GEORGIA DOT RESEARCH PROJECT 13-27

FINAL REPORT

Freight Movement, Port Facilities, and Economic Competitiveness – Supplemental Task: County-to-County Freight Movement (National and State Level)

OFFICE OF RESEARCH
15 KENNEDY DRIVE
FOREST PARK, GA 30297-2534

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.: FHWA-GA-16-1327	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: **Freight Movement, Port Facilities, and Economic Competitiveness – Supplemental Task : County-to-County Freight Movement (National and State Level)**	5. Report Date: April 2016	
	6. Performing Organization Code:	
7. Author(s): Dr. Catherine L. Ross, Dr. Amit Kumar, Fangru Wang, and Peter Hylton	8. Performing Organ. Report No.: 13-27	
9. Performing Organization Name and Address: Georgia Tech Research Corporation College of Architecture Center for Quality Growth and Regional Development (CQGRD) 760 Spring Street, Suite 213 Atlanta, GA 30332-0790	10. Work Unit No.:	
	11. Contract or Grant No.: 02-141, RES PROJ 13-27	
12. Sponsoring Agency Name and Address: Georgia Department of Transportation Office of Research 15 Kennedy Drive Forest Park, GA 30297-2534	13. Type of Report and Period Covered: Draft Final; December 2013 – June 2016	
	14. Sponsoring Agency Code:	
15. Supplementary Notes: Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.		

16. Abstract: While commodity movement data is available from several sources, notably from the U.S. Department of Transportation's Freight Analysis Framework (FAF), origins and destinations are heavily spatially aggregated into FAF zones, which comprise multiple counties. This makes it difficult to use FAF data for freight studies below the national level, including at the megaregion level, whose boundaries differ from FAF zones. This study disaggregates FAF3 data from the FAF zone level to the county and TAZ levels. The study uses data from multiple sources for this purpose that includes economic and sociodemographic data from the 5-year American Community Survey and the Decennial Census, transportation network locations from the U.S. Department of Transportation, tonnage-truck conversion from the Southern California Association of Governments, and geospatial boundaries from a variety of sources. The majority of these data have the advantage of being publicly available. The study employs regression analysis to establish relationships between commodity or truck trip productions and attractions, and logically related variables. These relationships are determined at the known level using ordinary Least Square (OLS) regression or spatial regression models depending on the level of autocorrelation. The regression results are then applied to the disaggregation level to obtain each area's share of known productions and attractions. The study generates O-D tables (available as a separate file) explaining the number of truck trips between US counties and TAZs of Georgia.

17. Key Words: Spatial Regression Truck models, County-to-County Truck Demand, Disaggregated Truck Demand		18. Distribution Statement:	
19. Security Classification (of this report): Unclassified	20. Security Classification (of this page): Unclassified	21. Number of Pages: 62	22. Price:

Form DOT 1700.7 (8-69)

GDOT Research Project No. 13-27

Final Report

FREIGHT MOVEMENT, PORT FACILITIES, AND ECONOMIC COMPETITIVENESS – SUPPLEMENTAL TASK: COUNTY-TO-COUNTY FREIGHT MOVEMENT (NATIONAL AND STATE LEVEL)

By
Dr. Catherine L. Ross (Principal Investigator)
College of Design
Center for Quality Growth and Regional Development (CQGRD)
Georgia Tech Research Corporation

Contract with

Georgia Department of Transportation

In cooperation with

U.S. Department of Transportation
Federal Highway Administration

July, 2016

Freight Movement, Port Facilities, and Economic Competitiveness – Supplemental Task: County-to-County Freight Movement (National and State Level)

Table of Contents

About the Center for Quality Growth and Regional Development

The Center for Quality Growth and Regional Development (CQGRD) is an applied research center of the Georgia Institute of Technology. The Center serves communities by producing, disseminating, and helping to implement new ideas and technologies that improve the theory and practice of quality growth.

For more information visit www.cqgrd.gatech.edu

LIST OF FIGURES

LIST OF TABLES

Abbreviations

ACS – American Community Survey

CBP – County Business Patterns

DOT – Departments of Transportation

FAF – Freight Analysis Framework

FAF3 – Freight Analysis Framework version 3

FAF4 – Freight Analysis Framework version 4

FHWA – Federal Highway Administration

GDOT – Georgia Department of Transportation

GPS – Global Positioning System

i.i.d. – Independent and Identically Distributed

LRTP – Long-Range Transportation Plans

MPO – Metropolitan Planning Organization

NAICS – North American Industrial Classification System

NCFRP – National Cooperative Freight Research Program

O-D – Origin-Destination

OLS – Ordinary Least Squares

SCTG – Standard Classification of Transported Goods

TAZ – Traffic Analysis Zone

TIP – Transportation Improvement Program

EXECUTIVE SUMMARY

Economic globalization has caused many changes in domestic and international freight movements, particularly as its pace has accelerated over the past several decades. Formerly, people consumed what was grown or produced close to home, and producers sold mostly on local and regional markets. Globalization has extended production networks ever farther over space and called upon freight transportation to interlink network nodes. Freight matters to economic wellbeing allowing local consumers to access global goods, and conversely in allowing local producers to link with globalized production networks and reach far away markets. However, reliable freight movement depends on the construction and management of appropriate transportation infrastructure across modes, which raises the need for freight studies and detailed freight data. Proper freight planning based on dependable analysis and forecasting can lead policymakers to develop suitable strategies for infrastructure investment to strengthen state economic competitiveness. Freight modeling and forecasting can serve as a basis for private and public sector decision making for freight infrastructure and other long-term investments.

Although Federal law requires state Departments of Transportation (state DOTs) and metropolitan planning organizations (MPOs) to account for freight in their long-range transportation plans (LRTP), transportation improvement programs (TIP), and annual work elements, in reality many MPOs and DOTs have faced difficulties in practice mainly due to a lack of data and appropriate models (Federal Highway Administration, 2007). Regional freight models, which are typically constructed with coarse levels of freight data, are unable to accurately forecast the impacts of freight on transportation systems, thus limiting the possibilities for policies and improvements to solve expected problems.

This study helps to incorporate freight movement in general and commercial trucks in particular into the transportation planning process by (1) constructing a nationwide county-level origin-destination (O-D) truck trip database; and (2) constructing a traffic analysis zone (TAZ)-level O-D truck trip database. The databases employ freight movement data from the U.S. Department of Transportation's Freight Analysis Framework version 3 (FAF3), which is disaggregated to county and TAZs levels based on regression methods using economic, sociodemographic, infrastructure, and transportation network variables that drive freight movement. The two databases will provide detailed Georgia Department of Transportation (GDOT) freight data and establish a methodology that can be used to reliably disaggregate freight movement data.

ACKNOWLEDGEMENTS

The research presented in the following report was sponsored by Georgia Department of Transportation through Research Project Number 13-27. The authors would like to acknowledge Yusuf H. Ahmed, Transportation Engineering Associate, Georgia DOT for his assistance in completing this project.

The authors would also like to acknowledge the assistance from researchers and graduate students at the Center for Quality Growth and Regional Development (CQGRD) at Georgia Tech. In particular, we are thankful to Dr. Timothy F. Welch, Assistant Director, CQGRD, for providing his guidance that helped us to complete the project and achieve the desired objectives.

The opinions and conclusions expressed herein are those of authors and do not represent the opinions, conclusions, policies, standards or specifications of the Georgia Department of Transportation or of the cooperating organizations.

CHAPTER I. INTRODUCTION

Background

From 2012 to 2014, researchers at the Center for Quality Growth and Regional Development (CQGRD) at the Georgia Institute of Technology undertook a study in partnership with the Georgia Department of Transportation (GDOT) to assess the impact of the Panama Canal expansion on Georgia's ports and roads. The canal expansion allows larger ships to traverse the canal, and is expected to both lower costs for trade between the U.S. east coast and East Asia, and increase freight volumes correspondingly. This study also analyzed likely impacts on three east and gulf coast ports (New Orleans, Savanah, and Norfolk) and their associated megaregions (Texas Triangle, Piedmont Atlantic, and DC-Virginia) by using truck movements observed with global positioning system (GPS) devices to understand truck movements from ports.

This GPS-enabled method proved to be a feasible, economical, and data-rich means to understand truck movement. However, it is not suited for every purpose since it concentrates on port-related traffic. Nationwide analyses or all-purpose freight studies will need to supplement GPS data with other information.

While commodity movement data is available from several sources, notably the U.S. Department of Transportation's Freight Analysis Framework (FAF), origins and destinations are heavily spatially aggregated into FAF zones, which comprise multiple counties. This makes it difficult to use FAF data for freight studies below the national level, including at the megaregion level, whose boundaries differ from FAF zones. However, there is an established track record of methods in literature to disaggregate national commodity movement data to lower levels and convert it to on-the-road truck flows, but these methods suffer from several gaps since they are not directly applicable below the county level, ignore spatially correlated errors, and sometimes use illogical independent variables in regression-based disaggregation methods.

This study provides a more robust commodity-to-truck disaggregation and conversion methodology amenable to county and sub-county geographic units. The proposed methodology is simple to understand and easy to implement. The contributions of this research are important from both theoretical and practical perspectives. From a theoretical point of view, the study sets out a generic method to disaggregate the FAF3 commodity flow data, first to the county level and then to the level of use to a state DOT. Further, the study employs spatial regression and nonlinearity transformation techniques to improve the currently identified relationships between commodity productions/attractions with employment. In addition, the study provides detailed documentation of the disaggregation process and easily replicable and widely applicable techniques for overcoming implementation challenges.

The strongest methods to disaggregate national data conceptually relates to the underlying economic processes motivating freight movement. Freight is a derived demand from the need to move goods to facilitate economic activity, and these derived demand characteristics mean that the strongest methods to forecast changes in freight movement patterns must account for the dynamics and economics driving freight. This study provides local freight movement estimates based on freight's relationships with underlying economic phenomena.

Objectives and Scope

The objective of this study is to disaggregate FAF level commodity flow data to create nationwide truck movement estimates at the county level (task 1) and the level of GDOT traffic analysis zones (TAZs) (task 2). The study employs a series of data sources, including the FAF version 3 (FAF3) commodity flow data, County Business Pattern (CBP) data, and U.S. Census data to estimate the relationship between the commodity and truck flows with local employment, network, and socio-demographic characteristics. The relationships established at higher spatial levels forms the basis for disaggregation to the more precise levels of counties and TAZs. The disaggregated commodity flow data thus allows for further estimation of truck trips, which will be useful for both regional and local level freight demand forecasting.

Specifically, the study comprises two main tasks outlined below:

Task 1: Construct a nationwide county-level truck origin-destination (O-D) trip database.

> The Federal Highway Administration (FHWA) provides commodity flows across the entire United States through its FAF data. The FAF zones have a spatial resolution based on the 123 zones of the Commodity Flow Survey. These zones delineate metropolitan areas and nest within states, but the geographical scale of the FAF zones is too large for regional or state freight flow analysis since each FAF zone may comprise multiple counties. Moreover, the FAF zones' boundaries do not conform to megaregion boundaries, which limits our ability to forecast and analyze freight at this emerging scale. Disaggregating the FAF zone commodity flows to individual counties and constructing a nationwide county-level commodity flow database, with the corresponding truck trip flows, will help fill these gaps. The deliverable of task 1 is a nationwide county-level truck trip O-D table.

Task 2: Construct a statewide TAZ-level truck O-D trip database for Georgia

Although the county-level truck trip database improves freight data's spatial specificity, county-level data are still too coarse for freight demand modeling within a metropolitan area or a state, where freight demand is often modeled with TAZs. Task 2 employs the result of task 1 and further disaggregates the county-level flows to GDOT TAZs. The deliverable of task 2 is a TAZ-level truck trip O-D table.

Research Significance

The results of this study will allow for useful transportation analysis at multiple scales, including travel demand models, forecasts, and performance measures. This study also establishes a methodology that can be applied to other publicly available transportation information to derive related datasets. For example, the Freight Analysis Framework (FAF) provides freight forecasts for years through 2040 for many commodity types and movement among multi-county FAF zones. The method used in this study for observed FAF data can also be used for forecasts or future FAF releases, including the Freight Analysis Framework version 4 (FAF4), which is being made progressively available over the 2015 calendar year. Finally, this method provides a reasonably simple low-cost yet high-quality approach for computing disaggregated freight movement data. It is a low-cost option because all of the data sources used are publicly available and are mostly collected and distributed by federal agencies. It is high-quality not only because the methods are benchmarked to industry best practices gathered from leading publications (NCFRP, 2013), but also because the initial data used in the analysis is regarded as a gold standard in the transportation research industry.

Report Organization

This report has five chapters. Chapter 1 includes background information. The second chapter describes the baseline conditions in the study area. The third chapter details the methodology used to disaggregate commodity movement data. The fourth chapter presents the results and explains data characteristics and format. The final chapter provides a summary of the report and identifies limitations and future extensions of this study.

CHAPTER II. STUDY AREA

Nationwide Counties in the US

There are two different study areas corresponding with the two tasks of this study. The first study area is nationwide, which refers to all counties or equivalent units in the 50 American states and the District of Columbia, and this area is defined by FAF3's geographic coverage. The 2007 county definitions were selected for alignment with the year that the FAF3 data was produced. Figure 1 presents the study area of task 1, showing the delineation of FAF zones and counties.

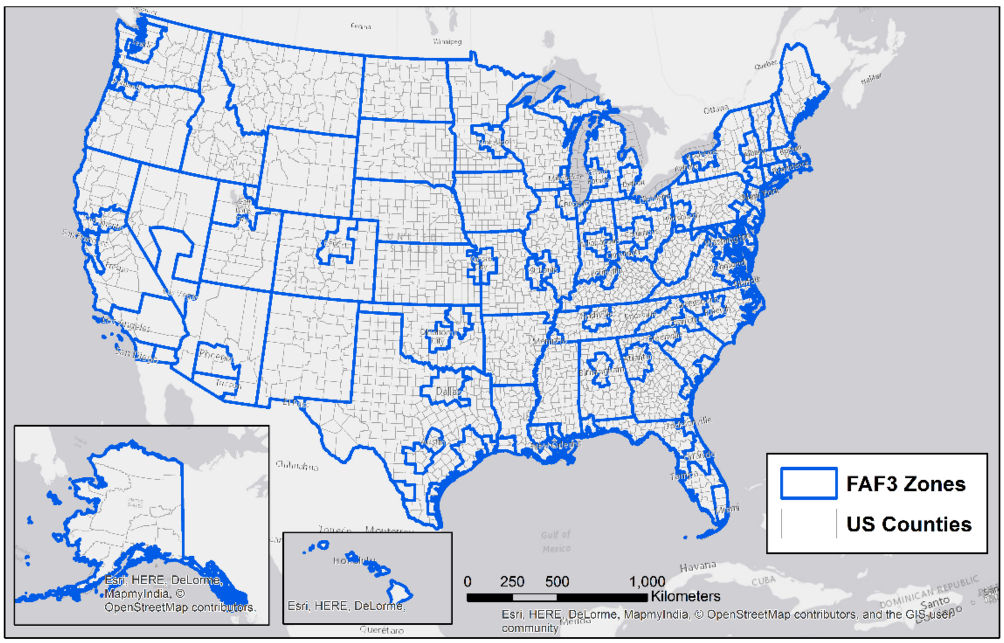

Figure 1. Task 1 Study Area.

Traffic Analysis Zones (TAZs) in Georgia

The second study area is the contiguous United States, which is encompassed in GDOT's TAZs. The study only focuses on deriving truck trips for TAZs that are delineated in the Georgia Department of Transportation's (GDOT) travel demand model, which cover the entire contiguous United States, albeit with highly variable sizes. In six southeastern states (Alabama, Georgia, Florida, North Carolina, South Carolina, and Tennessee), TAZs tend to be smaller than counties, which requires disaggregation of county-level truck flows. This fine-grained study area corresponds with the geographic area over which GDOT has prime responsibility and to which its other analytical and forecasting models are tuned. However, elsewhere TAZs correspond with state boundaries, meaning that county-level flow data are aggregated to the state level. Figure 2 presents the complete study area for task 2 and illustrates the highly variable size of TAZs. The details of the areas where TAZs are smallest are presented in Figure 3.

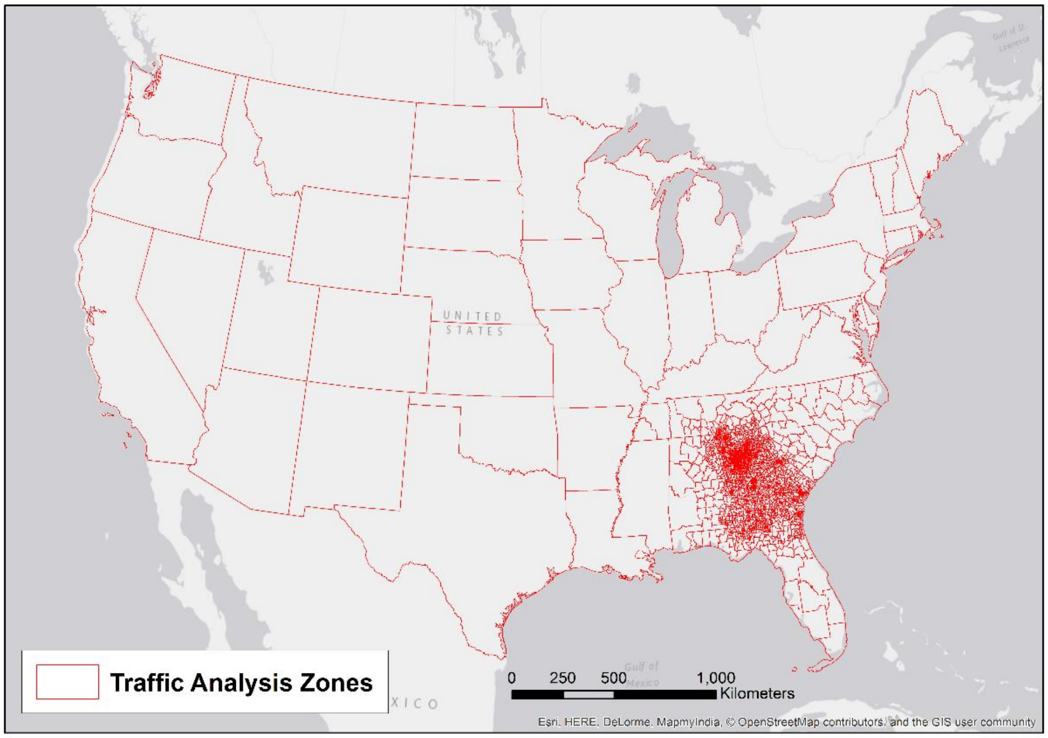

Figure 2: Nationwide Traffic Analysis Zones

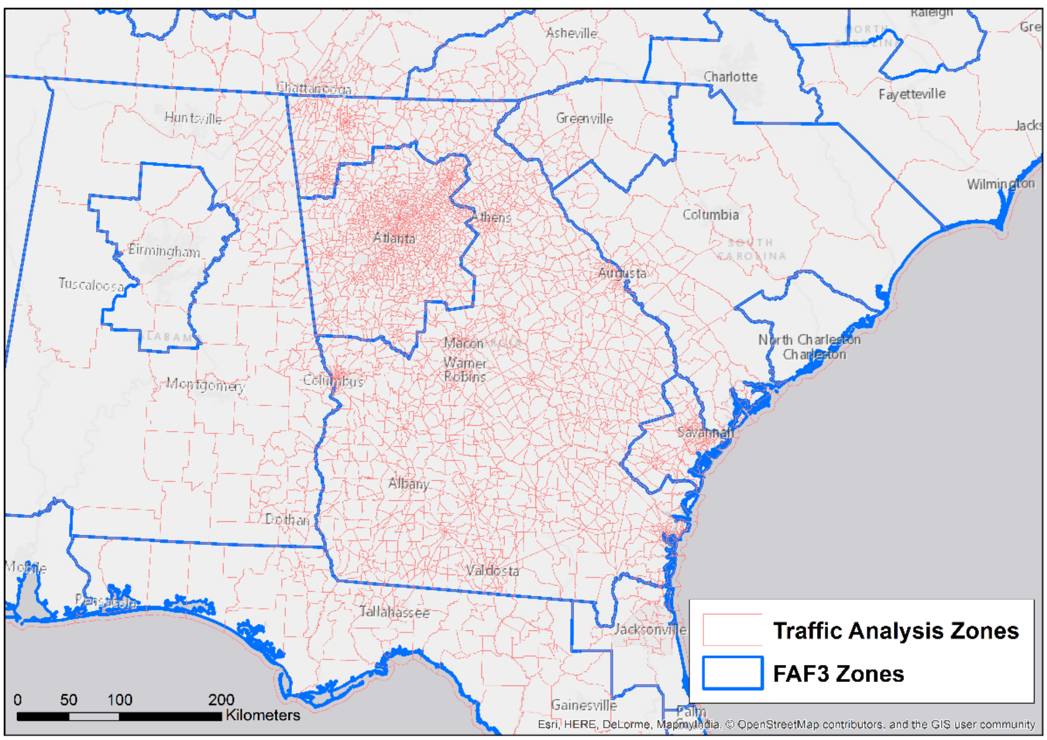

Figure 3: Southeastern Traffic Analysis Zones

CHAPTER III. DATA AND METHODOLOGY

Data Sources

This chapter introduces the data sources that are employed in this study. Since the data sources for task 1 and task 2 vary, the chapter alternates between two subsections, one for each task. The first subsection introduces the data used for disaggregating the FAF3 commodity to the county-level, and the second subsection introduces the data sources used for TAZ-level disaggregation.

Data Sources for Task 1: Disaggregation from FAF zone to County

The county-level disaggregation employs two primary datasets and a series of supporting data, detailed in Table 1 below. The primary datasets are FAF3 from the U.S. Department of Transportation and the County Business Patterns (CBP) from the U.S. Census Bureau. FAF3 provides the commodity flow data to be disaggregated, and CBP provides economic information that links commodity movements to economic activity. Commodity movement is reflected in specific patterns of economic activity at both production (origin) and attraction (destination) locations. FAF3 data includes FAF commodity categories at the scale of the FAF zones. County Business Patterns provide county-level employment for industries according to the North American Industrial Classification System (NAICS) codes, which range from two to six digits, reflecting increasing level of specificity. Three-digit NAICS codes were selected for this analysis because they provide industrial information that is specific enough to be tied to production or attraction of specific commodities without being too narrow and prescriptive.

Table 1: Data Sources, Section 1 (County-level)

Name	Source	Data	Year	Notes
Freight Analysis Framework, version 3 (FAF3)	U.S. Department of Transportation	Commodity movement tonnage among FAF zones	2007	
County Business Patterns (CBP)	U.S. Census Bureau	Employment	2007	
Decennial Census	U.S. Census Bureau	Population	2010	Unabridged county-level population not available for years between Decennial Census.
American Community Survey (ACS), 5-year (2006-2010)	U.S. Census Bureau	Mean Household Income	2010	
National Highway Freight Network	U.S. Department of Transportation	Shapefile	Current edition	
County Shapefile	U.S. Census Bureau, TIGER/Line	Shapefile	2008	2008 is the closest year to the FAF3 year for which a shapefile is available.
Tonnage - Truck Conversion Table	South California Association of Governments	Conversion tables between commodity tonnage and truck trips		

Additional datasets used in task 1 include the Decennial Census for population variables, the 5-year American Community Survey (ACS) for income variables, and shapefiles from the U.S. Department of Transportation and the U.S. Census Bureau.

Data Sources for Task 2: Disaggregation from County to TAZ

The second task (TAZ-level disaggregation) includes different datasets as a function of the data that is available at very fine geographic scales. The primary data sources are the county-level truck flows from section 1 and the 5-year American Community Survey, which provide the primary dependent and independent variables used in disaggregation to the TAZ level. Supporting data includes several shapefiles that were used to establish relationships between the FAF zones, counties, TAZs, and block groups. Table 2 provides a complete list of the data sources for task 2.

Table 2: Data Sources, Section 2 (TAZ-level)

Name	Source	Data	Year
Output from section 1	Calculated from section 1.	County-to-county truck movement estimates	2007
American Community Survey, 5-year (2007-2011)	Social Explorer	County and block group-level socioeconomic variables	2011
Traffic Analysis Zones	Georgia Department of Transportation	Shapefile	Current
County Shapefile	U.S. Census Bureau, TIGER/Line	Shapefile	2008
Block groups (AL, FL, GA, NC, SC, TN)	U.S. Census Bureau	Shapefile	2011

Methodology

Both tasks of this study employ similar methodology to perform disaggregation. The NCFRP report 26 called *Guidebook for Developing Subnational Commodity Flow Data* describes four methods for disaggregating data: geographic allocation, regression methods, iterative proportional fitting, and input-output models. Each is appropriate in different situations depending on the project goals and the known relationships (NCFRP, 2013). Regression methods were selected for this study because they allow productions and attractions to be disaggregated based on economic activity (i.e., industry-specific employment) in task 1, which was available for all counties in this study. Moreover, economic data has a strong conceptual link with commodity movement since freight demand is derived directly from economic activity. For task 2, the regression methods allows disaggregation based on the socioeconomic and transportation network data available at the local level. Iterative proportional fitting is another common disaggregation method that was also considered, but it would not add value in this case because the freight flows are doubly constrained. In other words, since both productions and attractions at the FAF3 level are known, iterative proportional fitting solves the allocations produced by the regression method in only one step, which is the same as applying a multiplier.

The two tasks of this study relied heavily on regression analysis to explore the relationships between the commodity or truck flows of a place and its industry, socioeconomic, and infrastructure characteristic to generally estimate the commodity movement or trucks going into/out of an area. The regression methods establish relationships between the variable being disaggregated and other conceptually related explanatory and control variables. In task 1, this means establishing a relationship between commodity productions and attractions (categorized by Standard Classification of Transported Goods – SCTG codes), economic activity (employment categorized by three-digit NAICS codes), and other explanatory variables like population and highway access. In task 2, the relationship is established between truck trip production and attractions, socioeconomic variables, and transportation network access. The relationship is first determined at the level at which freight movement data is currently aggregated (i.e., the FAF zones in task 1 and counties in task 2), and then the relationship is applied to economic activity and control variables at the disaggregation level (i.e., counties in task 1 and TAZs in task 2) to estimate lower-level productions and attractions.

The next step is to apply these estimates to the lower-level disaggregation areas. The study estimates commodity-specific production and attraction for each county as a share of the FAF zone totals (task 1) and TAZ truck productions or attractions as a share of county totals (task 2). There are multiple differences in the methodological details of the two tasks due to the distinct scale of counties and TAZs, consequently, the following sections detail the methods for tasks 1 and 2 separately.

Methodology for Task 1: Disaggregation from FAF zone to County

For task 1, the methodology includes four main steps: regression, initial estimation, O-D flow disaggregation, and commodity to truck conversion. Figure 4 illustrates the sequence of the steps, which are detailed below.

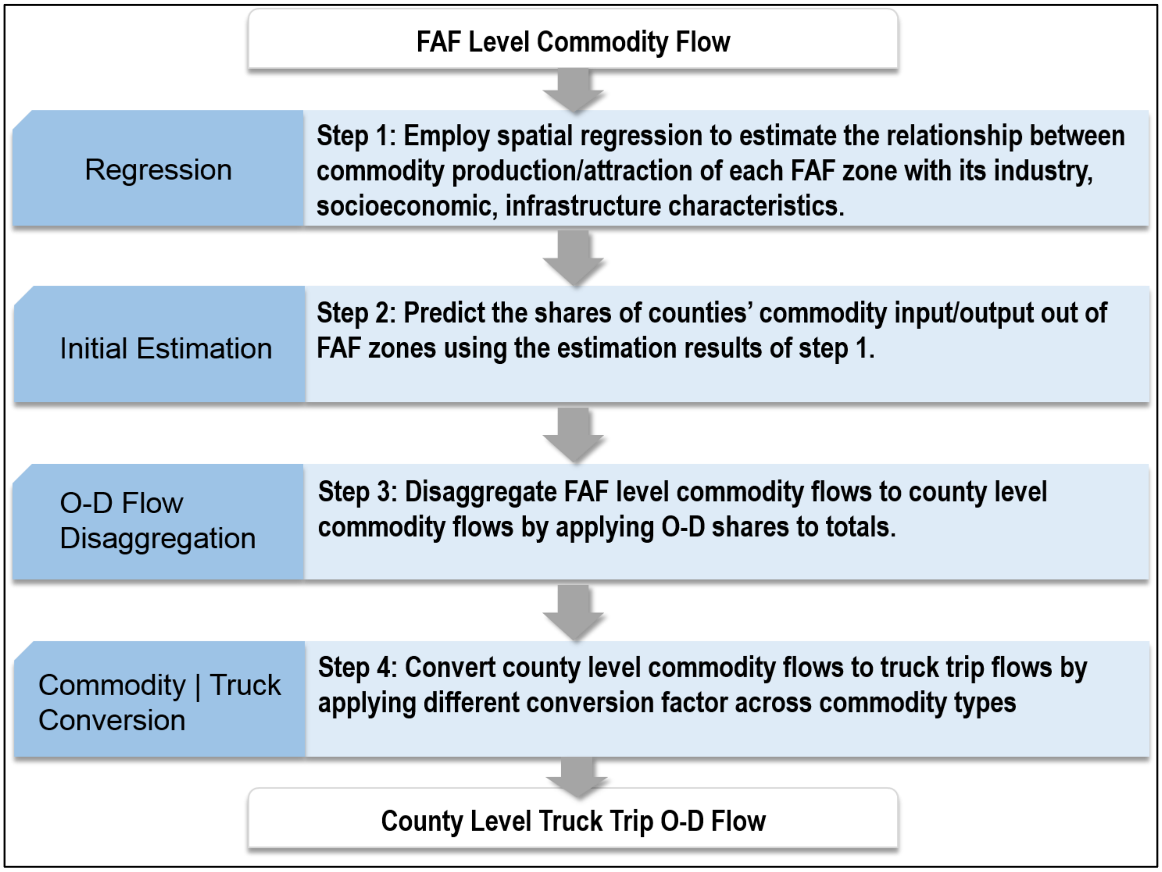

Figure 4: Task 1 Methodology

Task 1 – Step 1: Regression

Regression quantifies the relationship between a FAF zone's commodity productions or attractions, and its employment, socioeconomic variables, and infrastructure characteristics, which are all conceptually related with commodity flows. Sector-specific employment reflects economic activity that supplies or demands commodities, while socioeconomic variables relate to the population's size and purchasing power, which affect commodity use. Moreover, access to transportation infrastructure acknowledges the ways in which access to regional and national markets shape economic location. The first step explores and quantifies these relationships using regression analysis.

Task 1's dependent variable is the commodity tonnage by SCTG code at the FAF zone level. Commodities are not only treated by code, of which there are 43, but also by productions and attractions, making for a total of 86 regressions (i.e., 43 commodities x 2 directions). A series of explanatory variables was included in the model (see Table 3) and all the variables were aggregated to FAF level to assure consistency with the dependent variables.

Table 3: Variables for Task 1

Name	Description	Role
[43 SCTG codes]	Commodity tonnage by SCTG code for both productions and attractions.	Dependent Variable
[87 NAICS codes]	Estimated employment by NAICS code (3 digit)	Explanatory (independent)
WeightedHHInc	Weighted mean household income	Explanatory (independent)
ln_WeightedHHInc	Natural log of the weighted mean household income	Explanatory (independent)
NHFN_km	Length of National Highway Freight Network (km)	Explanatory (independent)
ln_NHFN_km	Natural log of the length of National Highway Freight Network (km)	Explanatory (independent)
Density_km_per_km2	National Highway Freight Network density (length divided by area)	Explanatory (independent)
ln_Density_km_per_km2	Natural log of the National Highway Freight Network density (length divided by area)	Explanatory (independent)
Area_km2	Area in square kilometers	Explanatory (independent)
ln_Area_km2	Natural log of the area in square kilometers	Explanatory (independent)
Pop2010	Population (2010)	Explanatory (independent)
ln_Pop2010	Log transformed population (2010)	Explanatory (independent)

Extensive data processing was required to assemble a single dataset with all the variables. The first data processing step was to convert FAF3 and CBP data to the same geographic scale. An equivalency table was created between counties and FAF zones using 2007 county shapefiles from the U.S. Census Bureau (TIGER/Line) and a shapefile for FAF zones from the U.S. Department of Transportation. Counties contained within a given FAF zone were assigned to that zone, where each county is assigned to only one zone. FAF zones and counties were visually inspected to ensure that no counties are included in more than one FAF zone. This equivalency table was then used to aggregate employment for all three-digit NAICS codes from the county to the FAF-zone level. Simultaneously, the commodity movement data is edited to only include truck movement (coded as mode '1' in FAF3 data). Separate tables are made to sum all commodity productions and attractions, and to sum the tonnage for each of these areas by the 43 commodity types.

The primary explanatory variables are the employment data across different industry sectors recorded by three-digit NAICS codes. However, CBP's reported employment is incomplete because the U.S. Census Bureau redacts some data for privacy, reporting instead the number of establishments in that sector for each of several employment size ranges. The reported establishments were used to estimate the number of employees missing from the reported data in each industry. This estimation involved three steps. First, employment is estimated for all three-digit industries and all counties (including those with reported employment) by multiplying the number of firms of each size by the mean employment size and summing the result for each county-industry pair (Equation 1). The estimated employment was compared to the actual employment in those counties with reported employment for each NAICS code, and a scaling factor was created for each NAICS code (Equation 2). The scaling factor adjusts the sum of estimated employment to equal the sum of reported employment for those counties that reported employment. Finally, the industry-specific scaling factors were applied to the estimated employment in those counties without reported employment (Equation 3).

$$\widehat{e_{nm}} = \sum_{\forall \mu\_\phi} (\mu + (\phi - \mu)/2)\, f_{\mu\_\phi} \tag{1}$$

$$\theta_n = \frac{\sum_m \widehat{e_{nm}} - \sum_m e_{nm}}{\sum_m e_{nm}} \quad \forall e_{nm} \in \text{non null} \tag{2}$$

$$\text{If } e_{nm} = null, \text{ then assign } e_{nm} = \frac{\widehat{e_{nm}}}{1+\theta_n} \tag{3}$$

Where,

$\widehat{e_{nm}}$ is estimated employment for NAICS code *n* in county *m*;

e_{nm} is actual employment for NAICS code *n* in county *m*;

θ_n is the adjustment factor for NAICS code *n*;

μ is the lower boundary for the establishment employment range;

ϕ is the upper boundary for the establishment employment range (10,000 employees assumed for the uppermost unbounded range);

f_{μ_ϕ} is the number of establishments (firms) with between μ and ϕ employees.

Several additional explanatory variables were also computed. They include the weighted mean household income for each FAF zone. ‘Weighted’ here signifies that counties with more people are given more consideration than counties with fewer people. To determine the weighted mean household income for a FAF zone, mean household incomes and populations of counties falling inside the boundary of that FAF zone was first extracted from the 5-year American Community Survey. Each county’s mean household income was then multiplied by the county population, these products were summed over the counties in that FAF zone to obtain total FAF zone level household income, and the sum was divided by the FAF zone’s total population (also obtained by summing the population of counties in that FAF zone). This produced the FAF zone-level weighted mean household income. It is important to mention that here the weighted mean is also the true mean as the boundaries of counties inside a FAF zone matched exactly and exclusively with the FAF zone.

Another important variable is the centerline length of the National Highway Freight Network (NHFN) contained in each county and each FAF zone. This variable was calculated in ESRI's ArcMap for both counties and FAF zones in kilometers. The FAF zone and county area in square kilometers is a fourth control variable. Finally, the NHFN centerline length and area in square kilometers were combined to form another control variable: freight network density. Freight network density equals NHFN length divided by area.

With all the employment, socioeconomic, and infrastructure-related independent aggregated or calculated at the FAF level, there is great consistency between the geographical scale of the dependent variable and the independent (explanatory) variables, which is necessary for conducting regression analysis. However, since all the variables involve spatial information in them, simple linear regression might suffer from spatial autocorrelation. Therefore, the analysis considered both ordinary least squares (OLS) regression and spatial error models for each regression, and spatial error models were used whenever spatial dependence was detected in the data. The simple linear regression models take a form described by Equation 4.

$$y = X(\beta) + \varepsilon \tag{4}$$

Where,

y is the dependent variable;

X is a matrix of observations on independent variables;

β is a vector of coefficients that can be estimated with the given data, representing the strength and type of relationship between X and y;

and ε represents the residuals or disturbance terms which are assumed to be non-correlated among themselves (non-autocorrelated) or with regressor and homoscedasticity normally distributed with zero mean.

The simple OLS linear regression model is widely used to statistically model the relationship between a dependent variable and explanatory variables due to their desirable property termed as BLUE (Best Linear Unbiased Estimator). It is important to mention that OLS estimates are BLUE only if OLS assumptions are not violated (for these assumption see (Washington et al., 2010)). However, in this study, all the variables are spatial variables or involve spatial information, which may result in the spatial-autocorrelation in the residuals. This means an important assumption of linear OLS regression that the residuals should be random, non-autocorrelated and normally distributed, may not hold if the residuals are spatially autocorrelated. As a consequence the OLS estimates lose efficiency (hence no longer are BLUE) although they are unbiased and consistent. Under such circumstances it is better to use spatial regression model whenever spatial dependence is detected. The spatial error model is a type of spatial regression model that takes the form described by Equation 5.

$$y = X(\beta) + \varepsilon,\ \varepsilon = \lambda(W)\varepsilon + \mu \tag{5}$$

Where,

W is a spatial weights matrix developed based on spatial positions of the units of analysis;

X is a matrix of the observations on explanatory variables;

β is a vector of the coefficients of explanatory variables;

ε is a vector of spatially autocorrelated error terms;

μ is a vector of i.i.d errors;

λ is a coefficient parameter on the spatially correlated error terms.

A simple logic, depicted in Figure 5, determines the regression type that is most appropriate. The analysis begins with a set of tests such as Moran's I and the Lagrange multiplier to assess the level of spatial dependence. Whenever a significant amount of spatial dependence was found, a spatial error model was used rather than OLS regression to address the spatial autocorrelation issue.

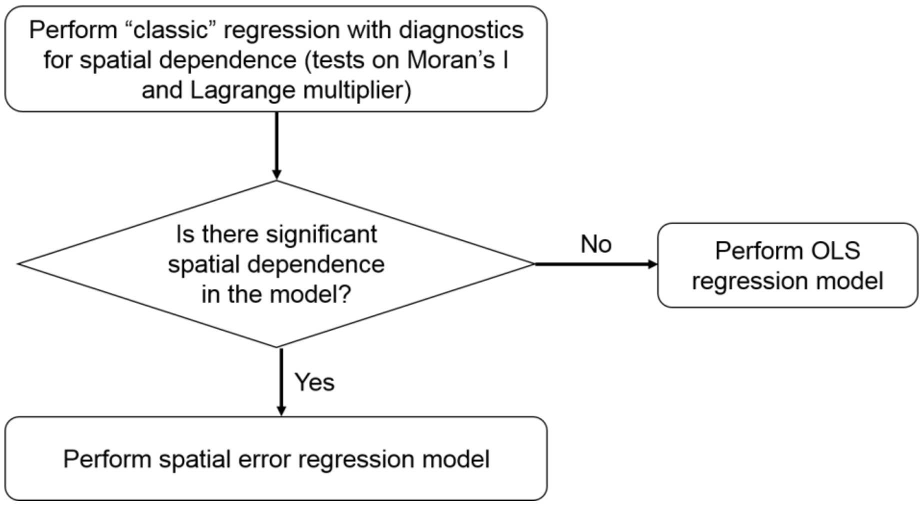

Figure 5: Logic of Choosing Between Simple Linear Regression and Spatial Error Regression Models

Another challenging task was to select useful explanatory variables for each regression model. Four steps directed the selection of variables for the regression models. First, independent variables with at least a moderately strong conceptual relationship with the dependent variable were added. Second, variables that were proved to be statistically insignificant or with the opposite sign of coefficient from what was expected were removed. Third, independent variables with high correlations with other independent variables were removed to avoid the collinearity between the explanatory variables. Multiple combinations of independent variables were tested to obtain models that make conceptual sense and provide high explanatory power evidenced by R-square statistics approaching 1.

Following this logic, we developed 86 regression models (for commodity productions and attractions across 43 commodity types) at the FAF level. The 86 regression models are either simple linear OLS regression or spatial error models, depending on the result of tests of the data’s spatial dependence. The regression results are presented in chapter 4 of this report.

Task 1 – Step 2: Initial Estimation

The next step is to estimate initial allocations for disaggregated produced and attracted tonnage by each commodity type at the county level. Because freight movement is a derived demand that is reflected in economic processes both at its origin and destination, it is possible to use industrial presence to estimate both the demand for and the availability of goods within each FAF zone. The total productions and attractions within each FAF zone remains a control total, and productions and attractions can be allocated based on the proportion of relevant employment in the county. The result from the previous step allows us to quantify the relationship between a zone's commodity productions and attractions, and its employment, socioeconomic, and transportation network characteristics. In this step, we apply the coefficients from those regression models to the county-level independent variables to derive the county's predicted values. The predicted values reflect the counties' 'relative potential' to produce or attract different types of commodities.

Since neither OLS nor spatial error models guarantee positive predicted values, the resulting county-level predicted values were sometimes negative. However, commodity movement must be either zero or positive, so negative values are impossibilities resulting from simplifications inherent to statistical models. Therefore, three techniques were jointly applied to eliminate negative predicted values. The first is to simply convert all negative values to zero; the second technique compresses the range from the minimum value to the maximum value until the minimum negative value reaches zero, holding the maximum value constant and moving all intermediate values upward proportionately. The third method adds the difference between the minimum negative value and zero to all values such that the minimum negative value becomes zero and all other values increase by the same amount. The average of the three adjusted values was used as the final predicted values in this step. These county level 'predicted values' were then used to calculate a county's share of commodity productions and attractions out of the total of its FAF zone.

Figure 6 illustrates the process with a simplified case, where an FAF zone includes just four counties. The production or attraction shares of the four counties can be noted as P_i ($i = 1, 2, 3, \dots, 43$) and A_i ($i = 1, 2, 3, \dots, 43$) respectively and for each of the 43 commodity types. The FAF productions and attractions are equal to $\sum_{n=1}^{4}(P_i) = 1$ and $\sum_{n=1}^{4}(A_i) = 1$ because the FAF zone contains four counties. The production and attraction shares of a county always vary across commodity types since their shares were estimated using regression-derived models from the previous step. The fact that each county attracts or produces different shares of each commodity is expected since real counties have unique industrial specializations and population characteristics that result in distinct freight commodity mixes.

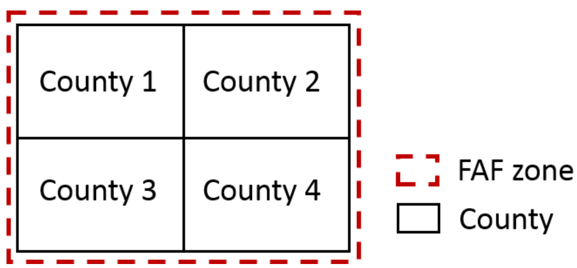

Figure 6: Conceptual Example of Initial Estimation

Task 1 – Step 3: O-D Flow Disaggregation

After estimating counties' production and attraction shares, the next step is to disaggregate the FAF zone level commodity flows to county-level commodity flows. The disaggregation is based on Equation 6.

$$C_{shk} = F_{sod} \times P_h \times A_k \tag{6}$$

Where,

C_{shk} represents the tonnage of commodity type s going from county h to county k;

F_{sod} represents the tonnage of commodity of type s going from FAF zone o to zone d;
County h is located in zone o, and county k is located in zone d;
P_h is the production share of county h out of FAF zone o, and A_k is the attraction share of county k out of FAF zone d.

This method balances practical and theoretical considerations. On the one hand, this disaggregation process guarantees that the county O-D commodity volume completely aligns with the FAF O-D flows for each commodity type. On the other hand, this disaggregation takes into consideration the fundamental and functional relationships between a locations' commodity productions / attractions and its economic, demographic, and transportation characteristics. Therefore, this disaggregation method is practically valid and theoretically sound. The output of this step is an extremely large table (over 400 million records) showing the tonnage for each of the 43 commodity types going between each of the over 9 million county-county O-D pairs and are not reported in this report.

Task 1 – Step 4: Commodity to Truck Conversion

The final step in task 1 is to convert commodity flows to truck trips. We applied a set of truck payload factors across different types of commodity, as shown in Table 4. The tons-per-truck conversion factors were applied to the county-level commodity flows to estimate the number of trucks for each type of commodity for each O-D county pair, which were then summed to produce total annual truck trips between county pairs.

Table 4: Commodity to Truck Conversion Table

SCTG Code	Description	Heavy Heavy Duty Truck (HHDT) Payload Factor (tons per truck)
01	Live animals/fish	16
02	Cereal grains	16
03	Other ag prods.	16
04	Animal feed	16

SCTG Code	Description	Heavy Heavy Duty Truck (HHDT) Payload Factor (tons per truck)
05	Meat/seafood	16
06	Milled grain prods.	16
07	Other foodstuffs	15
08	Alcoholic beverages	15
09	Tobacco prods.	15
10	Building stone	14
11	Natural sands	14
12	Gravel	14
13	Nonmetallic minerals	16
14	Metallic ores	24
15	Coal	18
16	Crude petroleum	15
17	Gasoline	15
18	Fuel oils	11
19	Coal-n.e.c.	18
20	Basic chemicals	14
21	Pharmaceuticals	14
22	Fertilizers	14
23	Chemical prods.	14
24	Plastics/rubber	12
25	Logs	14
26	Wood prods.	16
27	Newsprint/paper	15
28	Paper articles	13
29	Printed prods.	15
30	Textiles/leather	11
31	Nonmetal min. prods.	16
32	Base metals	24
33	Articles-base metal	15
34	Machinery	9
35	Electronics	8
36	Motorized vehicles	11
37	Transport equip.	11
38	Precision instruments	10
39	Furniture	9
40	Misc. mfg. prods.	8
41	Waste/scrap	14
43	Mixed freight	7
99	Unknown	13

Source: Revised from (SCAG, 2012)

Task 2 Methodology: Disaggregation from County to TAZ

Task 2 employs a similar method as employed for task 1, and is shown in Figure 7. However, because most TAZs are much smaller than counties, some variables that were included in the regression models in task 1 are not available for TAZs. Commodity-specific disaggregation to TAZs, which can approach the size of census blocks, is also extremely difficult to accurately perform based on regression because of the disparity across different TAZs (e.g., some TAZs might not have produced or attracted certain commodities in any amount). Therefore, task 2 regression was performed with truck trips as the dependent variable rather than commodity tonnage. Otherwise, the task 2 methodology follows a nearly identical logic to the one used in task 1. The primary difference in the approach is that the final step of conversion from commodity truck tons to truck trips is omitted since all calculations in task 2 are in truck trips.

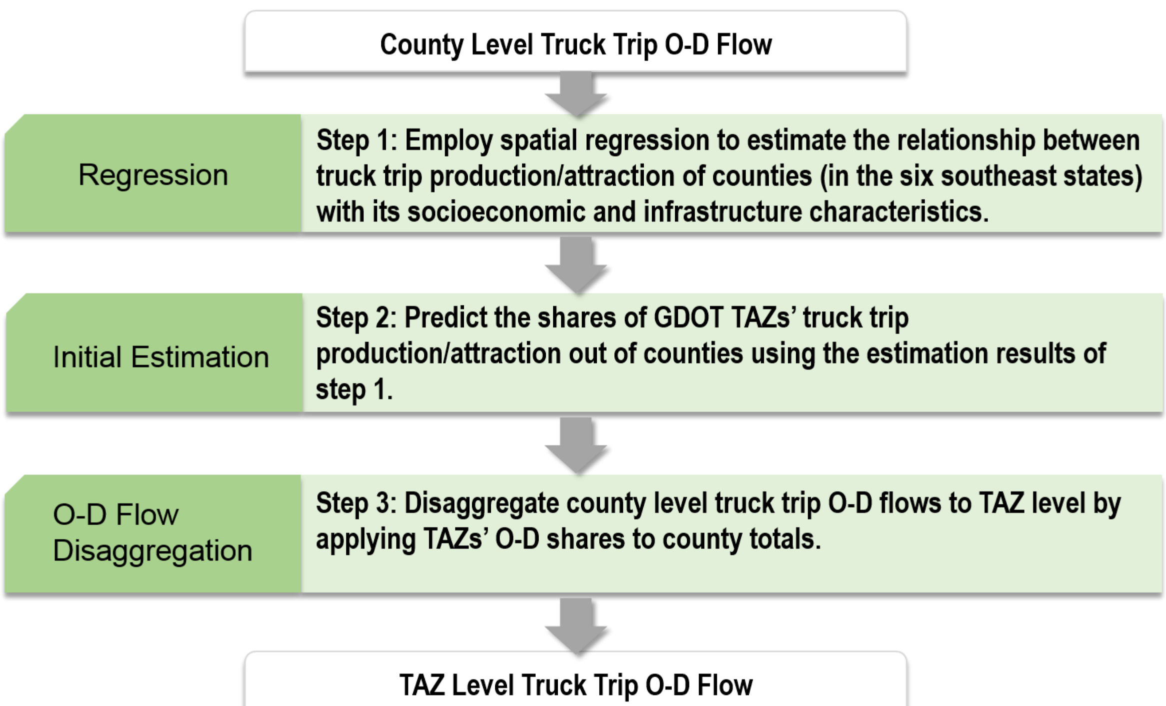

Figure 7: Task 2 Methodology

Task 2 analysis begins at the county-level to establish relationships between freight tonnage and socio-economic variables, and then we apply those relationships at the TAZ level. Therefore, independent variables in the regression section were collected and processed for both the county level and the TAZ level (detailed in chapter 3). The output from task 1 (truck movements by county pairs) is converted into county truck productions and attractions, and it becomes the dependent variables for task 2. A large set of other variables were tested and considered, as evident from Table 5. Many of these variables are available from the U.S. Census Bureau's CBP, and others were calculated based on spatial characteristics. All of these variables were collected at the county level, and then thus that were ultimately retained for the models were collected for U.S. Census Block Groups and converted to TAZs.

Table 5: Variables for Task 2

Name	Description	Role
*Number of trucks originating from a county (productions)	Result from task 1	Dependent Variable
*Number of trucks arriving in a county (attractions)	Result from task 1	Dependent Variable
Area	Area of the county (square kilometers	Explanatory (independent)
Pop2010	Population (2010)	Explanatory (independent)
*Sqrt_Pop2010	Square root of population (2010)	Explanatory (independent)
Rail_km	Length of rail in county (km)	Explanatory (independent)
*Sqrt_Rail_km	Square root of length of rail	Explanatory (independent)
Dist_NHFN	Distance from the TAZ centroid to the nearest National Highway Freight Network segment (km)	Explanatory (independent)
*Sqrt_Dist_NHFN	Square root of distance from the TAZ centroid to the nearest to NHFN segment (km)	Explanatory (independent)
PopDenSqKm	Population density (people per square kilometer)	Explanatory (independent)
HHInc	Average household income ($ per year)	Explanatory (independent)
*Sqrt_HHInc	Square root of average household income	Explanatory (independent)
ColGradRate	Rate of college graduates in adult population	Explanatory (independent)
Ln_ColGradRate	Natural log of the rate of college graduates	Explanatory (independent)
UnempRate	Unemployment rate	Explanatory (independent)
Ln_UnempRate	Natural log of unemployment rate	Explanatory (independent)
AvgComMinutes	Average commuting time (minutes)	Explanatory (independent)
Gini_Index	Gini index	Explanatory (independent)

Name	Description	Role
AirportName	Dummy variable designating whether one of the largest 100 cargo airports is in the county	Explanatory (independent)
AirportRank	The ranking of the airport (only for top 100 airports)	Explanatory (independent)

*Variables retained in the final regressions.

Block group-level data requires substantial processing to convert it to TAZs. As visible in Figure 2 and Figure 3 (page 8), TAZ size is highly variable. They are smallest in Georgia and directly adjacent states to provide the greatest spatial precision over GDOT's jurisdiction, and they are even smaller in the most populated parts of Georgia. By contrast, TAZs outside of the southeastern United States cover entire states, which are as large or larger than FAF zones. Most TAZs conform to census blocks (the most spatially precise census area), and many contain one or more block groups (which are composed of several joined blocks). However, the overlap is imperfect, with most TAZs containing multiple block groups and some block groups spanning several TAZs. The block group is normally the lowest level at which the U.S. Census Bureau releases socioeconomic data. Therefore, block groups were used to collect and compile socioeconomic data to the TAZ level.

There are two general types of variables that had to be treated differently in data conversion from block groups to TAZs. The first is variables in absolute numbers. They represent characteristics of the area that are spread over the block group in a way that is assumed to be even. When the block group is divided, each remaining section maintains a share of the variable total consistent with its portion of the entire block group area. These variables include population, and they are aggregated from the block group to the TAZ level according to Equation 7 below.

$$Pop_r = \sum_{i \in n_r} \left(\frac{A_i'}{A_i} \right) \times Pop_i \tag{7}$$

Where,

Pop_r is the total population of TAZ *r*;

A_i' is the area of block group *i* that intersects with TAZ *r*;

A_i is the total area of block group *i*;

Pop_i is the total population of block group *i*;

and n_r is the set of block groups having non-zero area of intersection with TAZ *r*.

The second variable type has characteristics of rates or averages, which are assumed to be equally true over the block group's entire area. Examples include mean household income and unemployment rate. These variables are indivisible over the block group but must nonetheless be combined at the TAZ level to reflect the average of the block groups that compose it. This aggregation was done on the basis of each block group's portion of the TAZ's total area, as described in Equation 8. While Equation 7 scales variables based on their portion of the block group's area, Equation 8 scales variables by the portion of the TAZ's area.

$$Prop_r = \sum_{i \in n_r} \left(\frac{A_i'}{A_r} \right) \times Prop_i \tag{8}$$

Where,

A_i' is the area of block group *i* that intersects with TAZ *r*;

A_r is the total area of the TAZ *r*;

$Prop_i$is the property (such as mean annual household income) of block group *i*;

and $Prop_r$ is the property of the TAZ *r*.

Figure 8 below illustrates the relationships among the different units and areas used to scale variables. It shows a TAZ (in orange) composed of three block groups. Two block groups (in black) are entirely contained inside the TAZ, and the third block group (in blue) is only partially continued inside the TAZ, although its intersected area (in green) is entirely contained within the

TAZ. The areas of the TAZ, the block groups, and the intersects between TAZs and block groups provide the multipliers needed to aggregate partially overlapping block group characteristics to the TAZ level.

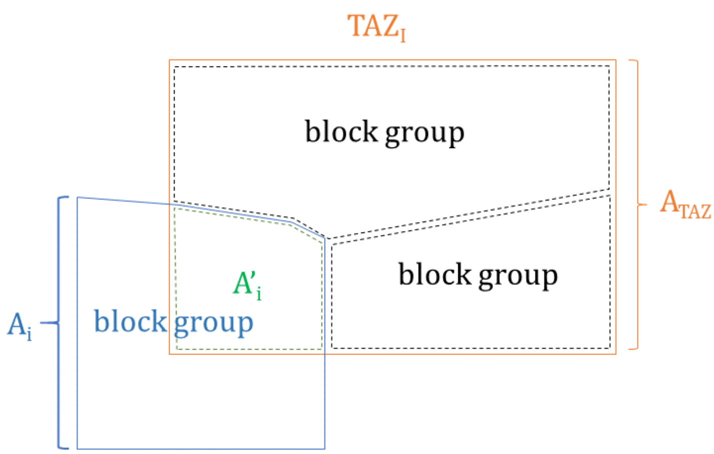

Figure 8: Conceptual Diagram of Data Aggregation from Block Groups to TAZs

Task 2 – Step 1: Regression

As in task 1, the first step is to establish a relationship between the productions / attractions, and relevant variables. This is performed on the six states of the southeastern region (AL, FL, GA, NC, SC, TN) since this is the only area with TAZs smaller than states. Determining the relationships based only on this region increases internal validity since they are based on data from the region where they will be applied rather than national data, which would obscure regional trends.

Two regression models have been estimated, one for county-level truck trip productions and another for county-level truck trip attractions. Independent variables are selected by the same method as in task 1. Both models demonstrate spatial autocorrelations in this case, so a spatial error model is used rather than OLS. The regression variables and results are presented in chapter 4 of this report.

Task 2 – Step 2: Initial Estimation

The constants and coefficients from step 1 are applied to TAZ-level socio-economic and network data to for TAZs smaller than counties, which is all TAZs in Georgia and TAZs in counties directly adjacent to Georgia. Then, these predicted values are used to compute the shares of productions and shares of attractions that is equal to the ratio of a TAZ's predicted value and the sum of predicted values of all TAZs in that county. Typically, calculated shares for productions and attractions for the zones were not equal.

For TAZs as large as a county or larger, there is no need for initial estimations. Indeed, truck trip productions and attractions are already known at the county level from step 1, and the productions or attractions for TAZs containing multiple counties are obtained by summing the productions and attractions of all their constituent counties. These TAZs with known values are assigned a share of 1.

Task 2 – Step 3: O-D Flow Disaggregation

The O-D flow disaggregation proceeds similarly to task 1, with the primary difference being that TAZs of the size of a county or larger are assigned a share of 1. The equation used multiplies production and attraction shares for each TAZ by the truck trips between the county pairs. If one or both TAZs are larger than counties, then all truck trips between the concerned counties have been summed to the TAZ level. Equation 9 below details the disaggregation procedures.

$$t_{rv} = T_{hk} \times P_r \times A_v \tag{9}$$

Where,

t_{rv} represents the truck trips moving from TAZ r to TAZ v;

T_{hk} represents the truck trips moving from county h to county k. When a TAZ is larger than a county, the truck trips of all constituent counties are summed;

TAZ r is located in zone h, and TAZ v is located in zone k. When a TAZ is larger than a county, the TAZ is substituted for the county. The equation is not repeated for the remaining counties in that TAZ;

P_r the production share of TAZ r out of county h, and A_v is the attraction share of TAZ v out of county k when TAZ r and TAZ v are smaller than counties.

This equation produces a table of annual truck trips between TAZs. Since there are 3,505 TAZs, the resulting table includes over 12 million unique pairs. This is the primary deliverable for task 2, which is detailed in Chapter 4.

CHAPTER IV. RESULTS AND ANALYSIS

Result of Task 1: Nationwide County to County Truck Trip Volume

Regression Results for Task 1

As explained in previous chapter of the report, for both productions and attractions, several industries were identified that are logically expected to be correlated with the zone's production or attraction of a given commodity. Industries were selected from the set of industries to obtain high model fit (e.g., R-Squared), emphasis was to use only a small number of relevant industries for each commodity, and to ensure that all model coefficients are positive. Some industries were found to be heavily spatially clustered by analyses such as the global Moran's I, visual inspection, and the Lagrange multiplier. When the commodity was shown to be highly spatially clustered, a spatial error regression model was used to correct for spatial correlation in the error terms, otherwise, OLS regression was employed.

Table 6 and Table 7 below summarize the model type and variables used to establish the relationships respectively between commodity productions and attractions, and industrial employment. Almost all of the production and attraction commodity tonnages can be predicted with reasonable accuracy at the FAF zone level using FAF zone employment in key industries, as shown by the mean productions R-Squared of 0.65 and the mean attractions R-squared of 0.71. Moreover, each model was checked for regression assumptions, including a linear relationship, normal distribution of errors, and homoscedasticity. When these conditions were not automatically met, adjustments were made to the variables (mostly by transformation of variables) to meet as many regression assumptions as possible. For instance, many regressions predicted the square root of the dependent variable or took the natural log of control variables. Moreover, the models with high spatial correlation also accounted for it in model selection. Afterwards, the models for each commodity type were applied to county-level data to establish predicted tonnage for productions and attractions of each commodity.

Table 6: Variables Used in Regressions for Task 1 Productions

SCTG	SCTG Description	Spatial Error Model (S) or OLS regression (O)	Explanatory Variables			R-Squared
01*	Live animals/fish	S	WeightedHHInc	Area_km2	Food Manufacturing	0.70
02*	Cereal grains	S	WeightedHHInc	Area_km2	Food Manufacturing	0.64
03	Other ag prods.	S	Pop2010	Support Activities for Agriculture and Forestry	Food Manufacturing	0.83
04*	Animal feed	S	WeightedHHInc	Forestry and Logging		0.64
05	Meat/seafood	S	Pop2010	Food Manufacturing		0.78
06	Milled grain prods.	O	Pop2010	Food Manufacturing	Food Services and Drinking Places	0.63
07	Other foodstuffs	O	ln_Density_km_per_km2	Food Manufacturing	Food Services and Drinking Places	0.81
08	Alcoholic beverages	O	ln_WeightedHHInc	Beverage and Tobacco Product Manufacturing	Food Services and Drinking Places	0.69
09*	Tobacco prods.	S	ln_Pop2010	Beverage and Tobacco Product Manufacturing		0.31
10*	Building stone	O	ln_Pop2010	Construction of Buildings		0.56
11	Natural sands	O	Pop2010	Heavy and Civil Engineering Construction		0.46
12*	Gravel	S	ln_Pop2010	Nonmetallic Mineral Product Manufacturing		0.54
13*	Nonmetallic minerals	O	ln_Pop2010	Printing and Related Support Activities	Petroleum and Coal Products Manufacturing	0.35
14*	Metallic ores	S	Density_km_per_km2	Primary Metal Manufacturing		0.30
15	Coal	O	ln_Density_km_per_km2	Mining (except Oil and Gas)	Support Activities for Mining	0.42
16	Crude petroleum	S	Area_km2	Utilities	Petroleum and Coal Products Manufacturing	0.71
17*	Gasoline	O	ln_Pop2010	Oil and Gas Extraction Sector	Miscellaneous Manufacturing	0.57
18*	Fuel oils	S	ln_Pop2010	Support Activities for Mining	Chemical Manufacturing	0.46
19	Coal-n.e.c.	O	Pop2010	Support Activities for Mining	Miscellaneous Manufacturing	0.55
20*	Basic chemicals	O	Pop2010	Oil and Gas Extraction Sector	Chemical Manufacturing	0.36
21	Pharmaceuticals	O	ln_Pop2010	Nonmetallic Mineral Product Manufacturing	Merchant Wholesalers, Nondurable Goods	0.52

SCTG	SCTG Description	Spatial Error Model (S) or OLS regression (O)	Explanatory Variables			R-Squared
22*	Fertilizers	O	ln_NHFN_km	Food Manufacturing	Chemical Manufacturing	0.49
23*	Chemical prods.	S	ln_Density_km_per_km2	Chemical Manufacturing	Nonmetallic Mineral Product Manufacturing	0.66
24*	Plastics/rubber	S	Pop2010	Petroleum and Coal Products Manufacturing	Chemical Manufacturing	0.83
25*	Logs	S	NHFN_km	Forestry and Logging		0.78
26*	Wood prods.	O	Pop2010	Forestry and Logging	Wood Product Manufacturing	0.87
27	Newsprint/paper	O	ln_Area_km2	Forestry and Logging	Paper Manufacturing	0.76
28*	Paper articles	O	ln_Pop2010	Paper Manufacturing		0.71
29	Printed prods.	O	NHFN_km	Printing and Related Support Activities		0.68
30	Textiles/leather	S	ln_Density_km_per_km2	Textile Manufacturing	Textile Product Mills	0.73
31*	Nonmetal min. prods.	S	ln_Pop2010	Construction of Buildings	Nonmetallic Mineral Product Manufacturing	0.79
32*	Base metals	O	ln_Density_km_per_km2	Primary Metal Manufacturing	Machinery Manufacturing	0.68
33*	Articles-base metal	S	ln_Pop2010	Fabricated Metal Product Manufacturing		0.74
34*	Machinery	O	Pop2010	Machinery Manufacturing		0.82
35*	Electronics	S	Pop2010	Electrical Equipment, Appliance, and Component		0.77
36*	Motorized vehicles	S	Pop2010	Transportation Equipment Manufacturing		0.79
37*	Transport equip.	O	Truck Transportation			0.42
38*	Precision instruments	O	Pop2010	Miscellaneous Manufacturing	Motion Picture and Sound Recording Industries	0.77
39*	Furniture	O	Pop2010	Furniture and Related Product Manufacturing		0.75
40	Misc. mfg. prods.	O	WeightedHHInc	Miscellaneous Manufacturing		0.62
41*	Waste/scrap	O	WeightedHHInc	Nonmetallic Mineral Product Manufacturing	Waste Management and Remediation Services	0.81
43	Mixed freight	O	ln_Pop2010	Warehousing and Storage		0.66
99	Unknown	O	Pop2010			0.98

*square root taken of the dependent variable

Table 7: Variables in Regressions Used for Task 1 Attractions

SCTG	SCTG Description	Spatial Error Model (S) or OLS regression (O)	Explanatory Variables					R-Squared
1*	Live animals/fish	S	Food Manufacturing	NHFN_km	Pop2010			0.72
2*	Cereal grains	S	Food Manufacturing	Area_km2	NHFN_km	Pop2010		0.73
3*	Other ag prods.	S	Support Activities for Agriculture and Forestry	Food Manufacturing	Area_km2	NHFN_km	Pop201 0	0.75
4*	Animal feed	S	Food Manufacturing	Pop2010				0.67
5*	Meat/seafood	O	Food Manufacturing	NHFN_km	Pop2010			0.80
6*	Milled grain prods.	O	Food Manufacturing	Miscellaneous Store Retailers	Area_km2	NHFN_km		0.80
7*	Other foodstuffs	O	Food Manufacturing	Miscellaneous Store Retailers	Area_km2	NHFN_km		0.85
8	Alcoholic beverages	S	Health and Personal Care Stores	Pop2010				0.94
9*	Tobacco prods.	S	ln_Pop2010					0.17
10*	Building stone	O	Construction of Buildings	ln_Pop2010				0.62
11	Natural sands	S	Heavy and Civil Engineering Construction	Pop2010				0.55
12*	Gravel	S	Heavy and Civil Engineering Construction	ln_NHFN_km				0.52
13*	Nonmetallic minerals	O	Nonmetallic Mineral Product Manufacturing	ln_Pop2010				0.45
14	Metallic ores	O	Support Activities for Mining	Primary Metal Manufacturing	Density_km_per_km2			0.45
15*	Coal	S	Mining (except Oil and Gas)	Support Activities for Mining	Primary Metal Manufacturing	ln_NHFN_km		0.48
16*	Crude petroleum	O	Petroleum and Coal Products Manufacturing	Area_km2				0.44
17*	Gasoline	O	Gasoline Stations	Pop2010				0.60
18	Fuel oils	S	Support Activities for Mining	Gasoline Stations	Transit and Ground Passenger Transportation	WeightedHHI nc		0.60

SCTG	SCTG Description	Spatial Error Model (S) or OLS regression (O)	Explanatory Variables					R-Squared
19	Coal-n.e.c.	O	Petroleum and Coal Products Manufacturing	Pop2010				0.70
20*	Basic chemicals	S	Heavy and Civil Engineering Construction	Chemical Manufacturing	Plastics and Rubber Products Manufacturing	ln_Weighted HHInc		0.71
21*	Pharmaceuticals	O	Ambulatory Health Care Services	Nursing and Residential Care Facilities	Pop2010			0.63
22	Fertilizers	S	Support Activities for Agriculture and Forestry	WeightedHHInc	NHFN_km	n_Pop2010		0.49
23*	Chemical prods.	S	Plastics and Rubber Products Manufacturing	Area_km2	ln_NHFN_km	Pop2010		0.73
24*	Plastics/rubber	S	Chemical Manufacturing	Plastics and Rubber Products Manufacturing	Pop2010			0.81
25*	Logs	S	Furniture and Related Product Manufacturing	ln_WeightedHHInc	ln_Pop2010			0.82
26*	Wood prods.	S	Wood Product Manufacturing	Furniture and Home Furnishings Stores	ln_Pop2010			0.88
27*	Newsprint/paper	S	Paper Manufacturing	Administrative and Support Services	ln_Pop2010			0.77
28*	Paper articles	O	Paper Manufacturing	ln_Pop2010				0.74
29*	Printed prods.	O	Paper Manufacturing	Administrative and Support Services	ln_Density_km_per_km2			0.72
30*	Textiles/leather	S	Textile Manufacturing	Textile Product Mills	Clothing and Clothing Accessories Stores	ln_Pop2010		0.89
31*	Nonmetal min. prods.	S	Furniture and Home Furnishings Stores	General Merchandise Stores	Pop2010			0.81
32*	Base metals	O	Primary Metal Manufacturing	Machinery Manufacturing	Transportation Equipment Manufacturing	NHFN_km	Pop2010	0.82
33*	Articles-base metal	S	Heavy and Civil Engineering Construction	Fabricated Metal Product Manufacturing	ln_Pop2010			0.87
34*	Machinery	O	Machinery Manufacturing	Transportation Equipment Manufacturing	Miscellaneous Manufacturing	Pop2010		0.86
35*	Electronics	S	Electronics and Appliance Stores	ln_Density_km_per_km2				0.80

SCTG	SCTG Description	Spatial Error Model (S) or OLS regression (O)	Explanatory Variables					R-Squared
36*	Motorized vehicles	O	Transportation Equipment Manufacturing	Motor Vehicle and Parts Dealers	ln_Pop2010			0.80
37*	Transport equip.	S	Support Activities for Mining	Gasoline Stations	Couriers and Messengers	Density_km_per_km2		0.59
38*	Precision instruments	O	Computer and Electronic Product Manufacturing	Health and Personal Care Stores	ln_Pop2010			0.69
39*	Furniture	O	Furniture and Related Product Manufacturing	Furniture and Home Furnishings Stores	Warehousing and Storage	Rental and Leasing Services		0.89
40	Misc. mfg. prods.	O	Miscellaneous Store Retailers	ln_Pop2010				0.83
41*	Waste/scrap	O	Waste Management and Remediation Services	ln_Pop2010				0.85
43*	Mixed freight	S	Miscellaneous Store Retailers	ln_Pop2010				0.87
99	Unknown	O	Miscellaneous Store Retailers	WeightedHHInc	Pop2010			0.98

*square root taken of the dependent variable

Data Characteristics of Task 1

This section presents data characteristics through a number of descriptive figures and statistics about a portion of the task 1 county-to-county truck trip pairs, first for nationwide data and then for Georgia and neighboring states. They indicate that nationwide, both truck productions and attractions appears to cluster heavily around large metropolitan areas. Largely rural counties anchored by a large city also produce and attract disproportionate truck trips. This shows truck trips to be largely motivated by activity occurring in metropolitan areas, although the trucks oftentimes pass through rural areas in transit. Some counties also include very high volumes of truck movement, including the Hawaiian Islands which is due to their geography and are obviously constrained from using trucks for external trade. Figure 9 and Figure 10 show broadly similar truck trip productions and attractions with local distinctions.

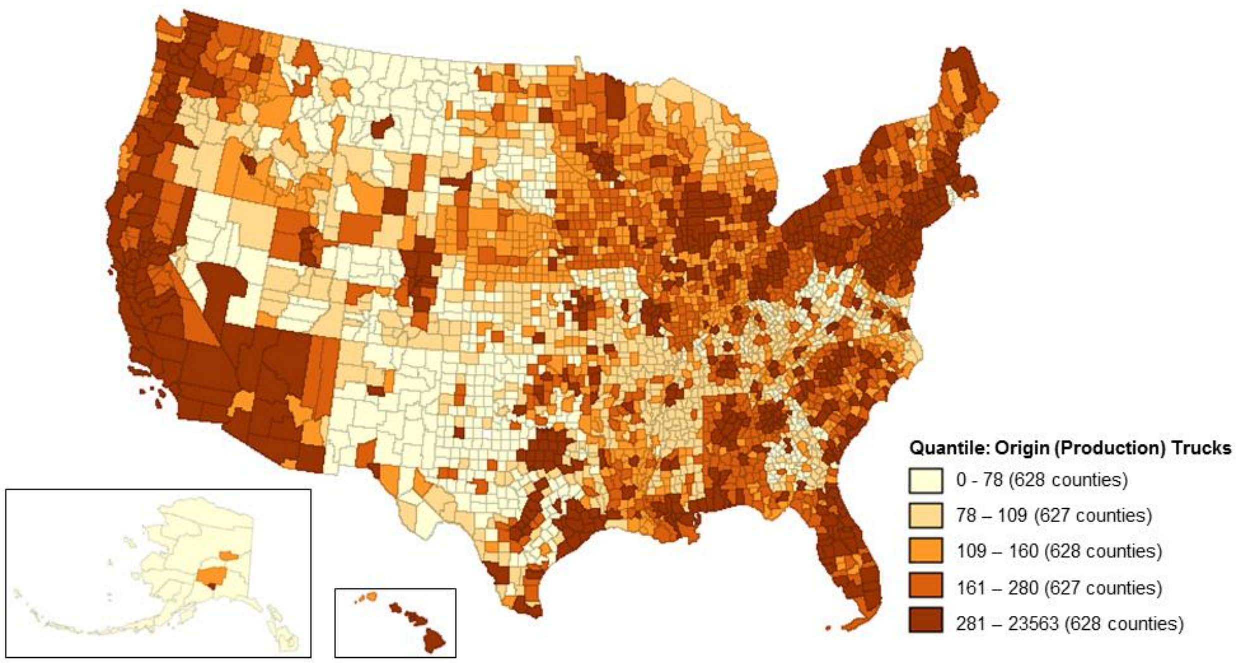

Figure 9: County-level Truck Trip Origins (Productions)

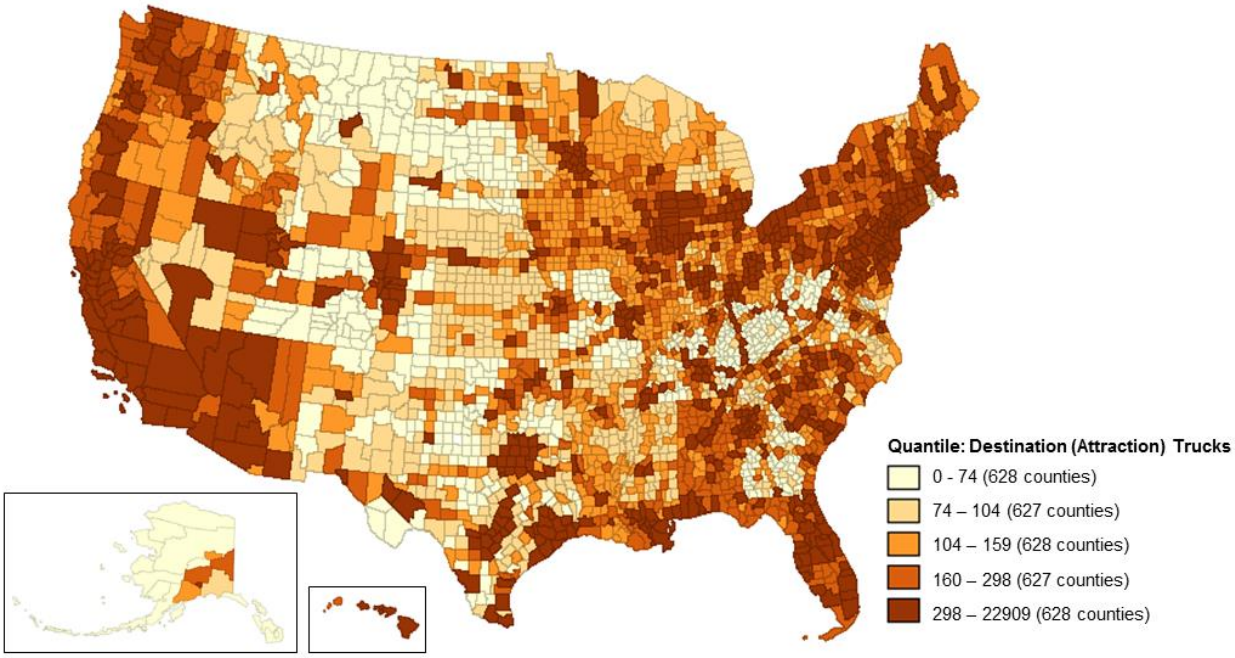

Figure 10: County-level Truck Trip Destinations (Attractions)

A more detailed analysis of Georgia and neighboring states shows a similar pattern with clustering around metropolitan areas. Figure 11 shows that the counties with the most trucks leaving them in Georgia are around Atlanta, Savannah, Columbus, and Macon. Truck origins in Georgia are centered more in highly urbanized counties compared with its neighbors, especially Florida, which has many high truck origins in rural and suburban areas. Truck destinations follow a similar pattern to origins, although the exact values and their distribution differs, as visible in Figure 12. Most importantly, the apparent patterns in these figures (Figure 9-12) also speaks to the importance of using a method that accounts for spatial autocorrelation, and justifies the use of the spatial error model for most commodities.

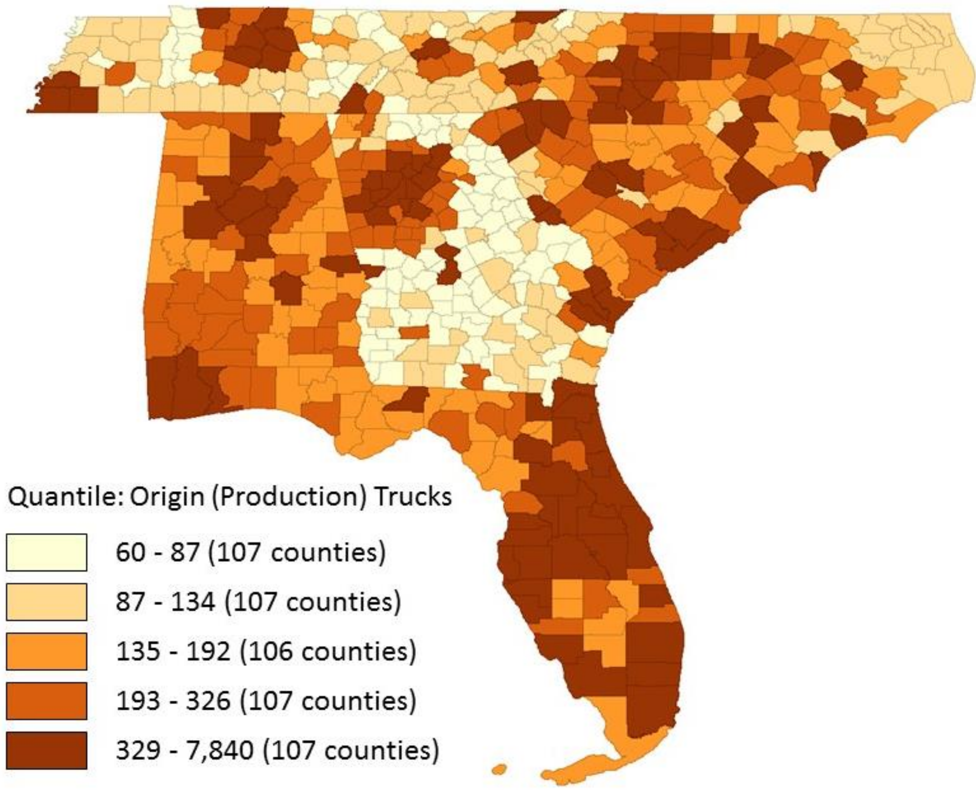

Figure 11: County-level Truck Trip Origins (Productions) for Select Southeastern Counties

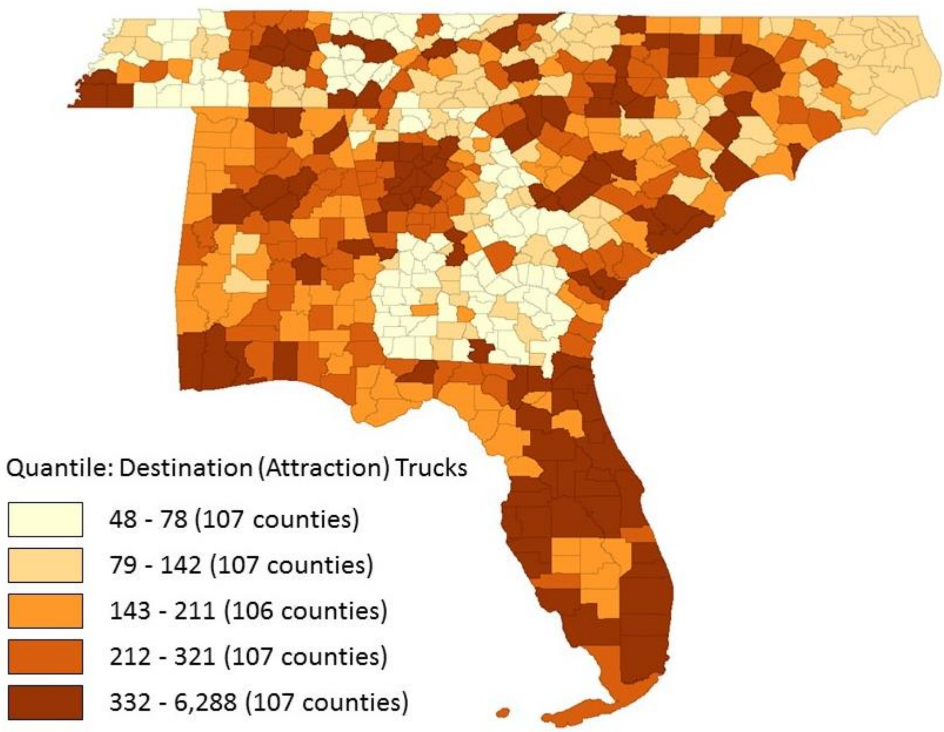

Figure 12: County-level Truck Trip Destinations (Attractions) for Select Southeastern Counties

In Georgia and neighboring states, the vast majority of counties have a very small number of truck trip productions (Figure 13) and attractions (Figure 14). Instead, there are just a few counties with productions or attractions exceeding 1,000 annual trips, and a very small number reaches up to nearly 8,000 truck trips (productions) or over 6,000 annual trips (attractions). This distribution is also indicative of the spatial concentration of high freight-producing economic activity.

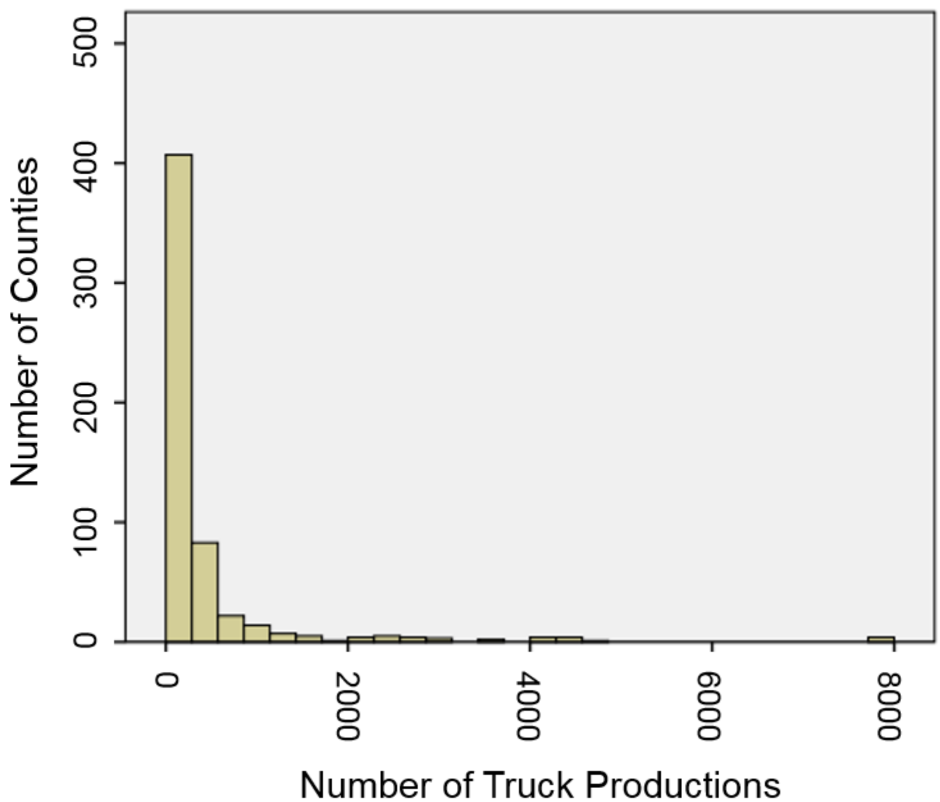

Figure 13: Histogram of Truck Production by County

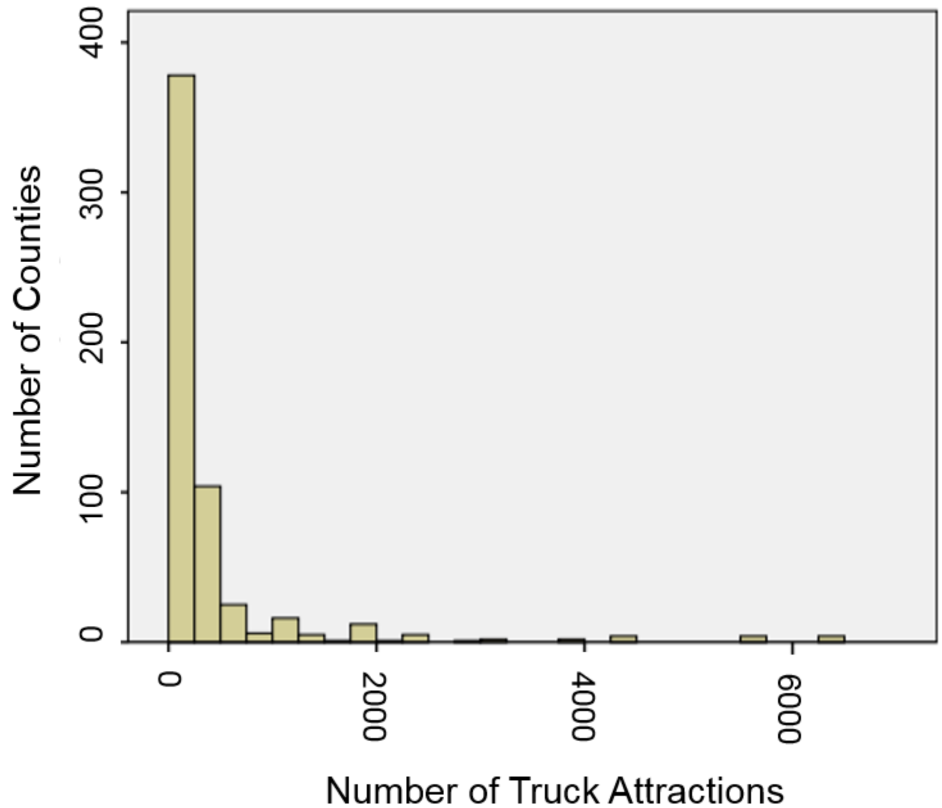

Figure 14: Histogram of Truck Attraction by County

The distribution of productions and attractions differs in that attractions are slightly more evenly distributed, although both southeastern U.S. productions and attractions are highly clustered. The median value for attractions is higher than for productions, while the maximum value is slightly lower (as shown in Table 8). However, the main takeaway is that the distribution is very uneven and displays broadly similar albeit not identical patterns among productions and attractions.

Table 8: Descriptive Statistics of County-level Truck Trips in Georgia and Neighboring States

	Truck Trips (Productions)	Truck Trips (Attractions)
N	570	570
Mean	448	438
Median	171	185
Std. Deviation	942	881
Minimum	60	48
Maximum	7,839	6,285

Output Data from Task 1

The result of task 1 is a nationwide county-level origin to destination truck trip volume table (The table is not included in report due to its very large size and is provided as separate file in a zipped folder). Figure 15 presents the format of the output and data from task 1. There are five columns, with the following definitions in the output data table.

- **O_State_County:** A unique identifier for the state and county where the truck is originating. The form of "1_101" means that the truck begins in the state with Federal Information Processing Standards (FIPS) code "01" (Alabama) and the county with FIPS code "101" (Montgomery County).
- **D_State_County:** A unique identifier for the state and county where the truck is destined. The form is identical to 'O_State_County."
- **O_CountyFIPS:** The FIPS code of the originating county.
- **D_CountyFIPS:** The FIPS code of the destination county.
- **Annual_Trucks:** The annual number of trucks forecasted to travel from the origin county to the destination county.
- **Daily_Trucks:** The daily number of trucks forecasted to travel from the origin county to the destination county.

O_State_County	D_state_County	trucks	OrigCountyFIPS	DestCountyFIPS
1_1	1_1	3.371769638	1001	1001
1_101	1_1	7.967435216	1101	1001
1_103	1_1	5.136949083	1103	1001
1_105	1_1	2.135786522	1105	1001
1_107	1_1	2.542175719	1107	1001
1_109	1_1	2.886212536	1109	1001
1_11	1_1	2.29242016	1011	1001
1_111	1_1	2.571469731	1111	1001
1_113	1_1	3.260474597	1113	1001
1_115	1_1	2.218617815	1115	1001
1_117	1_1	3.537042433	1117	1001
1_119	1_1	2.420980044	1119	1001
1_121	1_1	4.088638928	1121	1001
1_123	1_1	2.981736524	1123	1001
1_125	1_1	7.718043802	1125	1001
1_127	1_1	2.202140649	1127	1001
1_129	1_1	2.848354509	1129	1001
1_13	1_1	2.921990703	1013	1001
1_131	1_1	2.563940963	1131	1001
1_133	1_1	2.705305415	1133	1001
1_15	1_1	5.028566638	1015	1001
1_17	1_1	3.194436273	1017	1001
1_19	1_1	2.599097417	1019	1001
1_21	1_1	1.887167385	1021	1001
1_23	1_1	2.634566303	1023	1001
1_25	1_1	3.362343638	1025	1001
1_27	1_1	2.442232217	1027	1001
1_29	1_1	2.430376684	1029	1001
1_3	1_1	1.272338621	1003	1001
1_31	1_1	3.277736318	1031	1001
1_33	1_1	3.248490916	1033	1001
1_35	1_1	2.815641521	1035	1001
1_37	1_1	2.186683623	1037	1001
1_39	1_1	2.881508107	1039	1001

Figure 15: Format of Task 1 Deliverable

Result of Task 2: TAZ-Level Truck Trip Volume

Regression Results for Task 2

Task 2 also used the regression method to link truck trips with characteristics of the TAZ. Production and attraction models were both found to have a very strong fit with a set of three and four variables respectively: the square root of population, the square root of mean household income, the square root of the kilometers of railroad in the county, and for attractions the square root of the distance between the county centroid and the nearest segment of the NHFN (as summarized in Table 9). The variable transformation was used to improve model fit and meet regression assumptions that includes natural log of the dependent variable and the square root of all independent variables.

Table 9: Variables in Regressions Used for Task 2

Dependent Variable	Spatial Error Model (S) or OLS regression (O)	Explanatory Variables				R-Squared
Natural Log of Truck Trip Productions	S	Sqrt_Pop2010	Sqrt_HHInc	Sqrt_Rail_km		0.89
Natural Log of Truck Trip Attractions	S	Sqrt_Pop2010	Sqrt_HHInc	Sqrt_Rail_km	Sqrt_Dist_NHFN	0.86

Data Characteristics of Task 2

TAZ-level truck trips provide additional spatial detail than the county-level trips produced in task 1. Therefore, the TAZ-level productions and attractions follow the same macro patterns as task 1 and differ only at the sub county level. In Georgia, the TAZs that have higher truck trip productions and attractions compared with the rest of the county are more populated areas with higher incomes and proximity to one or several types of transportation infrastructure. The four variables, which were used in the disaggregation, are also reflected in the locations of TAZ-level productions and attractions. Moreover, the patterns of productions and attractions are very similar with only minor differences in scale between them. As illustrated in Figure 16 and Figure 17, almost all of the TAZs with a high volume of truck trip productions have a very similar volume of truck trip attractions.

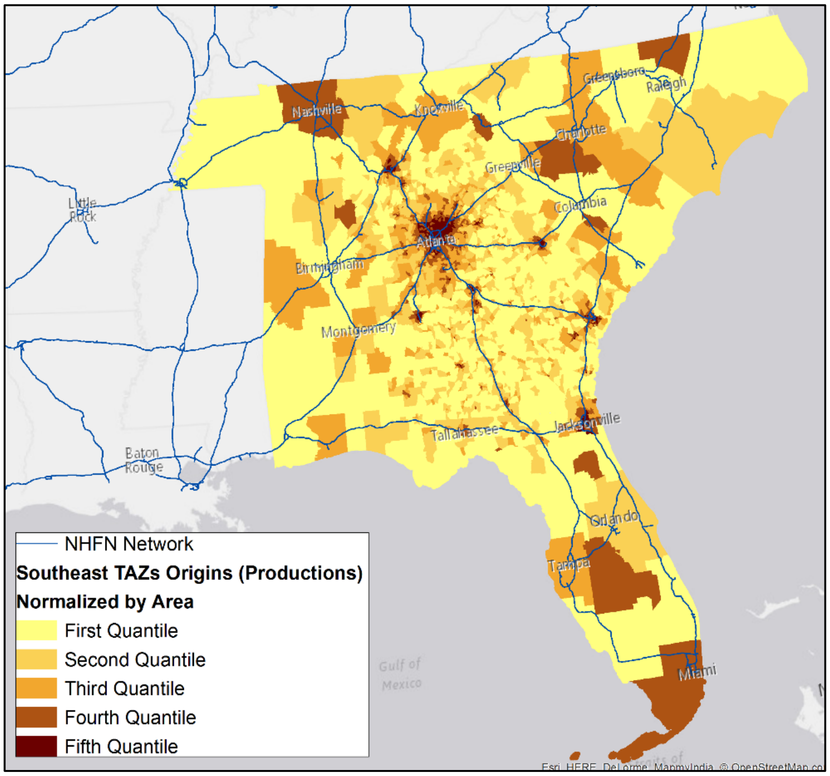

Figure 16: Southeast TAZs Origins (Productions) – Number of Trucks Normalized by Area

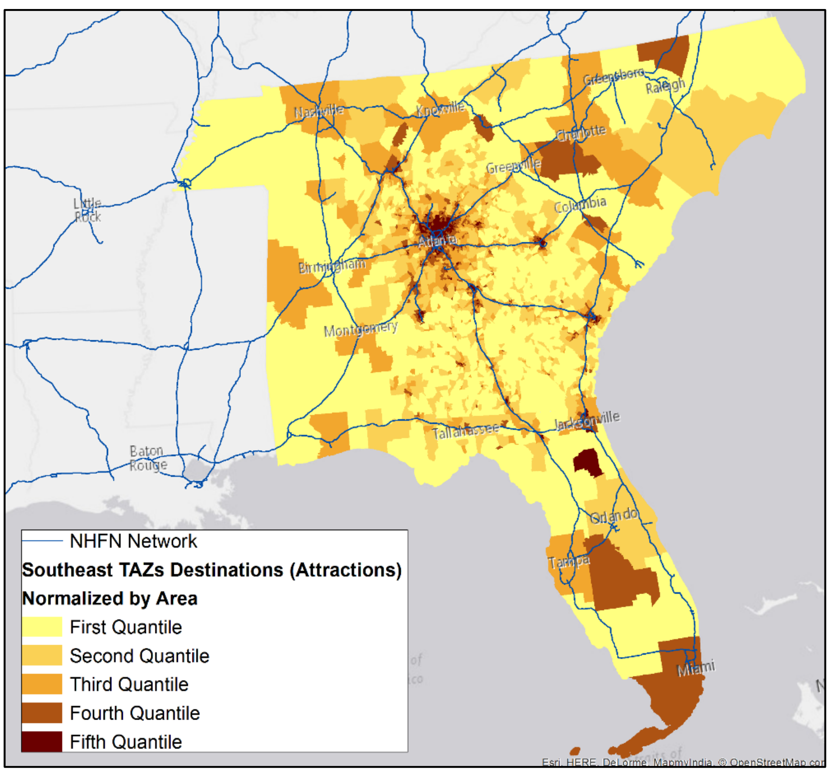

Figure 17: Southeast TAZ Destinations (Attractions) – Number of Trucks Normalized by Area

Output Data for Task 2

The results for task 2 follow the same format as for task 1. The final output table includes the following variables.

- **O_TAZ:** The origin TAZ number.
- **D_TAZ:** The destination TAZ number.
- **Annual_Trucks:** The estimated annual number of trucks from the origin to the destination TAZ.
- **Daily_Trucks:** The estimated daily number of trucks from the origin to the destination TAZ.

CHAPTER V. SUMMARY AND CONCLUSIONS

Summary

This study disaggregates FAF3 data from the FAF zone data to the county (task 1) and TAZ levels (task 2). It employs economic and sociodemographic data from the 5-year American Community Survey and the Decennial Census, transportation network locations from the U.S. Department of Transportation, tonnage-truck conversion from the Southern California Association of Governments, and geospatial boundaries from a variety of sources. The majority of these data types have the advantage of being publicly available, which facilitates future work.

Disaggregation relies on the regression method to establish relationships between commodity or truck trip productions and attractions, and logically related variables. These relationships are determined at the known level using OLS regression or spatial regression models depending on the level of autocorrelation. The regression are then applied to the disaggregation level to obtain each area's share of known productions and attractions. The higher an area's share, the more of the original freight flow is assigned to it. At the end of task 1, commodity tonnage is converted to truck trips, and the resulting county-to-county truck trip database serves as the input for task 2.

The output from both tasks is an O-D truck trip database among counties in all fifty states and the District of Columbia (for task 1) and GDOT's TAZs in the contiguous United States (for task 2). Total truck trips match those estimated via raw FAF3 data, and truck trips closely follow highly clustered development patterns. The study generates an O-D table (available as separate file) for each task explaining the number of truck trips between counties of USA and TAZs of Georgia.

Limitations and Further Study

This study generates disaggregated high-quality commodity movement data in the form of truck trip estimates at the county and TAZ levels according to conceptually strong and statistically valid relationships using economic, transportation network, and socioeconomic variables. This provides both a dataset that can be used immediately for understanding truck movements as well as a methodology that uses publicly available data that can be applied in the future to generate new datasets with updated data, new forecasts, or at other spatial levels.

Nonetheless, there remain features that would not ideally be present with perfect data sources. Among the limitations of this study are the year of the available data. As the Freight Analysis Framework version 4 (FAF4) is being released progressively over calendar year 2016, the commodity movement estimates were not available at the time calculations were performed. Therefore, the most recent FAF commodity movement data based on observations rather than forecasts is from 2007. However, this can be quickly updated with forecasted data and new observed data as FAF4 is released using the method adopted in this study.

The study employed the regression model to disaggregate commodity and truck origins and destinations. While the resulting models almost universally have high explanatory power as evidenced by high R-squared and alignment with expected relationships, the data distributions are such that many models show characteristics of heteroscedasticity and some deviation from normally distributed errors. This suggests that models may still be improved by adding variables that are not currently available or other possible variables transformations.

Travel demand modelers are familiar with the idea that some locations generate vehicle traffic out-of-proportion with their quantifiable characteristics. For freight, these ‘special generators’ include airports, seaports, and rail intermodal centers, among other sites. This study does not include special generators as their impact cannot be determined through data disaggregation without using truck counts or other data beyond the scope of this study. Therefore, conclusions about truck trips involving special generators should be reached by combining this study with truck counts focused specifically on the special generators.

In all cases, data for the same years was preferred. As the commodity movements are observed in 2007, all other variables were obtained for 2007 wherever possible. In some cases, 2007 data is not available, so the next closest year was substituted. For example, household income data is not available in complete form (i.e., without redacted data) at the county or lower level outside of the Decennial Census or the 5-year ACS estimates. Moreover, 5-year ACS estimates are the primary data source for block group-level data. These estimates average multiple years of data together, so they include 2007 data among other years. However, these slight changes in year are not expected to have a large effect on model output. Most economic, social, and demographic data are serially correlated, meaning that a previous year's value is a strong predictor of future values. In other words, the obtainable years are expected to closely resemble 2007 data. If the models attempted to provide causal explanations, having independent variables from a later year than dependent variables would be a concern because it would imply that the effect preceded the cause. By contrast, these models do not make causal assertions, but rather seek correlations useful in disaggregation, so the different years are not a major concern.

This study opens the door to future freight data disaggregation and analysis. One of the motivating factors in the report to which this study is a supplement was to analyze the impact of macro-scale supply chain changes onto the state economy and the transportation system. With truck and commodity movement data disaggregated to the county level, it is possible to study these supply chain changes in a new light. Moreover, this data allows for detailed studies about how Georgia counties, metropolitan areas, the state, or the megaregion of which Georgia is a part, trade among themselves and with other parts of the country. One side of a transaction can be aggregated (e.g., to a metropolitan level) while maintaining county-level specificity on the other side. The TAZ-level data also allows for detailed analysis of specific road projects and forecasted demands in Georgia to supplement the GDOT Statewide Travel Demand Model.

REFERENCES

Federal Highway Administration, 2007. Quick Response Freight Manual I and II. Freight Management and Operations (No. FHWA-HOP-08-010 (2007)).

NCFRP, 2013. NCFRP Report 26: Guidebook for Developing Subnational Commodity Flow Data. Transportation Research Board.

SCAG, 2012. SCAG Regional Travel Demand Model and 2008 Model Validation.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and econometric methods for transportation data analysis. CRC press.