Issues and Approaches to the Utilization of Highly Aggregated Databases in Freight Planning and Modeling

Topic Area: Models for transportation, port, air, intermodal operation; impact analysis

Gregory A. Harris, PhD., P.E.
Director, Office for Freight, Logistics & Transportation
University of Alabama in Huntsville
Shelby Center for Science & Technology, Room 145
256-824-6060
harrisg@uah.edu

Michael D. Anderson, Ph.D., P.E.
Associate Professor of Civil Engineering
Department of Civil and Environmental Engineering
The University of Alabama in Huntsville
Huntsville, AL 35899
256-824-5028
mikea@cee.uah.edu

Niles Schoening, Ph.D.
Professor of Economics
College of Business Administration
The University of Alabama in Huntsville
Huntsville, AL 35899
256-824-7314
schoenn@uah.edu

Lauren Jennings Neppel
Research Scientist
Office for Freight, Logistics & Transportation
College of Business Administration
The University of Alabama in Huntsville
Huntsville, AL 35899
256-824-6652
jenningl@uah.edu

ABSTRACT
As one of the four original pilot studies into disaggregation of the Freight Analysis Framework Version 2 program funded by the Federal Highway Administration, the research team at the University of Alabama in Huntsville (UAHuntsville) has experienced many of the issues and problems associated with the use of such a highly aggregated database. The UAHuntsville team has also experienced the frustration with attempting to analyze, plan and model freight movement without the actual data necessary for accurate results. Rather than stopping, or being satisfied with the current state of freight data and analysis, the UAHuntsville team developed approaches, processes and methodologies to utilize available data, and develop additional data sources to provide the freight information needed to model transportation systems at the local Metropolitan Planning Organization (MPO) level and at the statewide level. This paper describes the issues encountered throughout this process and the approaches that have proved beneficial and have improved the modeling performed at both levels.

INTRODUCTION
It is difficult to incorporate freight information into transportation models and plans mainly because freight data is considered proprietary by private companies and the release of that data is considered to be detrimental to the company's competitive position. Due to the difficulty in acquiring freight data, the inclusion of freight in most transportation plans and models has either been limited in scope or based upon limited sample sizes without knowledge of contents.

The use of national freight data at the local level is challenging due to the high level of aggregation. Many national freight databases aggregate information to the individual states or major communities. Most methods of utilizing freight data depend on applying proxy factors to allocate freight to the transportation system. The planning factors used in freight system analysis must be capable of describing the freight generation and attraction characteristics of the region. The use of employment as a planning factor has come under scrutiny mainly due to its inability to estimate accurately the effect of productivity improvements made by a company to increase production without increasing employment.

This research has shown that local economic data from many different sources can successfully be used to allocate freight volumes into smaller zones from the future freight traffic volumes provided by highly aggregated national databases. The output of this effort is used as input to the modeling of freight, and the integration of that freight into existing transportation planning and modeling activities at the state and local level. This has been accomplished in Alabama at the statewide and Metropolitan Planning Organization level, resulting in validated transportation models that integrate freight into the planning activity. The methodology described in this paper can easily be replicated by other states and metropolitan planning organizations.

BACKGROUND
Freight transportation is vital to the growth and economic development of a region or state. Local transportation professionals often have problems incorporating freight into transportation plans and models because freight data is proprietary at local levels and requires extensive aggregation to national levels before being released to the public. Understanding freight and the

factors affecting freight activity is important for planning infrastructure to transport freight demand and for evaluating potential investment and operational strategies. In the United States, many national freight databases aggregate information to the individual states, or major communities in the states. For example, the Freight Analysis Framework, Version 2 Database (FAF2) developed and distributed by the Federal Highway Administration (FHWA) contains freight flows for 114 zones at the national level, as shown in Figure 1. A benefit of using the FAF2 database for analysis of freight is that the database includes freight flow data for base-year, 2002, as well as forecasts for 2010 through 2035 in 5-year increments [1].

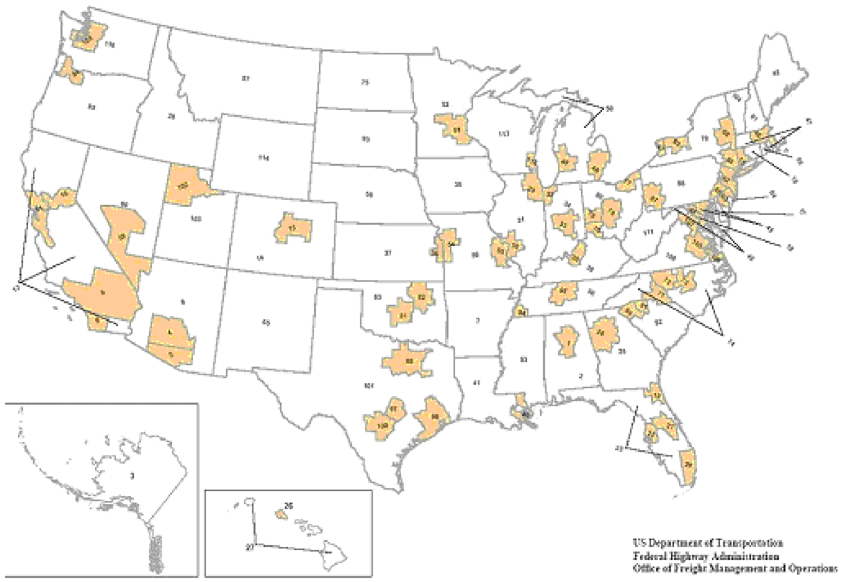

Figure 1 Geographic Locations for FAF2 data [2].

FAF2 project

The research team at the University of Alabama in Huntsville (UAHuntsville) was one of the four original pilot studies into methods for the disaggregation of the Freight Analysis Framework Version 2 program funded by the Federal Highway Administration. Until now, the disaggregation of freight from national levels for use in local areas has been primarily based on the relative employment in the local area to the total employment in the zone. This disaggregation technique has limitations in that productivity improvements which allow manufacturers to produce more products that require more freight shipments using fewer employees are often undetected [3]. The UAHuntsville research team developed a new methodology using a national freight origin/destination database and various socio-economic factors to perform disaggregation to the local level. The factors considered in this research include Value of Shipments (or Sales), Personal Income, Population, and Employment.

Lack of Data

Even as freight models have begun to emerge as valuable tools for decision analysis in policy, design and investment, the development of freight models are still far behind passenger models. The consensus of the freight and modeling communities is that one of the main reasons for this stalled development is the lack of sufficient, high-quality data [4].

This poor state of affairs in freight data can be traced to issues associated with gathering and utilizing the data. The freight system is very complex, and is characterized by the number of participants, the quantity of interactions and the proprietary nature of the data itself. Most of these interactions are unobservable to the researcher. The cost of acquiring the data, either through purchase or collection, can be extremely expensive [5]. Due to the difficulty in acquiring freight data, the inclusion of freight in most transportation plans and models has either been limited in scope or based upon limited sample sizes without knowledge of contents [6].

FAF2 Issues

Finding data to estimate freight origin and destination movements at the sub-state freight analysis zone (FAZ) level [7] is often complicated by both availability and comparability issues. The primary sources of sub-state FAZ data are economic censuses published by the Census Bureau and by state agencies. Data collected by the Census Bureau for the Census of Manufacturing and the Census of Mining are often repressed to prevent divulging information about specific companies. This problem is especially prevalent in small, rural sub-state FAZs. Researchers are forced to use estimates rather than actual figures for many manufacturing and mining operations due to this problem. Data are much more complete in the Agriculture and Retail Trade Censuses and in sub-state FAZs which include a metropolitan area. State agencies often provide reliable and complete data for economic sectors, such as logging and oil and gas drilling which are not covered by the Census Bureau's economic censuses.

Estimating sub-state FAZ freight destination movements is more difficult. Demand arising from households can be determined by utilizing the proxy of personal income or retail trade data that are routinely published by the Commerce Department's Bureau of Economic Analysis and by state Departments of Revenue respectively. Data on the movement of intermediate goods originating from manufacturing, agriculture and mining activity are not readily available from published sources and must be estimated. Adding to this problem is the growing complexity of the nation's distribution systems. Today's distribution systems often are very complicated and uniquely configured for each user. In addition, information about how these systems operate and their volumes of freight traffic are proprietary [8]. Lack of these data often leaves the researcher with a major survey task to complete the picture of destination traffic movements in a sub-state FAZ.

Besides dealing with data availability issues at the sub-state FAZ level, the researcher must also cope with comparability problems. Freight in the FAF2 Commodity Database is defined by the Standard Classification of Transported Goods (SCTG). Economic data is classified by the North American Industrial Classification System (NAICS). Of the 43 SCTG codes, only 14 have identical counterparts under the NAICS classification system. In other categories either the NAICS or the SCTG systems provide more detail and in three SCTG categories – mixed freight; waste and scrap; and other there are no NAICS counterparts [9].

Another problem is translating data reported in the economic censuses into truckloads. In some cases, as in extractive industries, the translation is easy since both the censuses and FAF2 are reported in tons. In other cases, such as poultry or logging, a translation factor must be found. The Agricultural Census reports the number of chickens sold which then requires a translation to the average number of chickens per truckload. Logs are recorded in board feet by state forestry agencies so total board feet of lumber per truckload must be calculated.

These data availability and comparability issues make the job of compiling freight origin and destination movements difficult at the sub-state FAZ level. Published and readily available data has to be supplemented by estimations and data collected by survey, but proven techniques are available which assure that the aggregate results possess an acceptable level of accuracy.

USING THE AVAILABLE DATA

Even with all of the issues associated with existing freight data, and rather than stopping, or being satisfied with the current state of freight data and analysis, the UAHuntsville team developed approaches, processes and methodologies to utilize available data, and develop additional data sources to provide the freight information needed to model transportation systems at the local Metropolitan Planning Organization and the statewide level.

Additional Data Sources

The base year for the sub-state economic database is 2002, the year corresponding to the FAF2 O-D matrices. The year 2002 is also when the US Census Bureau surveyed industries for its series of state economic censuses including the *Census of Manufacturing, the Census of Agriculture*, and the *Census of Mining* (Figure 2). The base year will change after the 2007 O-D matrices are released.

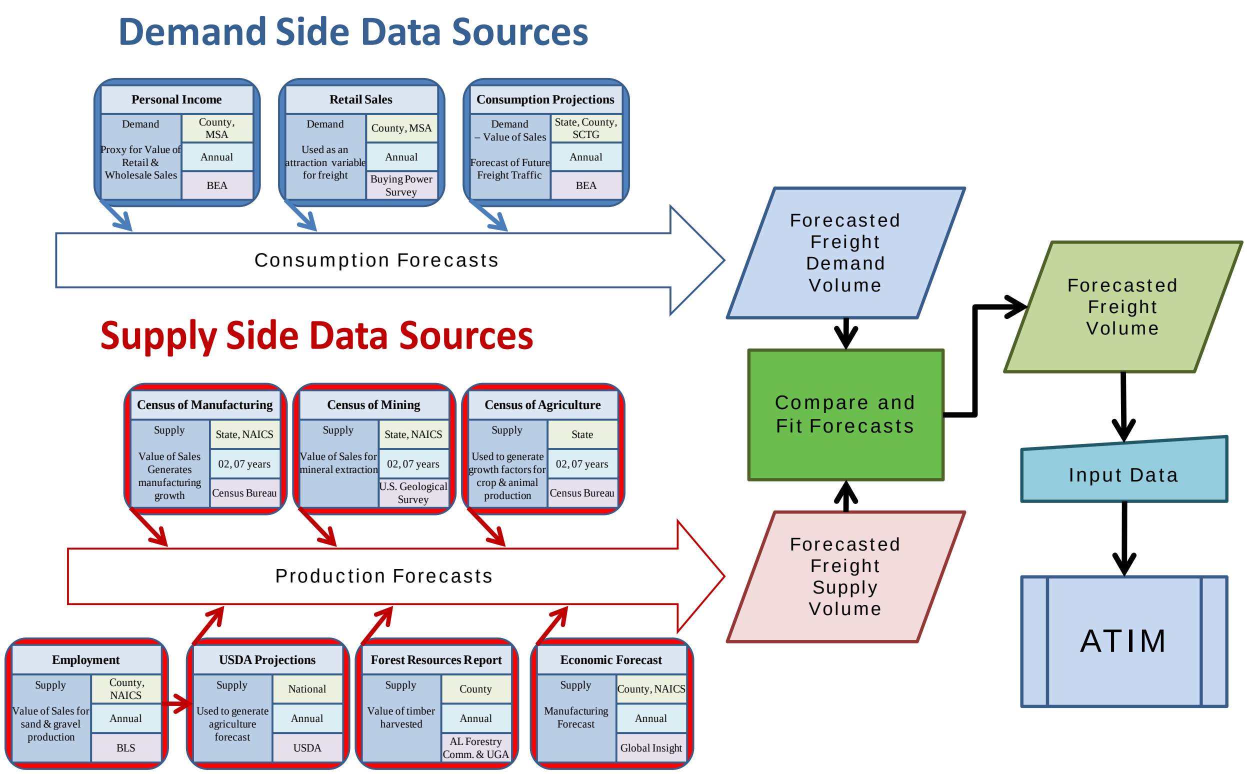

FIGURE 2 Data Sets and Interactions

Determining Variables to Use to Allocate Freight Traffic to Sub-state FAZs

Heretofore, employment growth has typically been used to generate a freight traffic forecast in a particular area. However, employment growth has been shown to be a poor predictor of freight traffic increases because it doesn't take into account productivity improvements in goods-producing industries [10]. The Value of Sales or Shipments has been shown to be a better predictor of freight generation activity. Thus, the Alabama sub-state economic database includes the value of sales from goods-producing industries (Figure 2). Using Value of Sales instead of employment factors in future productivity improvements and should provide a better forecast of future freight traffic.

Personal income was chosen to proxy the value of retail sales to households and wholesale sales to businesses in a sub-state region (Figure 2). The growth of personal income is highly correlated with the growth of household consumption expenditures and consequently should give a more accurate forecast of freight traffic than either population or employment growth.

Estimating the Value of Sales and Personal Income

The value of sales for manufacturing are published in the *Census of Manufacturing* (Figure 2) for each state, metropolitan area, and county that contains manufacturing enterprises [11]. If there are only a few manufacturers or one or two dominant firms, the value of sales data will be suppressed to protect the privacy of the firms. In Alabama the value of sales data was suppressed in 19 of 67 counties – nearly all of them sparsely populated rural counties with a single dominant company. A Value of Sales estimate must be prepared in these cases. The Census Bureau provides a range of employment for the plant(s) in these counties. Multiplying the mid-point in the employment range by the average value of sales per employee for the industry within the state provides a reasonable proxy for the actual Value of Sales in these counties. The Value of Sales in each county, including estimates that had to be made, can then be summed and compared to the actual total value of sales for the state. If the published total is larger or smaller than the total containing the estimates, the estimates, can be increased or reduced until they equal the published state total.

The *Census of Agriculture* (Figure 2) provides detailed Value of Sales data for each type of crop or animal sold from a particular county [12]. The US Geological Survey (USGS) periodically publishes a state geological survey which includes the Value of Sales for the mineral industry [13]. 2003 is the most recent USGS survey for Alabama. Production and sales data are provided by geological area rather than by county in this publication so it must be supplemented by information from the *Census of Mining* (Figure 2) to allocate the value of mineral extraction to each county in the state [14].

Sand and gravel operations are located in almost every county in Alabama and can be found using *County Business Patterns* (Figure 2) where the publication lists total employment by county [15]. Allocating sand and gravel sales by employment give an estimate of the contribution of the sand and gravel industry to total sales in each county.

The quantity of logs harvested in each county is released in an annual report (Figure 2) from the Alabama Forestry Commission [16]. The data are provided by type of log and by volume in board feet. The value of these logs was determined by translating board feet into tons and using 2002 pricing data for the South published by the Daniel B. Warnell School of Forestry Resources, University of Georgia [17]. Personal income (Figure 2) by county is released annually by the Bureau of Economic Analysis (BEA), US Department of Commerce and is a part of the Regional Economic Accounts database [18].

CONCLUSIONS

Rather than conceding the use of freight data, or being satisfied with the current state of freight data and analysis, the UAHuntsville team developed approaches, processes and methodologies to utilize available data, and develop additional data sources to provide the freight information needed to model transportation systems at the local Metropolitan Planning Organization level and at the statewide level.

This research has shown that local economic data from diverse sources can be employed to allocate freight volume into smaller FAZs from the commodity volumes provided by highly aggregated national databases, such as FAF2. The output of this effort is used as input to the modeling of freight, and the integration of that freight into existing transportation planning and modeling activities at the state and local level. This has been accomplished in Alabama at the statewide and metropolitan planning organization level, resulting in validated transportation models that integrate freight into the planning activity.

The methodology described in this paper can easily be replicated by other states and metropolitan planning organizations. Future research will focus on finding the set of economic variables that best predicts present freight movements into and out of these FAZs and consequently increase the accuracy of predicting future freight movements.

REFERENCES

1. FAF2 Internet Page. http://ops.fhwa.dot.gov/freight/freight_analysis/faf/index.htm (August 1, 2007).
2. FAF2 Areas. http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/cfs_faf_areas.htm (August 1, 2007).
3. UAH Office for Economic Development (2005), Transportation Infrastructure In Alabama - Meeting the Needs for Economic Growth, Final Report on the Requirements for Infrastructure and Transportation to Support the Transformation of the Alabama Economy. Prepared for the Office of the Secretary, U.S. Department of Transportation, Grant No. DTTS59-03-G-00008.
4. Mani, Akshay and Jolanda Prozzi. State-of-the-Practice in Freight Data: A Review of Available Freight Data in the U.S. *Project 0-4713: Development of Sources and Methods for Securing Truck Travel Data in Texas*, Center for Transportation Research, The University of Texas at Austin, February 2004.
5. Holguín-Veras, José, Ellen Thorson, and Kaan Ozbay. Preliminary Results of Experimental Economics Application to Urban Goods Modeling Research. *Transportation Research Record: Journal of the Transportation Research Board, No. 1873*, TRB, National Research Council, Washington, D.C., 2004, pp. 9–16.
6. UAH Office for Freight, Logistics & Transportation (2008), Transportation Infrastructure In Alabama - Bridging the Data & Information Gap, Final Report. Prepared for the Federal Transit Administration, U.S. Department of Transportation, Project No. AL-26-7262-01. November 20, 2008.
7. Harris, Gregory A., Phillip A. Farrington, Michael D. Anderson, Niles Schoening, and James Swain. Developing Freight Analysis Zones at a State Level: A Cluster Analysis Approach. *TRB 88th Annual Meeting Compendium of Papers CD*, Washington, D.C., January 11-15, 2009.

8. Anderson, Michael D. and Gregory A. Harris. Development of a Method to Forecast Freight Demand Arising from the Final Demand Sector and Examination of Federal Data to analyze Transportation Demand for Local Area Through Trips. Final Report for the Alabama Department of Transportation, Research Project 930-697. June 4, 2009.
9. Harris, Gregory A., Niles Schoening, Michael Anderson, Jeffrey Thompson, Robert Jilla, and Emmanuel Oranika. Development of a Freight Database for Use in Allocating Freight Traffic to Sub-State Traffic Zones. Submitted to the 89th Transportation Research Board Annual Meeting, Washington, D.C., January 2010.
10. Anderson, M., G. Harris, S. Jeereddy, S. Gholston, J. Swain and N. Schoening. Using a Federal Database and New Factors for Disaggregation of Freight to a Local Level. *Proceedings of the 10th International Conference on Application of Advanced Technologies in Transportation*, Athens, Greece, May 2008.
11. US Census Bureau, 2002 Census of Manufacturing: Alabama County Data.
12. National Agricultural Statistics Service, USDA, *2002 Census of Agriculture: Alabama County Data*, pp. 265-275.
13. US Geological Survey, *Geological Survey of Alabama*, 2003, pp. 3.1-3.5.
14. US Census Bureau, *2002 Census of Mining: Alabama.*
15. US Census Bureau, *2004 County Business Patterns: Geographic Area Series: State, County and Metropolitan Area Series: 2004 and 2003*, December, 2006.
16. Alabama Forestry Commission, *Forest Resources Report*, *2003*, pp.2-3.
17. Warnell School of Forest Resources, University of Georgia, Timber Mart – South, "South-wide Average Stumpage Prices Press Release," October, 2002.
18. Bureau of Economic Analysis, Regional Economic Accounts: Local Area Personal Income: Alabama, March, 2009.