data. While FAF2 data is available only at a regional level, employment by industry is more readily available at smaller levels of geography. The U.S. Census Bureau provides CBP employment by county by NAICS industry. Commercial and state data sources provide employment by NAICS or comparable industry classifications at smaller levels of geography. This employment data is aggregated to develop mathematical relationships between the FAF2 commodity shipments to and from a FAF2 region and the employment by industries in that FAF2 region. The availability of employment data by industry is used with these equations to estimate the expected production and attraction of freight tonnage in a FAF2 region and the units of smaller geography such as Counties in that FAF2 region. The shares of the smaller units of geography tonnage to the regional tonnage is then used to disaggregate the freight flows from FAF2 regions to the smaller units of geography within those FAF2 regions.

The steps to disaggregate the data from FAF2 regions to counties is summarized below and details applied to Florida can be found in (Viswanathan et. al., 2008) and at the national scale in an unpublished FHWA report (Cambridge Systematics, 2009).

- Step 1: The first step is to determine employment at the three digit NAICS county level.

- Step 2: The next step is to build a bridge between the commodities in the FAF2 reported using the two digit SCTG and three digit NAICS and other factors using regression methods. These regressions guide the development of factors for each commodity for the disaggregation of freight flow productions and attractions.

- Step 3: The third step involves determining the share of the originating and terminating tonnage by commodity for each of the counties within a specific FAF2 zone and is applied to the reported FAF2 regional tonnages, to obtain the disaggregated FAF2 O-D database for domestic freight shipment.

- Step 4: Subsequent to step 3, the modal share by county at the Port of Entry or Exit (POE) is determined and allocated to the county(s) where the POE is located.

- Step 5: The domestic origins and destinations for these exports and imports respectively are disaggregated in a manner similar to the flows (Step 3) that take place entirely within the United States.

- Step 6: The disaggregated import and export data (from POE origins destinations) is then merged with the domestic flow data and provides the final FAF2 disaggregated dataset.

The outcome of the above steps is a database that gives origin destination flows between counties by truck and all modes for commodity tonnage and value. The reason only the truck mode is disaggregated by itself is because of the presence of roads everywhere which allow trucks to traverse the network. However, with the increased emphasis on rail to move goods and people as a means of relieving congestion and mitigating the impacts of climate change, it is necessary to develop methods which allow for the assigning the disaggregated rail tonnage data between origin destination counties. The next section develops a method to assign the disaggregated data between the appropriate origin destination counties.

4. Allocating Rail and Truck–Rail Data

One of the issues in allocating disaggregated FAF2 tonnage data by rail is the lack of information about the location of rail transfer and intermodal centers and determining how goods movement take place by rail. While the information on the rail network can be obtained using the Oak Ridge National Laboratory (ORNL) CTA railroad network, it is not necessary that tonnage that moves between two FAF regions by rail, the same movement will apply when the data is disaggregated to counties. For rail only movements, it is obvious that for counties which do not have rail terminals goods are not being transported to those counties via rail. However, for truck–rail movements, distance of the shipment of the commodity producer or consumer from the closest intermodal center will, to a large extent dictate whether the movement occurs by truck–rail or only by truck. For example, the probability is higher that a shipper in Los Angeles County, CA might decide to use truck–rail mode to ship to the Chicago FAF2 region whereas a shipper in Imperial County, CA might decide to ship via trucks to the Chicago FAF2 region since it might not be economically viable to take it by truck from Imperial County to Los Angeles County and then put on a freight train and ship to the Chicago FAF2 region. Therefore, it is necessary to draw a distance buffer around the intermodal center and make the decision that all commodities produced or attracted to locations within the buffer will undertake part of their shipment via rail. Therefore, it is necessary to obtain establishment level employment data from either state Departments of Labor or private vendors such as InfoUSA and Woods & Poole and develop the relationships using the County level FAF2 disaggregation methods highlighted above.

5. Methodological Framework

Once the county–to–county truck–rail freight O–D Matrix has been estimated, as described in the previous section, the next step involves the assignment of these flows on links. The earlier state–of–practice involved an all–or–nothing (AON) assignment of freight flows between an O–D pair (FHWA, 2006). That is, all flows from an origin to a destination were assigned to the shortest path (calculated in terms of free flow travel costs) between them. The FHWA documentation on FAF2 (FHWA, 2006) illustrates that an all–or–nothing assignment will result in all the flow being assigned to *just one* highway link between the two FAF2 regions. But in reality, this is seldom the case. On a highway, truck and passenger–car traffic both interact and this interaction in turn affects highway congestion. Therefore, trucks originating from one county and destined to another, take routes that minimize their travel time. This implies that a user equilibrium (UE) assignment will be a better technique than an AON assignment.

5.1. Assignment of Freight Flows on Links

The assumption in the classical UE approach is that users have perfect information with regards to network congestion and delay; this assumption is relaxed in Stochastic User Equilibrium (SUE) models (Y. Sheffi, 1985). FHWA reports that Battelle developed an SUE modeling framework for assigning inter–zonal truck flows in FAF1 (FHWA, 2006). This method will be modified and used for the truck–rail flows assignment. Here, we are interested in modeling intermodal movement of freight (as the O–D flows estimated are truck–rail flows), which means that we have to ensure that while moving from an origin to a destination both modes (truck and rail) are used. It might be thought that, given this constraint, we need to identify a set of road links (a cut in the network) whose removal will ensure that the origin and destination are connected *only because of the rail link*. But this approach might not be a feasible one.

Consider the example shown in Figure 1. Here, cuts C_1 and C_2 both ensure that, when flow is assigned, the rail link gets loaded accordingly. That is when the arcs marked m or p are removed, the rail link is the only link that prevents the network from being disconnected. However, figuring out which links to remove (links belonging to C_1 or C_2) is not possible. Here, we can remove the links belonging to C_2 (marked p), but these links lead to the intermodal point and their removal will affect the movement of trucks to that point. In short, there is no way to determine that the actual links that need to be removed are the ones that belong to set C_1 (marked m).

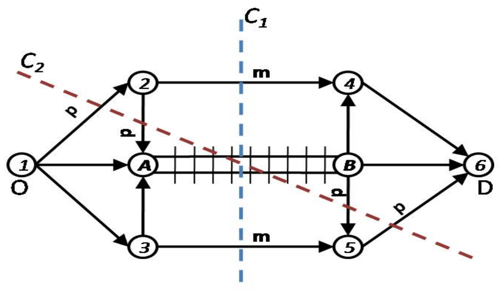

Figure 1: An example showing that cuts are non–unique

In the problem considered here, we assume that there is only one rail route from one intermodal point to another. For example, in going from Jacksonville to Seattle using rail, there can exist two rail routes: Jacksonville–Memphis–Seattle and Jacksonville–St. Louis–Seattle. But the choice of rail routes are dictated by various policy decisions and for the time being this additional constraint can be relaxed (the methodology proposed here can be trivially extended to the case where multiple rail routes between an origin–destination pair are allowed). Furthermore, it will also be assumed that all the truck–rail flow originating from a county will be assigned to the nearest rail intermodal center.

In light of these assumptions, the SUE can now be applied to the truck–rail O–D flows as follows. From a given origin we first assign all the flows to the nearest rail access point. And since it is assumed that there is only one rail route between a rail access and egress point, we then assign flows from the rail egress point to the county centroid where those flows were actually destined for. That is, given an O–D pair flow, we split that pair; first from origin county to nearest rail intermodal point and then from egress point of that rail route to the destination county. For example, consider a network as follows: $O \rightarrow R_1 \rightarrow R_2 \rightarrow D$, where O and D are the origin and destination respectively, and R_1 and R_2 are the rail intermodal nodes. Here, an SUE assignment will first be carried out from O to R_1, and then from R_2 to D.

5.2. Discerning non–unique Path Flows from Link Flows

In most commercial software, the 'select–link analysis' (Bar-Gera and Luzon, 2007) command returns the detailed split of the flows on the links. This command provides information on where a proportion of the flow on the link has originated and where it is destined for. After the assignment has been carried out, for truck–rail flows, this command will not give the desired information. In our case, the 'select–link analysis' command will only convey, for a link, where a proportion of flow has originated and to which nearest intermodal rail terminal it is destined for; or which intermodal rail terminal that flow is coming from and which county it is destined for. In short, we will be unable to discern, for a particular link, the *county* origin and destination split of flows. This is because, it can be recalled from the previous section, we modified the O–D table before the assignment stage. As a result of this, the paths (route flows) stored in the commercial software package will contain only partial information. That is, for one O–D pair (both origin and destination counties) there will be two sets of route flows. One from the origin to the nearest intermodal center and the second from the egress of that rail route to the destination county. This is the reason why the 'select–link analyses' will give only incomplete information with regards to a link.

Any commercial software or SUE code can store the paths on which it assigns flow at ever iteration and thus keep a track of the path flows. In the truck–rail flows case, these paths are either from the origin to the nearest intermodal center or from the egress of that rail route to the destination county. Two sets of most likely path flows can also be obtained by using the Maximum Entropy formulation (Bell and Iida, 1997).

We will present a methodology to find non–unique path (route) flows between an origin and destination county. We will use the Flow Decomposition Algorithm as the basic building for our proposed method.

- Step 0: Identify the intermodal rail terminal closest to the origin county. Do this for all counties. Now, given an O–D pair flow, split that pair; first from origin county to nearest rail intermodal point and then from egress point of that rail route to the destination county.

- Step 1: Perform a Stochastic User Equilibrium (SUE) assignment using the modified trip table from Step 0.

- Step 2: Find all shortest paths (in the equilibrated network) for all non–zero O–D pairs of the modified trip table. That is, we find all the shortest paths connecting an origin to its nearest rail terminal, and all shortest paths connecting a rail terminal (egress) to a destination. This is done for all origin and destination counties.

- Step 3: Store all such paths in a LIST and let the flow on each path be equal to the flow on the link that constricts flow on that path (link with least flow).

- Step 4: Execute the flow decomposition algorithm for one O–D pair at a time (This algorithm can be found in its entirety in any classical book on Network Flows, e.g. (Ahuja et. al., 1993)). Ensure that only paths and rail route related to that O–D pair are present in the network when executing the flow decomposition algorithm. All used paths are deleted from the LIST and all unused paths are maintained in the LIST.

- Step 5: Store the information for each route flow.

Note: (a) The flow on each rail link is set to be equal to the sum of all the flows originating from the counties that use that rail link. (b) If, at the assignment stage, all the paths are not recorded, then a modified version of Dijkstra's Algorithm (E. Dijkstra, 1959) can be used find all the shortest paths. This can be done because at equilibrium all paths which carry flow (between an origin and a destination) have equal and least cost. Enumerating all these paths using Dijkstra's might not be very fast. (c) The flow decomposition sub–routine will not find any cycles because we decompose flow one O–D pair at a time and explicitly define the set of paths before and after the rail route.

6. Illustrative Example

The procedure delineated above is shown on an illustrative example, the hypothetical network is shown in Figure 2.

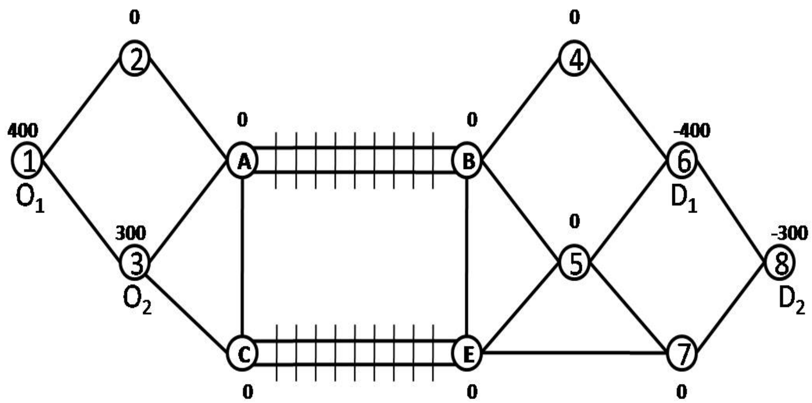

Figure 2: A hypothetical network with demands

- Step 0: The rail terminal closest to O_1 is A and the one closest to O_2 is C. The original O–D matrix and the modified O–D matrix is as show in Table 1 and Table 2 respectively.

	D_1	D_2
O_1	400	0
O_2	0	300

Table 1: The original O–D matrix

	A	C	D_1	D_2
O_1	400	0	0	0
O_2	0	300	0	0
B	0	0	400	0
E	0	0	0	300

Table 2: The modified O–D matrix

- Step 1: After performing the SUE assignment, the network and its link flows are shown in Figure 3.
- Steps 2 and 3: The shortest paths (the shortest paths connecting an origin to its nearest rail terminal, and all shortest paths connecting a rail terminal (egress) to a destination) and the flows on those paths are tabulated in Tables 3 and 4. The modified networks for both O–D pairs for the application of the flow decomposition algorithm is shown in Figure 4 and Figure 5.
- Steps 4 and 5: The flow decomposition algorithm is now applied to the networks shown in Figure 4 and Figure 5. The path flows from O_1 to D_1 and from O_2 to D_2 are shown in Table 5.

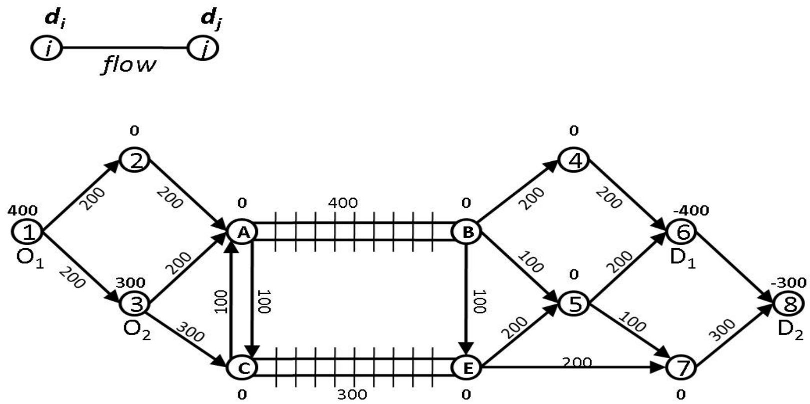

Figure 3: Link Flows after SUE

Path No.	From O_1 to A	Flow
I	O_1 – 2 – A	200
II	O_1 – 3 – C – A	100
III	O_1 – 3 – A	100
Path No.	From B to D_1	Flow
IV	B – 4 – D_1	200
V	B – 5 – D_1	100
VI	B – E – 5 – D_1	100

Table 3: Two sets paths for the first O–D pair

Path No.	From O_2 to C	Flow
VII	O_2 – C	200
VIII	O_2 – A – C	100
Path No.	From E to D_2	Flow
IX	E – 7 – D_2	200
X	E – 5 – 7 – D_2	100

Table 4: Two sets paths for the second O–D pair

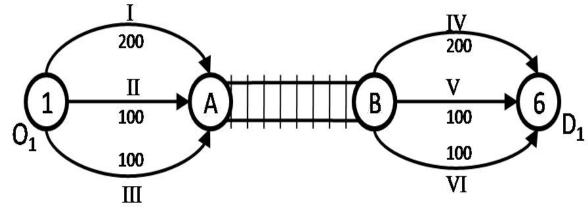

Figure 4: Network on which flow decomposition is applied (first O–D pair)

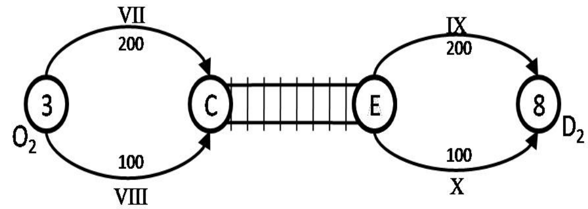

Figure 5: Network on which flow decomposition is applied (second O–D pair)

Path No.	From O_1 to D_1	Flow
(i)	O_1 – 2 – A – B – 4 – D_1	200
(ii)	O_1 – 3 – C – A – B – 5 – D_1	100
(iii)	O_1 – 3 – A – B – E – 5 – D_1	100
Path No.	From O_2 to D_2	Flow
(iv)	O_2 – C – E – 7 – D_2	200
(v)	O_2 – A – C – E – 5 – 7 – D_2	100

Table 5: Final non–unique path flows

7. Summary and Future Work

The key attractions of rail is the ability to mitigate congestion and reduce Greenhouse Gas (GHG) emissions while offering shipping prices that are not so closely tied to the price of gas. These attractions of rail as an economically viable means of goods transport have garnered the attention of investors. Therefore, it is necessary that decision makers develop a better understanding of freight movements by truck–rail and rail modes in order to prepare the appropriate policy responses to increased freight movement by rail. In this paper we have laid out a methodology that is feasible and cost effective to better understand how freight movements between origin and destination counties happen in the context of truck and rail combinations.

The next step is to test the methodology developed here with economic and commodity data from Florida which is a major consuming state and whose freight travel behavior is well understood by members of this team. The Florida DOT is at the forefront of trying to understand freight issues and develop appropriate and cost effective modal policy responses and this work effort will directly feed into those policy directions.

8. References

Ahuja, R.K, T.L Magnanti and J.B Orlin *Network Flows: Theory, Algorithms and Applications*, Prentice Hall, 1993.

Bar-Gera, H., and A. Luzon *Differences among Route Flow Solutions for the User-Equilibrium Traffic Assignment Problem*, Journal of Transportation Engineering, 133, pp.232–239, 2007.

Bell, M.G.H and Y. Iida, *Transportation Network Analysis*, J. Wiley, 1997.

Cambridge Systematics, *Development of A Computerized Method to Subdivide the FAF2 Regional Commodity OD Data to County Level OD Data*, Draft Report submitted to FHWA, 2009.

Dijkstra, E., *A note on two problems in connexion with graphs*, Numerische Mathematik, 1, pp.269–271, 1959.

FHWA, *Freight Analysis Framework Version 2.2, User Guide*, U.S. Department of Transportation, November 2006.

Sheffi, Y., *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Englewood Cliffs: Prentice-Hall, 1985.

Viswanathan, K., D. Beagan, V. Mysore, and N. Srinivasan, *Disaggregating Freight Analysis Framework Version 2 Data for Florida: Methodology and Results*, Transportation Research Record, 2049, pp.167–175, 2008.