Developing Freight Analysis Zones at a State Level: A Cluster Analysis Approach
Author(s): Gregory A. Harris, Michael D. Anderson, Phillip A. Farrington, Niles C. Schoening, James J. Swain and Nitin S. Sharma
Source: *Journal of the Transportation Research Forum*, Vol. 49, No. 1 (Spring 2010), pp. 59-68
Published by: Transportation Research Forum
Stable URL: http://www.trforum.org/journal

Developing Freight Analysis Zones at a State Level: A Cluster Analysis Approach

by Gregory A. Harris, Michael D. Anderson, Phillip A. Farrington, Niles C. Schoening, James J. Swain and Nitin S. Sharma

The ability to forecast freight to support transportation infrastructure decisions is limited by data availability at a level of detail meaningful to the transportation planner. The Freight Analysis Framework Version 2 is a national, comprehensive public freight database. The difficulty that transportation planners encounter when using this data is due to extensive aggregation. In this paper, the authors develop a methodology for creating freight analysis zones (FAZs) at a sub-state level by partitioning a state into meaningful zones that support freight transportation planning and analysis. The authors conclude that FAZs can be used effectively without degrading the quality of the forecasts.

INTRODUCTION

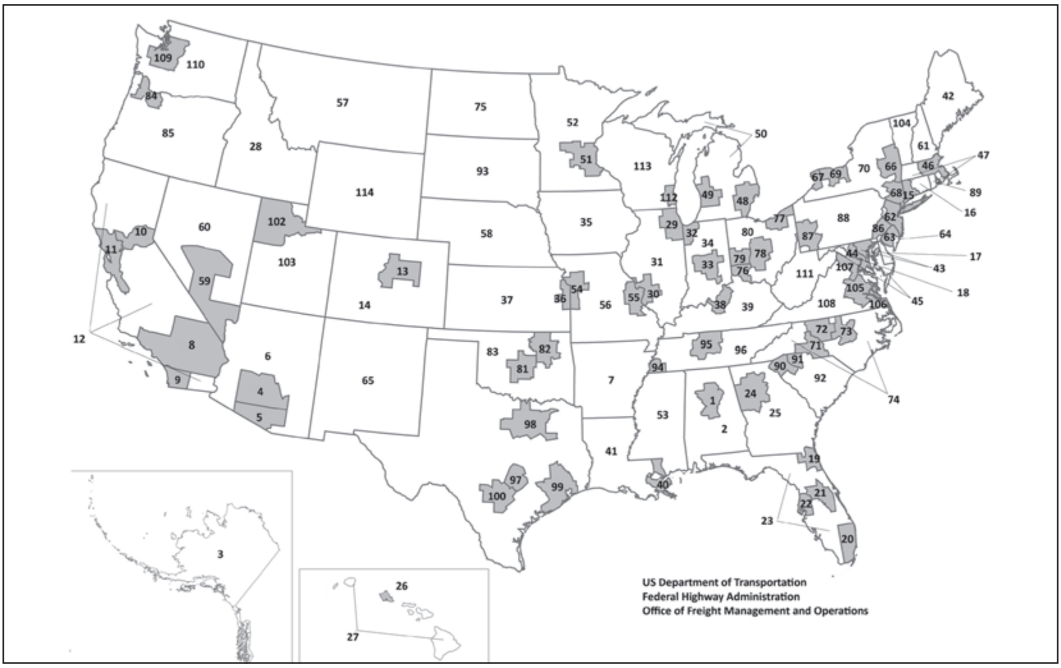
The Freight Analysis Framework 2 (FAF2) is a public freight database that has the Commodity Flow Survey as its basis (Federal Highway Administration 2007). It is aggregated at the national level with 114 origins and destinations as shown in Figure 1. In FAF2, Alabama is represented by two zones, an eight county area around Birmingham and the remaining 59 counties in the state. The high level of aggregation in this database is not conducive to analysis of the effects of freight traffic on the transportation systems at the state and local levels. As a result this data has limited use for metropolitan transportation planning. In 2006, the Federal Highway Administration (FHWA) funded four pilot projects to develop methods to disaggregate FAF2 to the county level (Tang 2006). The Office for Freight, Logistics and Transportation (OFLT) at the University of Alabama in Huntsville (UAH) was chosen to conduct one of these projects to disaggregate FAF2 to the county level within Alabama.

With Alabama only having two designated zones in FAF2, the data is too highly aggregated to provide freight flow information for local or sub-state planning purposes, thus requiring disaggregation. Although this disaggregation is possible because Alabama only has 67 counties, it could be arduous in states such as Texas and Georgia which have significantly more counties. Further, the OFLT research team believed that county-level disaggregation of FAF2 would be too detailed for most states to use for freight planning purposes and chose instead optimal origin and destination zones between two pairs and 67 pairs, the latter being the number of counties in Alabama. These zone pairs should provide the necessary level of information for freight analysis without excessive detail.

In 2002, 10 counties contributed nearly 60% of Alabama's total income and the top 20 counties accounted for three-fourths of all personal income (Office of Economic Development 2005). This situation is not significantly different from what is observed in other states. With resources for planning strained in most transportation budgets, efforts applied to freight planning for areas where insignificant economic activity exists are not a responsible use of funds. It is hypothesized that areas of low economic activity could be aggregated into regions that produce enough economic activity to justify expending resources to plan for freight there.

In this paper the authors present a cluster analysis approach to aggregating counties within a state to form sub-county regions appropriate for the disaggregation of FAF2 data. It briefly presents background on cluster analysis and the various approaches for doing so using economic

Figure 1: Geographic Locations for FAF2 Data (FHWA 2007)



Source: http://www.ops fhwa.dot.gov/freight/freight_analysis/faf/cfs_faf_areas htm

(employment, total value of shipments, personal income) and geographic (longitude, latitude, and distance from interstate) data. The authors provide an analysis of the results obtained when various levels of aggregation (individual counties and FAZs) are used and conclude that there is no significant reduction in accuracy when clustering to create FAZs in a statewide freight travel demand model.

BACKGROUND

The decision to investigate the development of FAZs emerged from research at the national level by Shin and Altman-Hall (2007). These authors suggested that it would be beneficial to increase the current number of FAZs in the FAF2 database and described several methods to do so, finally settling on approximately 400 zones, which they considered optimal by aggregating zip codes. This work triggered the idea that the socio-economic factors considered in the FHWA pilot study by Tang (2006) could be used to aggregate counties into FAZs for freight planning and analysis at a level that used single counties or combined counties depending on the level of economic activity within each of them. Apart from this initial study, there have been attempts to disaggregate FAF2 freight flow data to the county level with varied results (Viswanathan et al. 2008, Anderson and Harris 2009, Rowinski et al. 2008, Opie et al. 2009). Based upon the work of Shin and Altman-Hall (2007) and the reality of the distribution of economic activities among few counties within states, this disaggregation may not have been ideal (OED 2005), suggesting a need to examine alternative approaches.

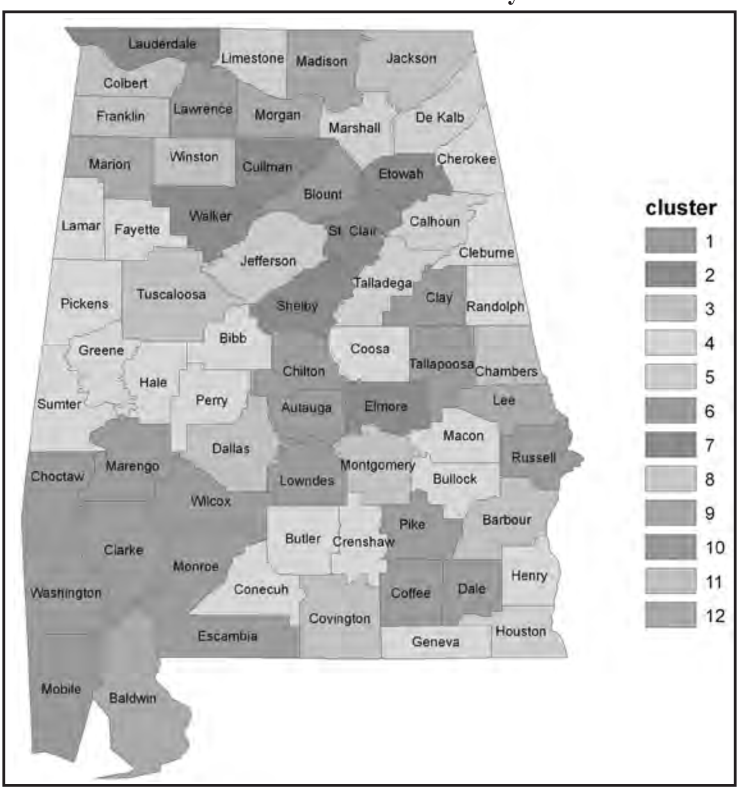
One such approach is cluster analysis. This is a multivariate technique that uses statistical procedures to form groups of entities called clusters based on certain pre-determined characteristics. Moudon et al. (2005) developed zones for metropolitan transportation planning using land use variables, such as density of activities, presence and agglomeration of destinations, block size, and transportation infrastructure attributes. COBWEB (Fisher 1987), CLUSTER/S (Stepp and Michalski 1986) and LABYRINTH (Thompson and Langley 1991) are examples of programs used for concept definition in cluster analysis.

In this research, Ward's hierarchical clustering method is used because it is effective when the intent is to minimize information loss associated with any iterative step in cluster formation (Lattin et al. 2003). The cluster analysis was performed using Minitab™ because it provides the user with a wide range of options for linkage methods and distance measures for standardized and non-standardized variable formats for entities. It also provides the user with the ability to manage the final number of clusters and options for forming clusters based on either a distance measure or similarity level. This study is unique in using zones larger than counties to disaggregate FAF2 and aggregating counties through cluster analysis. Additionally, it incorporates a modeling component to verify the results of the combined disaggregation and the cluster analysis approach.

DEVELOPMENT OF THE CLUSTERS

The FAZ development process began with identification of the basic set of economic data that can be used to define analysis zones. Data were obtained on employment, payroll, shipment value, population, and personal income for each of the 67 counties in Alabama. These data were evaluated to form clusters using Ward's method to minimize within-cluster variance (Lattin et al. 2003) and Euclidean distances between clusters were used for aggregation. Also, a variety of options for FAZ development were considered, ultimately settling on clustering counties based on economic data and eight potential solutions. Each possible solution utilized economic data and resulted in several clusters that while similar were often widely dispersed geographically, a result not conducive to effective freight planning and analysis. For example, Figure 2 shows 12 clusters based on population, value of shipments and personal income. From this figure it is clear that, without including proximity measures, the clusters contain counties that are geographically dispersed to a large extent. To resolve this problem the OFLT research team included one or more measures of geographic proximity in the clustering process.

Figure 2: Cluster Solution of Counties Based Only on Economic Variables



It was also noticed that some of the clusters often crossed interstate roadways, which is not suitable when defining traffic analysis zones. Therefore, the state was divided into zones before clustering to ensure that significant roadways were not contained in individual zones but served as their boundaries. The final solution builds clusters of counties within broad regions defined by the interstate highways traversing Alabama. This approach has several attractive features. Specifically, interstate highways provide natural traffic boundaries and meet the objective of picking up significant traffic flow on them and between zones for freight planning activities. Using these highways as boundaries, the state was divided into six planning sectors. Counties were then allocated to each sector based on their proximity to interstate highways. Although interstate highways are used as sector boundaries in this paper, in other states the boundaries may be other transportation systems such as railroads or waterways.

Using the six largest sectors created with interstate highways as boundaries, a solution was obtained that used some of the economic characteristics of a county and location information associated with the county's economic center where major employment is concentrated as clustering variables. These variables include the center's longitude and latitude and distances from interstate highways. The latter variable is important because counties closer to interstate highways appear to have more economic activities and freight traffic than those farther away from interstate roadways (OED 2005). The solution in Figure 4 is based on economic and proximity variables and distance from interstate highways and shows 34 clusters. An analysis of this solution shows that interstate sectors 3, 4, 5, and 6 have few counties. As a result, this solution was modified by combining sectors 3 and 4, and 5 and 6 thereby creating a total of four interstate sectors, which are fewer but homogenous. Figure 5 shows the cluster solution in the four interstate sectors based on economic and proximity data and each county center's distance from an interstate highway. This solution shows 27 clusters and has the most promise because the clusters are in close proximity to the natural boundaries provided by the interstate highways traversing Alabama. After the cluster analysis, the 27 clusters were evaluated based upon the type and growth of industry in each cluster to validate and refine the solution.

Figure 3: Interstate Based Sectors for Alabama

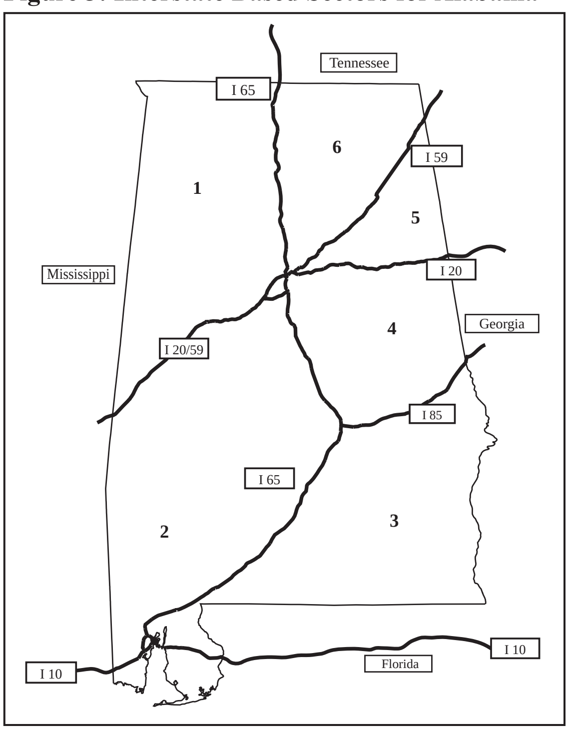


Figure 4: Cluster Solution within Interstate Sectors Based on Economic Variables, Longitude, Latitude and Distance from Interstate

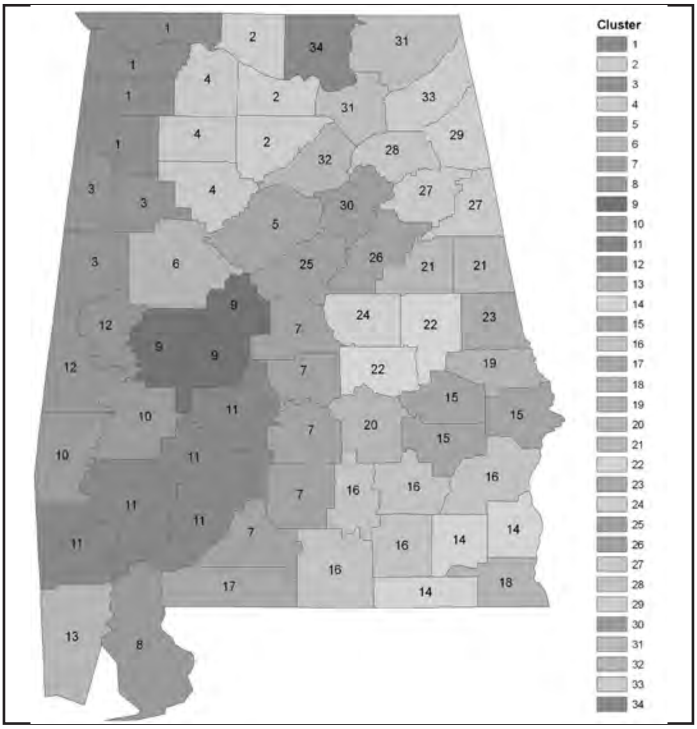
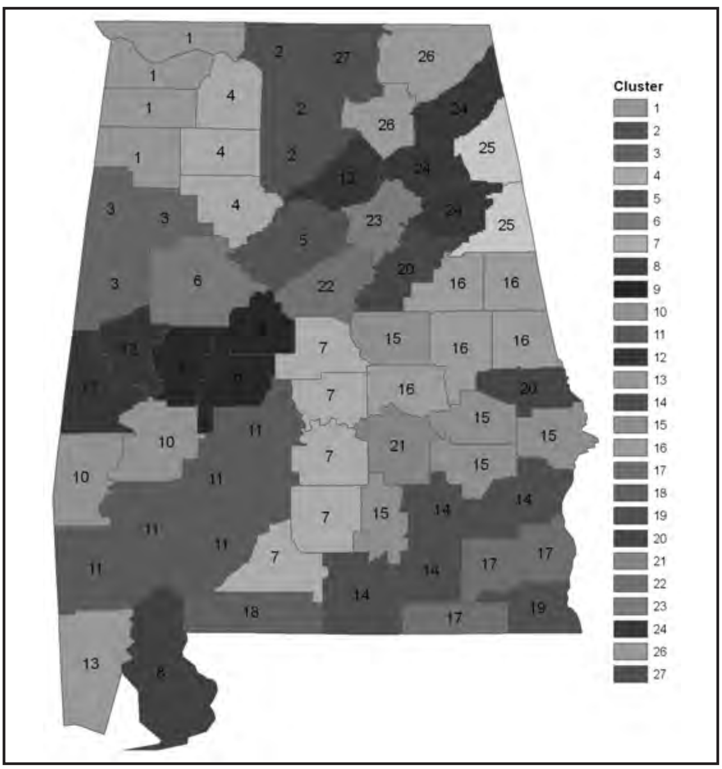


Figure 5: Cluster Solution within Modified Interstate Sectors Based on Economic Variables, Longitude, Latitude and Distance from Interstate Highway



COMPARISION OF FINAL CLUSTER SOLUTION

The differences between the FAZ approach and the county approach were analyzed by the development of a case study employing the Alabama Freight Model developed by UAH researchers (Office of Economic Development 2005). The OFLT team's first effort was to create a 67 county input file utilizing direct proportional disaggregation of the FAF2 data using population, total employment, value of shipment and average personal income within each county. By aggregating the various county data into the clustered zones that contributed to each FAZ, a 27 FAZ input file was created. These aggregated trips were subsequently assigned to the specific county that represented the economic center of the zone. Accordingly, the economic county center became the origin or destination for the FAZ. A freight distribution and assignment model was then used to calculate truck trip interchange and determine the trucks forecasted for each section of roadway in the state. The truck trip interchange was developed through the application of a gravity model that had as its inputs truck production and attraction into and out of the zones based on the disaggregation of the FAF2 data. The assignment of truck trips onto the road network was based on an all-or-nothing assignment procedure where all trips take the shortest travel path from origin to destination. The road network consists of approximately 5,000 miles of roadway in Alabama along with 15 roadways that serve as connections to surrounding states, with 250 nodes and 660 links (See Figure 6**).** Figure 7 shows the traffic on the network after the assignment.

To compare the performance of the two approaches (i.e., 67 counties versus 27 FAZs), a series of Alabama Department of Transportation (ALDOT) truck counts were added to the attributes of the roadway segments. The ALDOT values for all roadway segments where truck volume exceeds 1,000 trucks per day are shown in Figure 8. Figures 9 and 10 are scatter plots showing the relationship between the model's truck assignment results for all counties and the 27 FAZs and ALDOT's truck counts respectively.

To measure the difference between the truck assignment results from the two input levels (all 67 counties or the 27 FAZs) and actual truck traffic counts, the correlation coefficients between the truck assignment results and actual truck traffic data were calculated. The model with 67 counties resulted in a correlation coefficient of 0.507, and 0.494 for the model with 27 FAZs. Comparatively, the percent root mean square error between the assignment results and actual truck counts is 92.7 for the model with 67 counties and 95.2 for the model with 27 FAZs, demonstrating that the FAZ model is essentially as accurate as the county model. Finally, the Nash Sutcliffe (NS) coefficient (Nash and Sutcliff 1970) is 0.689 for the model that used all 67 counties and a 0.679 for the model with 27 FAZs. By utilizing the NS coefficient, a measure of fit of the model's results to actual data values is obtained, where a value of 1.0 indicates perfect alignment between model results and actual counts and a value of 0.0 indicates that an average of actual counts would provide similar results, and a negative value indicates that the model predicts at a level below the average of the actual values (Nash and Sutcliff 1970). After a review of all the statistics calculated for the comparison, the results demonstrate that while the statewide freight flow model is not perfect, its accuracy does not degrade when using the FAZ approach versus using all of the counties in the state.

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this paper is to outline a methodology for developing freight analysis zones at a state level. This is because the ability to plan and forecast freight demand for transportation infrastructure is limited by data availability at a level of detail that is meaningful to the transportation planner. FAF2 provides publicly available freight data for planning purposes. However, with 114 zones nationwide (and most states having two zones or less), the ability of a state or metropolitan planning organization to use it is significantly restricted. Consequently a disaggregation of the data to a more detailed level is needed. But, the fundamental problem is how to disaggregate the data to a usable level without reducing data quality and causing errors. The initial use of counties as the

Figure 6: Network for the Alabama Distribution and Assignment Model

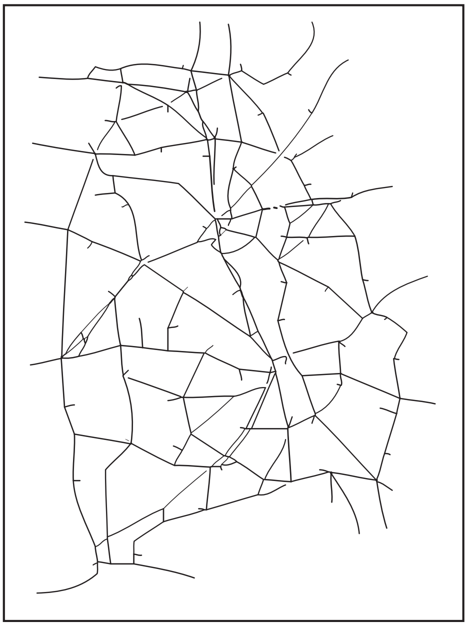


Figure 7: Network Assignment with Line Thickness Proportional to Assigned Volume

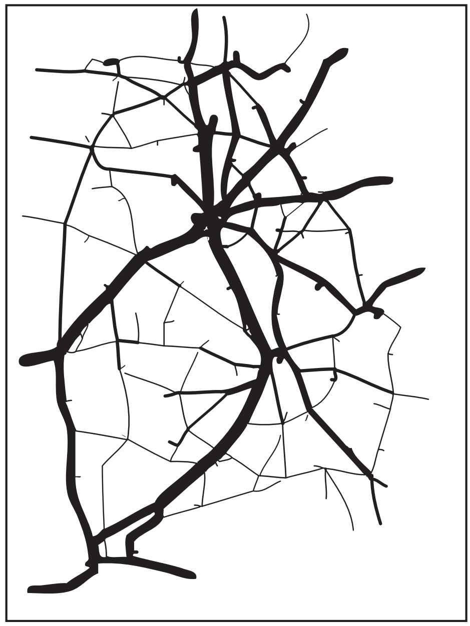


Figure 8: Location of ALDOT Truck Counts that Exceed 1,000 Trucks Per Day

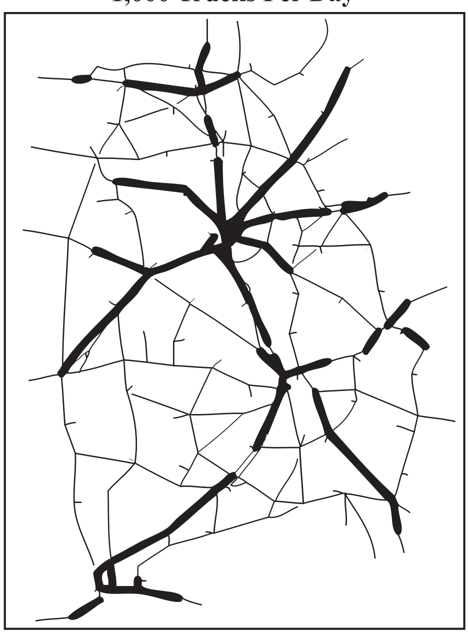


Figure 9: Scatter Plot for the 67 County Model

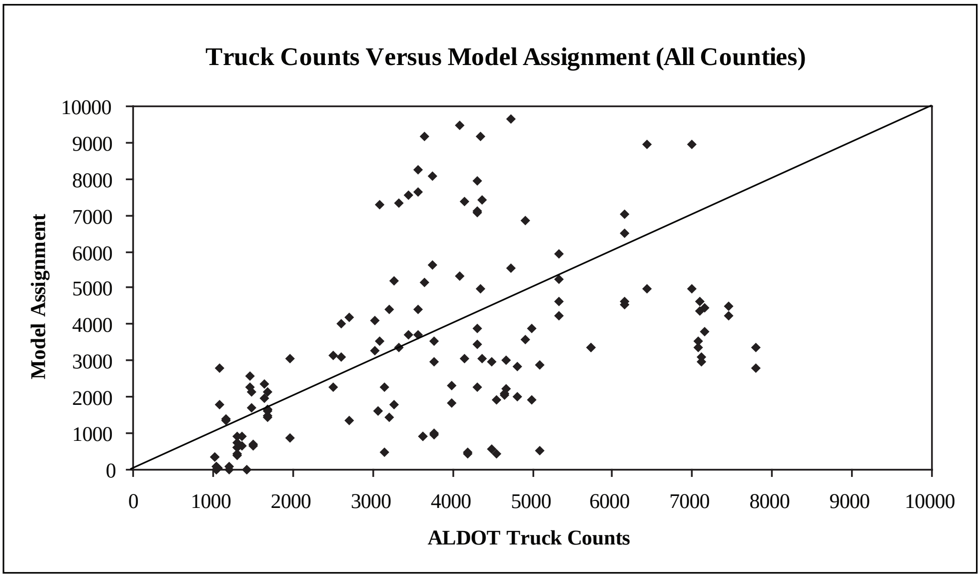
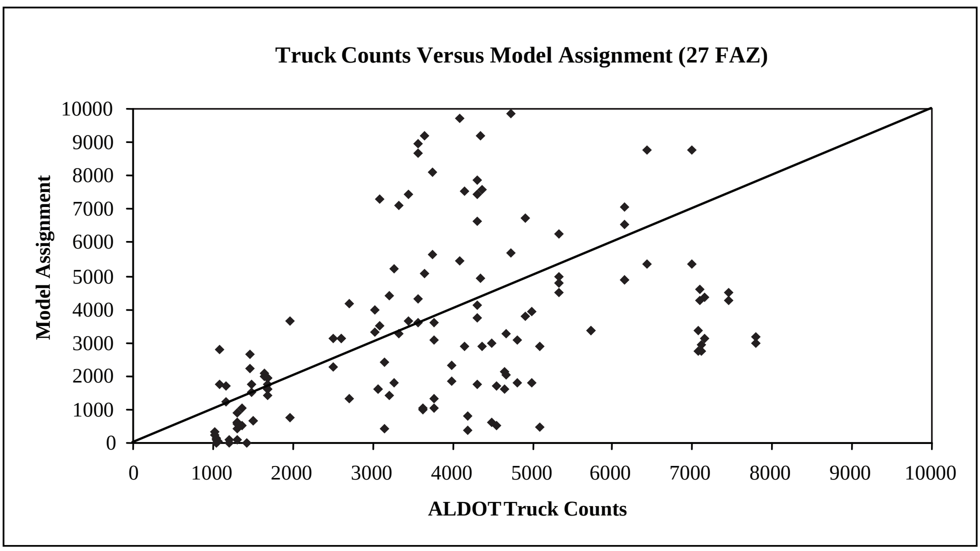


Figure 10: Scatter Plot for the 27 FAZ Model



level of disaggregation appears promising and has easy initial understanding until the number of counties creates a data matrix that becomes excessively large and unwieldy. The research examined in this paper supports the conclusion that FAZs can serve as an appropriate disaggregation level, providing the ability to utilize the FAF2 freight flow data in a statewide model without creating an unmanageable database.

Future research into the concepts of FAZs needs to continue through the examination of freight data disaggregation methods and travel model results. The various methodologies to disaggregate freight to the FAZs will help identify the impact of using these larger measurement units and the modeling of freight data will provide a mechanism to validate the various FAZs options.

References

Anderson, M. and G. Harris. "Developing Validated Freight Transportation Models Utilizing Highly Aggregated Data."*3rd Metrans National Urban Freight Conference Proceedings*, 3rd National Urban Freight Conference, Long Beach, CA. October 21-23, 2009.

Federal Highway Administration. Freight Analysis Framework Documentation. November 7, 2007. http://www.ops fhwa.dot.gov/freight/freight_analysis/faf/index.htm

Fisher, D.H. "Knowledge Acquisition via Incremental Conceptual Clustering." *Machine Learning* 2 (2), (1987): 139-17.

Lattin, J., J.D. Carroll, and P.E. Green. *Analyzing Multivariate Data*. Brooks/Cole – Thomson Learning, Pacific Grove, CA, 2003.

Moudon, A.V., S.E. Kavage, J.E. Mabry, and D.W. Sohn. "A Transportation-Efficient Land Use Mapping Index." *Transportation Research Record: Journal of the Transportation Research Board,* 1902, (2005).

Nash, J. E. and J.V. Sutcliffe. "River Flow Forecasting Through Conceptual Models. Part I: A Discussion of Principles." *Journal of Hydrology* 10, (1970): 282-290.

Office for Economic Development. *Transportation Infrastructure in Alabama: Meeting the Needs for Economic Growth.* The University of Alabama in Huntsville, 2005 Report to the U.S. Department of Transportation, 2005. Grant No. DTTS59-03-G-00008.

Opie, K., J. Rowinski, and L. N. Spasovic. "Commodity Specific Disaggregation of the 2002 FAF2 Data to the County Level for New Jersey." In the *Proceedings of the TRB 88th Annual Meeting Compendium of Papers DVD*, Paper No. 09-3754, Transportation Research Board Annual Conference, January 2009.

Rowinski, J., K. Opie, and L. N. Spasovic. "Development of a Method to Disaggregate the 2002 FAF2 Data Down to the County Level for New Jersey." In the Proceedings of the *TRB 87th Annual Meeting Compendium of Papers DVD*, Paper No. 08-0682, Transportation Research Board Annual Conference, January 2008.

Shin, H. and L. Aultman-Hall. "Development of Nation-Wide Freight Analysis Zones." In the Proceedings *of the TRB 86th Annual Meeting Compendium of Papers CD-ROM*, Paper No. 07-2289, Transportation Research Board Annual Conference, 2007.

Stepp, R. and R. Michalski. "Conceptual Clustering of Structured Objects: A Goal-Oriented Approach." *Artificial Intelligence* 28, (1986):43-69.

Tang, T. *Federal Highway Administration, FAF2 Pilot Project – Utilization of FAF2 Data by State and Local Governmental Agencies***.** February 28, 2006.

Thompson, K. and P. Langley. *Concept Formation in Structured Domains*. Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann Publishers Inc. San Francisco, CA (1991): 127-161.

Viswanathan, K., D. Beagan, V. Mysore, and N. Srinivasan. "Disaggregating FAF^2 Data for Florida: Methodology and Results." *Transportation Research Record: Journal of the Transportation Research Board* 2049, (2008): 167-175.

Acknowledgements

This research was sponsored in part by the U.S. Department of Transportation, Federal Transit Administration, Project No. AL-26-7262-00; the Federal Highway Administration, Project No. DTFH61-07-G-00007; and the Alabama Department of Transportation, Bureau of Research and Development, Research Project 930-682.

The authors thank the three anonymous reviewers for the helpful comments and suggestions on earlier versions of this manuscript. The authors are responsible for any remaining omissions or errors.

Gregory A. Harris*, P.E., is the director of the Office for Freight, Logistics and Transportation at the University of Alabama in Huntsville. His specialty is freight logistics and supply chain management. He received his Ph.D. in industrial and systems engineering from the University of Alabama in Huntsville, an M.B.A. from St. Edwards University, and a B.S. in industrial engineering from Auburn University.*

Michael D. Anderson*, P.E., is an associate professor of civil engineering at the University of Alabama in Huntsville specializing in transportation engineering, travel modeling, freight and public transportation systems. He received his Ph.D., M.S., and B.S. degrees in civil engineering from Iowa State University.*

Phillip A. Farrington *is a professor of industrial and systems engineering at the University of Alabama in Huntsville specializing in quality engineering and innovative product/process design. He received his Ph.D. in industrial engineering from Oklahoma State University and his M.S. and B.S. degrees in industrial engineering from the University of Missouri.*

Niles C. Schoening *is an emeritus professor of economics at the University of Alabama in Huntsville specializing in regional economics. He received his Ph.D. in economics from the University of Tennessee, an M.S. from Ohio State University and B.S. degree from Columbia University.*

James J. Swain *is a professor of industrial and systems engineering at the University of Alabama in Huntsville specializing in simulation modeling and applied statistics. He received his Ph.D. in industrial engineering from Purdue University, an M.S. in mechanical engineering and a B.S. in engineering science from the University of Notre Dame.*

Nitin S. Sharma *is a graduate student in industrial and systems engineering at the University of Alabama in Huntsville.*