2018

Fusing Freight Analysis Framework and Transearch Data: An Econometric Data Fusion Approach

Salah Uddin Momtaz
University of Central Florida

FUSING FREIGHT ANALYSIS FRAMEWORK AND TRANSEARCH DATA: AN ECONOMETRIC DATA FUSION APPROACH

by

SALAH UDDIN MOMTAZ
B.Sc. Bangladesh University of Engineering and Technology, 2010
M.S.C.E. University of South Alabama, 2015

A dissertation is submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2018

Major Professor: Naveen Eluru

ABSTRACT

A major hurdle in freight demand modeling has always been the lack of adequate data on freight movements for different industry sectors for planning applications. Freight Analysis Framework (FAF), and Transearch (TS) databases contain annualized commodity flow data. The primary motivation for our study is the development of a fused database from FAF and TS to realize transportation network flows at a fine spatial resolution (county-level) while accommodating for production and consumption behavioral trends (provided by TS). Towards this end, we formulate and estimate a joint econometric model framework grounded in maximum likelihood approach to estimate county-level commodity flows. The algorithm is implemented for the commodity flow information from 2012 FAF and 2011 TS databases to generate transportation network flows for 67 counties in Florida. The data fusion process considers several exogenous variables including origin-destination indicator variables, socio-demographic and socio-economic indicators, and transportation infrastructure indicators. Subsequently, the algorithm is implemented to develop freight flows for the Florida region considering inflows and outflows across the US and neighboring countries. The base year models developed are employed to predict future year data for years 2015 through 2040 in 5-year increments at the same spatial level. Furthermore, we disaggregate the county level flows obtained from algorithm to a finer resolution - statewide transportation analysis zone (SWTAZ) defined by the FDOT. The disaggregation process allocates truck-based commodity flows from a 79-zone system to an 8835-zone system. A two-stage factor multiplication method is proposed to disaggregate the county flow to SWTAZ flow. The factors are estimated both at the origin and destination level using a random utility factional split model approach. Eventually, we conducted a sensitivity analysis of the

parameterization by evaluating the model structure for different numbers of intermediate stops in a route and/or the number of available routes for the origin-destinations.

Dedicated to my beloved wife, Raha for being with me.

ACKNOWLEDGMENTS

I would like to thank Dr. Naveen Eluru for his recurrent guidance. I am glad for the opportunity of working on his research project sponsored by the Florida Department of Transportation. His mentoring enlightens me to develop me as a better researcher. Without financial support from his research, I might not end up with my graduate research. I express my gratitude to the funding agency for supporting the research project. The advice in the project from Dr. Sabreena Anowar improved the dissertation to a better shape. It is an honor to thank her for her help.

I am also grateful to God for this beautiful life with nice people around me. I am thankful to all those nice people including, my, parents, wife, family, friends, colleagues, and acquaintances, who are continuously supporting me in the way of my life. Especially, I am glad to have Mr. S. M. Momtaz Uddin Mia and Mrs. Khadiza Momtaz as my parents, who struggled all their life to facilitate me a better life. I am proud of them. I really express my gratitude to my wife, Eng. Sk Rahatul Jannat, for being there for me during my whole time during my Ph.D. in Florida.

I would like to extend my thanks to Dr. Min Wook Kang, for his mentoring during my masters. His guidance showed me the path for being a researcher. I will always remember him as my first guardian in the US. I would also like to thanks Dr. H. M. A. Aziz for his supports as a mentor in my very first research. Just a "thank you" will not be enough for my one of the closest friends, Dr. Dipesh Das, for his advice during the university selection during my graduate studies in the US. I would like to mention Dr. Moshuir Rahman and Mrs. Evana Ahmed for sharing a wonderful time in Orlando. Last but not least, I would like to extend my gratitude to another friend, Dr. Kollol Shams, for sharing his experience of searching for a job after the Ph.D.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 Background and Motivation

Freight movement is a defining aspect of a region's economic viability and livability. A region's economy substantially benefits from increased intra-and inter-regional freight flows between different trading partners and intermodal centers (e.g., ports, intermodal logistics centers). Implementation of strategies that support efficient freight movement is therefore essential not only for attracting new industries to move freight within, into, and out of the region but also for addressing the needs of existing businesses. The strategies should also take into account for the fact that increased movements bring challenges associated with added stress on already congested transportation networks and negative impacts to air quality. To address these challenges, detailed data on freight movements would provide a greater understanding of freight patterns and its impacts on the transportation network.

Florida is currently the third largest state by population in the United States with over 20 million residents. According to Viswanathan et al. (2008), between 2001 and 2030, population and employment in the state of Florida is predicted to increase by 46.5% and 110%, respectively. Understandably, freight transportation will also grow over time with the expansion of population and economic activity within the state. Hence, the issue of efficient freight movement is gaining increasing importance at all levels of government in the state. Towards better understanding the freight flows in Florida, Florida Department of Transportation (FDOT) has been at the forefront of acquiring and investigating new data sources for freight planning applications. However, freight

movement data comes in many different forms, from many different sources (public or proprietary), with varying temporal and spatial resolutions, with substantial differences in the sampling and/or data collection methods. To be sure, each data source contains a wealth of information, but each has its own sets of strengths and weaknesses. Therefore, instead of relying on a single source of data for modeling and other applications, a smarter approach would be to take advantage of data fusion techniques to create a fused data with expanded scope of information and then use it for planning and forecasting purposes. In this study, our goal is to link different Florida specific freight movement data sources using appropriate matching criteria to gain an in-depth insight on the full continuum of freight movement issues in the state.

1.2 Research Context

In recent years, there is growing recognition among travel demand modelers that freight planning is an important exercise for overall travel demand forecasting. Traditionally, the travel demand forecasting field has focused on estimating passenger travel demand. In the passenger realm, trip based, or activity-based travel demand models have been developed using household travel surveys conducted in urban regions. Compared to passenger travel demand, a major hurdle in freight demand modeling has always been the lack of adequate data on freight movements for different industry sectors for planning applications. The current research effort is geared towards addressing the data availability challenge through an innovative econometric methodology for data fusion.

Several data sources are available for freight planning purposes in the United States. Of these, the most commonly adopted sources include Freight Analysis Framework (FAF), Transearch (TS), American Trucking Research Institute (ATRI) truck route data, and Department of Transportation (DOT) weigh-in-motion (WIM) data. FAF and TS databases contain annualized commodity flow data that can be used in long range freight forecasting. FAF database is prepared based on the Commodity Flow Survey (CFS) conducted periodically. It is freely available to the public and can be downloaded from the Federal Highway Administration (FHWA) website (FHWA, 2012). It provides freight flows (by weight, value and mode) for 43 commodity types classified by Standard Classification of Transported Goods (SCTG 2-digit) code. The commodity flow information is available at a very coarse spatial resolution - 132 domestic zones across the United States and 8 foreign zones. The baseline year for current FAF data (FAF4) is 2012 and includes forecasts on freight flows between 2015 and 2045 at a 5-year interval.

Transearch database, a proprietary product developed by IHS Global Insight, provides detailed information on freight flows (by weight, value and mode). The database is constructed from various commercial and public sources including: Annual Survey of Manufacturers (ASM), Surface Transportation Board (STB) Rail Waybill Sample, Army Corps of Engineers Waterborne Commerce data, Federal Aviation Administration (FAA), Enplanement Statistics, and Airport-to-airport cargo volumes. However, the algorithm used to generate the final data product is not publicly available. The freight flows in TS are reported by commodity type based on the Standard Transportation Commodity Code (STCC) in more than 500 categories. The data can be purchased at a fine spatial resolution (such as county level). However, the database is expensive to acquire and requires substantial investment from transportation agencies.

Although both FAF and TS provide annual commodity flows in the United States, several differences exist between these sources. The most obvious difference arises from the variability in data collection mechanism employed; FAF relies on processing commodity flow data (such as CFS 2012) while TS employs various sources of data to generate county level flows using a proprietary algorithm. A second difference arises from what the commodity flows in each dataset represent. FAF flows represent actual transportation network flows while TS flows represent production-consumption commodity flow. To illustrate the difference, consider that X units of a commodity is shipped from location A (production zone) to location B (consumption zone) through an intermediate location C. The FAF flows would represent these flows as X units from A to C and X units form C to B. On the other hand, in TS, these flows are only represented as X units from A to B. Thus, FAF would report a total tonnage of 2X units transferred while TS would report only a transfer of X units. A more general summary of data sampling procedures in FAF and TS is presented in Figure 1. From the figure, it is evident that FAF flows are potentially sampled at more intermediate points such as warehousing locations while TS flows are considered only at origin and destination.

For understanding transportation network usage measured through network flows, FAF is a more appropriate database as the reporting is based on realized network flows. On the other hand, the flows represented in the TS database are annual production-consumption measures from the TS defined regions and do not represent the actual transportation network path flows. To be sure, there is significant value in understanding production and consumption trends to develop a behavioral framework of freight commodity flows in the future.

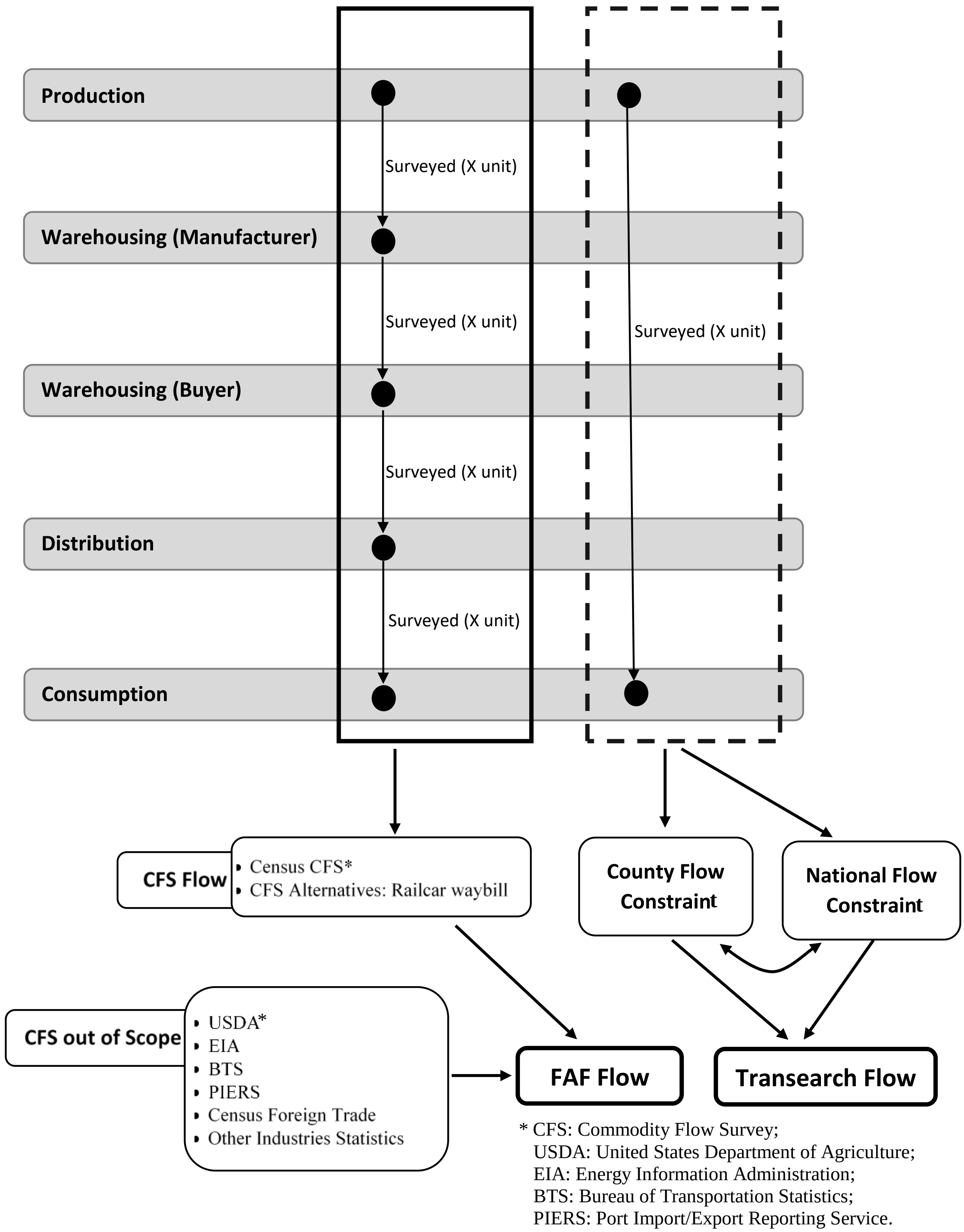


Figure 1 FAF and Transearch data collection methods and dataset generation

In terms of cost, FAF data is freely available while TS database is an expensive database and the algorithm employed is inaccessible to users. The commodity type definition across the two datasets is also different – 43 commodity types in FAF and over 500 commodity types in TS. Finally, the coarser spatial and commodity type resolution in FAF makes it challenging to generate reliable network flow estimates. While TS provides data at a fine spatial and commodity type resolution, the supply demand nature of the database requires additional analysis to realize transportation network flows. Overall, the comparison of the databases highlights the inherent strengths and weaknesses of the two databases.

1.3 Objective of the Dissertation

The major goal of the dissertation is to mitigate the above stated challenge of the absence of comprehensive data for freight transportation planning. The dissertation proposes an econometric data fusion algorithm to generate a comprehensive dataset using existing datasets. To achieve the goal of the dissertation, we focus on the following five specific objectives:

Objective I: *Development of Econometric Data Fusion Algorithm.*

The dissertation aims to develop a first of its kind fusion algorithm employing Freight Analysis Framework (FAF) and Transearch (TS) data to realize transportation network flows at a fine spatial resolution (county level) while accommodating for production and consumption behavioral trends (provided by TS). Towards this end, the dissertation formulates and estimates a

joint econometric model framework grounded in maximum likelihood approach to estimate county level commodity flows.

Objective II: *Implementation of Proposed Algorithm.*

The dissertation implements the proposed algorithm to generate the fused freight database for the state of Florida at the county level resolution for the base year 2011. The algorithm is also implemented to predict future year data for years 2015 through 2040 in 5-year intervals at the same spatial level.

Objective III: *Data Generation for Finer Spatial Resolution of Origin and Destination.*

The Florida Department of Transportation freight model employ freight truck flows at the resolution of state-wide transportation analysis zone (SWTAZ). In this dissertation, we propose a methodology to disaggregate the fused flow database to a finer spatial resolution by considering the disaggregation of flows from county level to SWTAZ. The disaggregation process allocates truck-based commodity flows from a 79-zone system (67 counties + 12 external zones) to an 8835-zone system.

Objective IV: *Sensitivity of Route Choice in the Algorithm.*

The proposed algorithm employs network routing to fuse the datasets. In the original approach, we consider travel distance for each path as the independent variable affecting route selection probability. We enhance the proposed approach by undertaking a comprehensive

sensitivity analysis to evaluate other alternative functional forms for the model structure and variables.

1.4 Outline of the Dissertation Proposal

The remainder of the dissertation proposal consists of seven chapters. The chapters are organized as follows.

CHAPTER Two: presents a comprehensive review of existing literature on freight data generation at a disaggregate spatial level. The chapter outlines the different alternative methodological frameworks and presents details of the exogenous variables used in these studies. Further, we highlight the limitations of existing work while positioning the current study in context.

CHAPTER Three: describes the datasets employed for the data fusion approach. The chapter presents the descriptive statistics and the characteristics of the datasets. A detailed comparison of the datasets and the inherent strengths and weaknesses are also discussed.

CHAPTER Four: presents the mathematical details of the proposed algorithm (contributing to Objective I). The proposed formulation is applied to the Florida data for truck mode. The data preparation procedures for the algorithm are discussed, as well. The chapter concludes with the results from the proposed data fusion algorithm along with validation of outputs for the within state truck flows.

CHAPTER Five: discusses an extension of the proposed algorithm from only within Florida flows to the entire country and neighboring foreign regions (contributing to Objective II).

Further, the application of the fusion approach for future years from 2015 through 2040 in five-year increments is also presented.

CHAPTER Six: shows an outline of the disaggregation of the flows to a statewide traffic analysis zone resolution along with the prediction for future years. (contributing to Objectives III).

CHAPTER Seven: describes the sensitivity of the route choice with the different functional form of the model framework. (contributing to Objectives IV).

CHAPTER Eight: finally concludes the dissertation and discusses the remarks on the research outcome. This chapter also explains the limitation and the future avenue of the research.

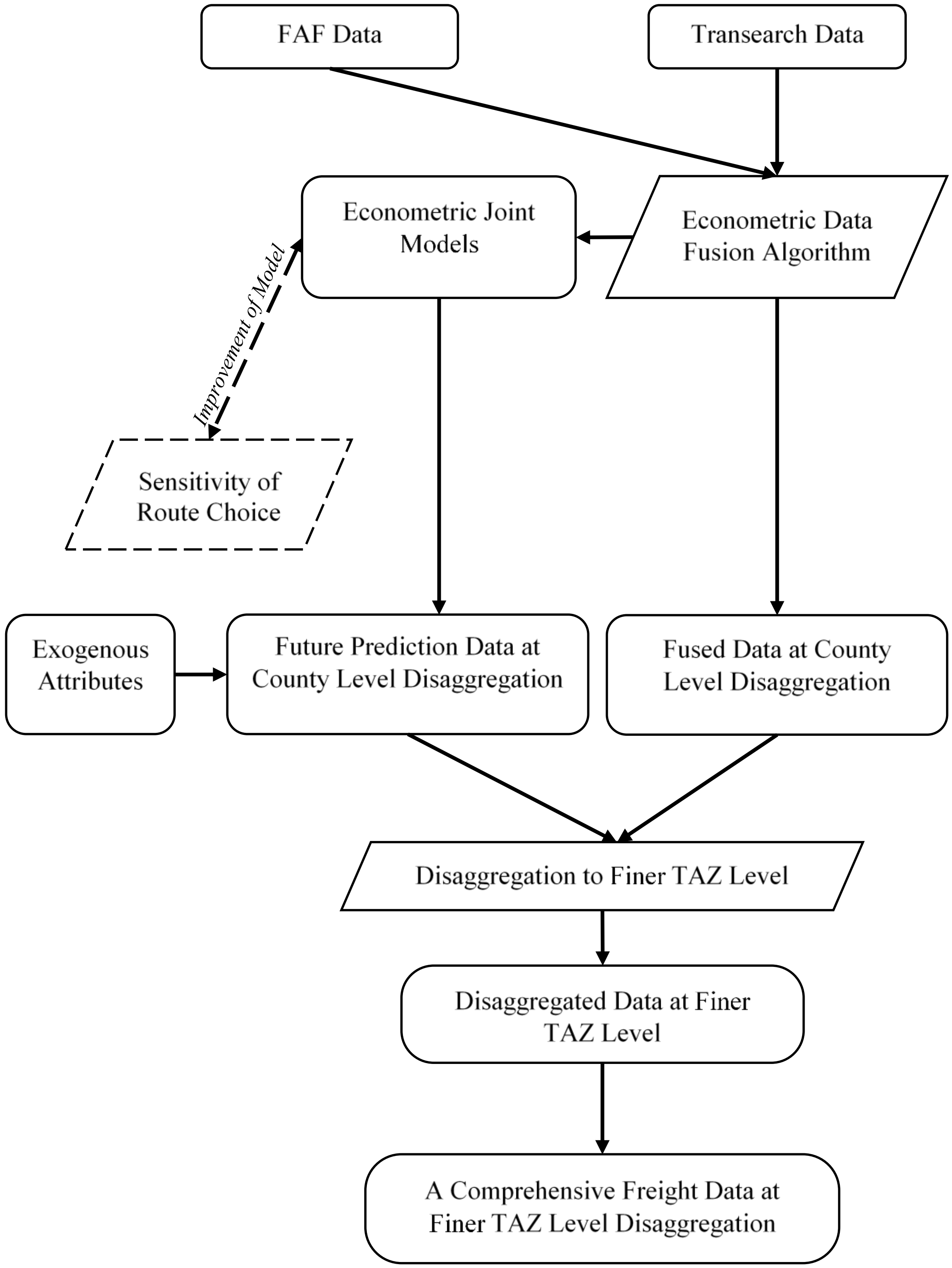


Figure 2 Outline of Dissertation Objectives and Specific Goals

CHAPTER TWO: LITERATURE REVIEW

2.1 Existing Literature on Disaggregation of FAF Flow

Several research efforts have attempted to address the spatial resolution challenge with FAF data. A summary of earlier literature summarizing freight data merging and/or disaggregating efforts is provided in Table 1. The table provides information on the study area, datasets employed, objective of the research effort, modeling methodology, and exogenous variables considered.

Several observations can be made from the table. First, the majority of the studies developed a procedure for disaggregating FAF data from the FAF zone level to a county level or traffic analysis zone (TAZ) level. Second, several states in the US have developed disaggregation procedures including Texas, California, New Jersey, Wisconsin, Georgia, North Carolina, and Florida. Third, the various methods considered to disaggregate FAF flows include: (i) proportional weighting method (applied for socio-economic variables or vehicle miles traveled (VMT)), and (ii) statistical methods. In the proportional weighting method, a "disaggregation factor" is estimated using various socio-economic variables (such as employment and population), land use and truck VMT variables by computing the ratio of the variables of interest at the disaggregate spatial resolution and aggregate spatial resolution. Using these factors, the freight flow allocation to the disaggregate spatial resolution is made. The disaggregation factors are considered to vary based on the type of origin and destination spatial configuration (such as internal – internal zonal pair, external – internal zonal pair). The statistical methods considered in freight modeling include linear or log-linear regression, structural equation modeling, economic input output models, and

fractional split methods that employ socio-economic and demographic variables such as employment and population as exogenous variables. The models developed are employed to generate freight flows at the desired disaggregate spatial resolution. The models are typically validated by aggregating freight flows at the finer resolution and comparing it to the observed flows at the aggregate resolution. Fourth, in disaggregation studies, the variable of interest includes tonnage, value and/or ton-miles. Finally, the variables considered to be of significance in the data merging process include employment, population, travel time and cost, business establishments, and transportation system characteristics.

Table 1 Review of Earlier Studies on Disaggregation of FAF

Study	Geographic Region	Dataset(s) Used	Research Objective	Methodology	Variables Used
Giuliano et al., 2010	Los Angeles	CFS, IMPLAN, WISERTrade, WCUS, SCAG	Estimate link specific truck flows	Data integration; I-O model; gravity model; user optimal network assignment	Small area employment data
Bujanda et al., 2014	Texas	FAF^3, Transborder freight flow, Maritime flow	Estimate state level flows from FAF^3 (import and export flows)	ArcGIS spatial analysis; network assignment	-
Aly and Regan, 2014	California	FAF^2	Disaggregate FAF commodity flow at the county level	Proportional weighting for both origin and destination	Truck VMT
Opie et al., 2009	New Jersey	FAF^2, Transearch (for validation)	Disaggregate FAF commodity flow at the county level	Proportional weighting	Total land area occupied by port (import and export flows); for domestic flows: commodity-specific employment, truck VMT, total employment, population
Ranaiefar et al., 2013	California	FAF^3	Develop structural commodity generation model at the FAZ level	Structural equation model	Employment, number of establishments, population, agriculture related variables (farm acreages), manufacturing sector GDP, energy-related data (refinery capacity)
Mitra and Tolliver, 2009	North Dakota	FAF^2, truck count data (validation)	Disaggregate truck flows (productions and attractions)	Proportional weighting (production); I-O model (attraction) gravity model (internal flow)	Two-digit NAICS employment count

Vishwanathan et al., 2008	Florida	FAF[2], Transearch (output cross-check)	Disaggregate FAF commodity flow at the county level	Proportional weighting; linear regression	Total employment, population, two/three-digit NAICS employment count
Ruan and Lin, 2010	Wisconsin	FAF[2], Transearch (validation)	Comparison of different data synthesis method for disaggregating FAF flows	Proportional weighting; direct regression; optimal disaggregation model	Employment by industry type, number of intermodal facilities
Ross et al., 2016	Georgia	FAF[3], CBP, Census data	Disaggregate FAF flows to county and TAZ level	Spatial regression; proportional weighting	Three-digit NAICS employment count, population, freight network density
Oliveira-Neto et al., 2012	USA	FAF[3], CFS (validation)	Disaggregate FAF flows at the county level; estimate ton-mile by mode	Log-linear regression; gravity model	Total employment payroll
Sorratini and Smith, 2000	Wisconsin	CFS, Transearch	Disaggregate truck flows at the TAZ level	I-O model	Employment
Lim et al., 2014	California	FAF[3], FAF[2], Transearch (validation)	Disaggregate FAF flows at the county level	Linear regression	Population, employment, farm acreage and crop sales

2.2 Existing Literature on Statewide Freight Models

The current research effort is geared towards developing a statewide freight planning framework while incorporating the influence of flows originating outside the study region (rest of US and international flows). In transportation literature, several regional models have been developed for such freight planning purposes in the US and Europe. In this section, we provide a brief review of the most relevant studies for our research.

2.2.1 Regional Models

In the US, multiple states developed customized statewide freight models including Florida, Indiana, Kentucky, New Hampshire, New Jersey, Virginia, Ohio, Alabama, Oregon, Wisconsin, Massachusetts, and Missouri (2, 3, 9–12). A summary of statewide model development efforts across the country is presented in Table 2. The table provides information on study region, study year, input data source, methodology adopted, spatial resolution of outputs, if flows outside the state were considered, if foreign flows were considered, if model was validated, if future flows were predicted. Several observations can be made from the table. First, the most commonly used data sources used for developing statewide models include Transearch, FAF, Commodity Flow Survey (CFS), and ATRI data. Only one study (6) employed multiple data sources as input for obtaining freight flows. Second, the spatial resolution levels considered in these frameworks include county, three-digit zip code, and statewide traffic analysis zones (STAZ). Third, for majority of the studies, external flows consisted of flows from rest of US while the others considered flows from foreign regions. Fourth, the most common validation datasets employed

include truck vehicle miles travelled, truck counts, CFS, and Transearch. Fifth, among all earlier studies, only Eatough et al. (2) developed estimates for future years for Virginia. Specifically, the study projected freight truck trips for a 7-year horizon using a growth factor approach. Finally, the approaches employed in these frameworks range from simple proportional allocation (or disaggregation factor) approaches, trip-based models, input-output models, fractional split models, to behaviorally motivated data-fusion approaches.

In Europe, several national freight models have been developed for various countries. While these are national freight models, the approach employed might be comparable to the statewide models for larger states in the US. The readers are referred to (13) and (14) for a detailed review of these freight models. Of particular relevance to our effort is a study by Ben-Akiva and de Jong (15) that describes a framework implemented for Norway, Sweden, and Denmark. In this effort, the authors develop a multi-stage model system that considers (1) production consumption flows at an aggregate level, (2) shipment size and transport chain decisions at the disaggregate level, and (3) assignment routine at an aggregate level. The resolution of analysis is driven by data availability. The authors indicate that a disaggregate shipment size and transport chain decision process is feasible due to the availability of such data in Europe. However, with the data availability challenge for freight transportation in the US, such frameworks are not readily applicable to the US context.

Table 2 Review of Earlier Studies on Statewide Freight Models

Study	Input Data	State	Spatial Resolution	Year	Flows from rest of the country considered (Yes/No)	Exports/ Imports considered (Yes/No)	Data validation source	Future Year Prediction (Yes/No)	Methodology used
Aultman-Hall et al. 2000	Transearch 1998 data	KY	County & 3-digit Zip Code for KY and four adjacent states; each state as one and international zone as one external zone	1998	Yes	Yes	None	No	Disaggregation factor
Anderson et al. 2007	Freight flow data from multiple surveys	AL	County	1997	Yes	No	None	No	Four Step Model, Regression Analysis
Xiong et al. 2997	Transearch 2003 data	FL	County	2003	Yes	Yes	Truck VMT	No	Gravity Model
Jin et al. 2012	FAF	UT	County	2009	No	No	CFS 2002	No	Enhanced Gravity Model
Eatough et al. 2000	CFS	VA	First BEA, then County	1993	Yes	No	None	Yes (7 Years Projection)	Use of conversion factor for BEA, then Gravity Model to distribution county level flow
Krishnan, and Hancock 1998	CFS	MA	3-digit Zip Code	1993	Yes	No	Highway Performance Monitoring System for MA (% commercial vehicles)	No	Capacity Restraint Link flow with iterative process of user equilibrium assignment

Sivakumar and Bhat, 2002	Transearch	TX	county	1996	Yes	Yes (Mexico)	None	No	Fractional Split for distributing flows
Bernardin et al. 2011	ATRI	IN	Statewide TAZ for IN	2010	Yes (up to 50 miles beyond state boundary)	No	Number of Freight trip from Indiana Statewide Travel Demand Model	No	Four-step model
Opie et al. 2009	FAF	NJ	County	2002	Yes	Yes	Transearch data (2001)	No	Disaggregation Factor
Mitra and Tolliver, 2009	FAF	ND	Statewide TAZ	2002	Yes	Yes	Truck VMT (NDDOT)	No	Input-Output model
Huang and Smith, 1999	CFS	WI	Statewide TAZ	1993	Yes	No	Truck trip count from video data	No	Gravity Model; Select Link Analysis
Jones and Sharma, 2002	CFS	NE	County	1992	No	No	Truck Trip	No	Input-output model

2.3 Current Study Context

Based on the literature review, it is evident that multiple research efforts have considered disaggregation of FAF commodity flow to a finer spatial resolution such as county TAZ. While the disaggregation is of immense value, the approach employed is purely a factoring exercise without any attempt to address production consumption relationships. FAF data inherently does not provide production consumption relationship and hence, using FAF alone to arrive at production consumption flows is not possible. To be sure, earlier research employed TS flows for evaluating FAF disaggregation outputs for validation purposes (Opie, Rowinski, and Spasovic, 2002; Viswanathan, et. al., 2008; Ruan, and Lin, 2010). In our study, we enhance earlier research attempts by developing a fusion framework that disaggregates FAF flows while accounting for production consumption relationships observed in TS.

The primary motivation for our study is the development of a fused database to realize transportation network flows at a fine spatial resolution (county-level) while accommodating for production and consumption behavioral trends. Thus, we undertake disaggregation of FAF flows while augmenting with production consumption-based TS flows. Towards this end, we formulate and estimate a joint econometric model framework grounded in maximum likelihood approach to estimate county level commodity flows. The framework has two separate modules to ensure matching estimated county level flows with commodity flows in FAF and TS at the appropriate spatial resolution. A third module generates a behavioral connection between FAF and TS. In our algorithm, we connect the flows between TS and FAF by generating potential paths between the origin and destination of interest for TS flows. Note that the inherent differences in the data cannot

be completely reconciled. Hence, the framework focuses on building a fused database that maximizes the match with the commodity flows in the two databases. The consideration of behavioral trends in the model framework can assist us in parameterizing TS flow relationships thus allowing us to circumvent TS for the future (if needed). The proposed algorithm is implemented for the commodity flow information from 2012 FAF data for five FAF zones and 2011 TS databases for 67 counties in Florida.

Subsequently, we extend the algorithm proposed to consider spatial regions within Florida, outside Florida (in the US and foreign regions). Thus, we address the major limitation of the proposed approach for freight flow predictions. Further, we develop a long-range planning framework for the state of Florida by generating freight flows for the base year at the county level resolution. The model accommodates for multiple spatial resolutions including a county level resolution inside Florida, eleven external zones outside Florida and one external zone for Canada, Mexico, and Alaska. Further, of particular importance to planning agencies, we implement our algorithm to generate flows for future years. The current study generates estimates for freight flows from 2015 through 2040 in five-year increments. The model was implemented for thirteen commodity type classifications (more on the Chapter Five).

Finally, we extend our proposed outputs and develop disaggregate estimation of freight flows at a statewide traffic analysis zone (SWTZ) level. The disaggregation from county level flows to SWTAZ level flows is accomplished using a fractional split approach at the origin and destination level. The fractional split models are estimated for all commodity types and employed for developing base year estimates as well as future year estimates.

CHAPTER THREE: DATA DESCRIPTION

The previous chapter discussed earlier research relevant to the present research. In this chapter, we focus our attention on the different datasets employed for our analysis.

3.1 Freight Analysis Framework (FAF) Dataset

This dissertation used Version 4 of the FAF database (FAF4). The data specific to the state of Florida was extracted and prepared for analysis. FAF4 provides freight flow information for tonnage, value, and domestic ton-miles by region of origin and destination, commodity type, and mode. The baseline year is 2012 and forecasts on freight flows until 2045 are available, starting from 2013, 2014, 2015 and then at a 5-year interval. In terms of the geographic dimension, the FAF4 data provides freight trading between 132 domestic zones and 8 foreign zones; five of which are in Florida: Jacksonville (121), Miami (122), Orlando (123), Tampa (124) and the rest of Florida to FAF region (129). The Zones are shown in the Figure 3. In terms of commodity classification, FAF4 reports freight flows using the same 43 2-digit Standard Classification of Transported Goods (SCTG) classes, as reported by the Commodity Flow Survey (CFS).

The flows are defined in various categories including the Domestic flow includes flows that originated and terminated within Florida, that originated in Florida but destined to regions outside Florida within the USA, and that originated in regions outside Florida within the USA but destined to Florida. Export refers to the freight volume traveling to foreign regions outside the USA from Florida while import refers to inbound flow of freight from foreign regions outside of the USA to Florida.

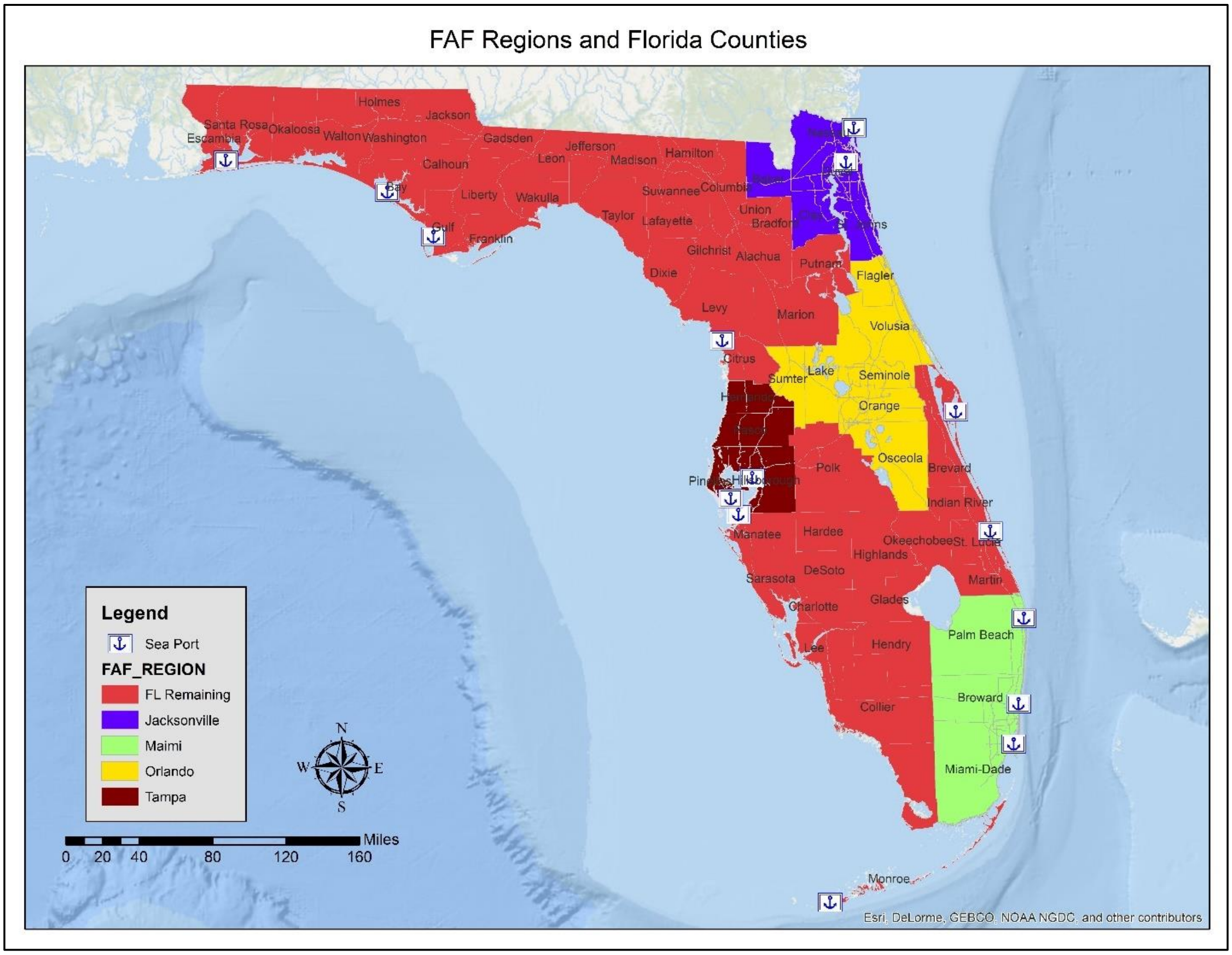


Figure 3 FAF and Transearch TAZ

3.1.1 Tonnage Share Analysis

In 2012, 706 million tons of freight valued at $903 billion moved into, out of, within the Florida region via its roads, railroads, waterways, and air freight facilities. Table 3 displays freight flows by weight, value and direction for 2012. The following observations can be made from the Table.

- Domestic freight accounted for 639 million tons or nearly 91 percent of the total tonnage valued at $686 billion. More than 39 million tons (2.73%) were exported while 40 million

tons (5.59%) were imported to and from the foreign regions. The total value of the exported ($69.56 billion) tonnage was higher than the imported tonnage ($71.99 billion).

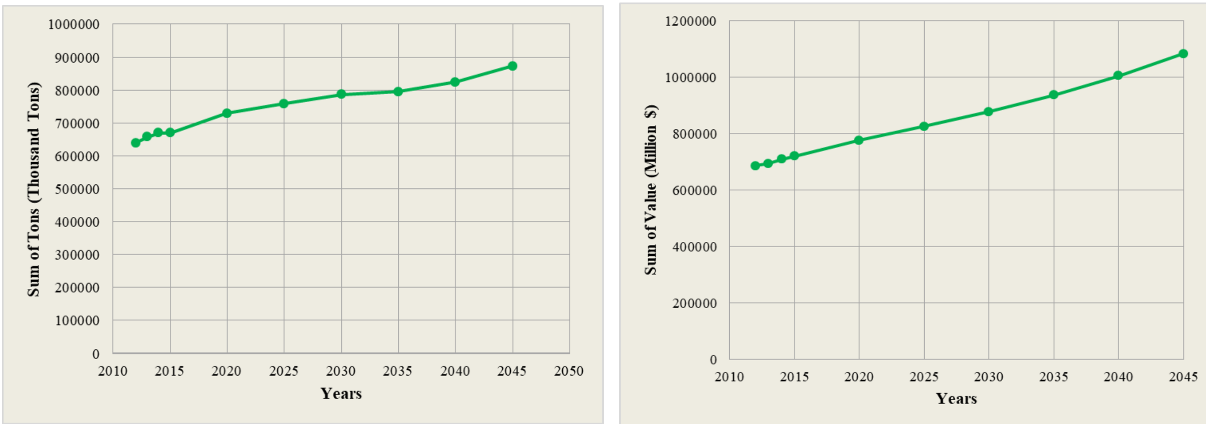
- Intrastate volumes (Florida-Florida) represented the largest group in terms of total tonnage shipped (approximately 469 million tons) followed by inbound volumes from the rest of USA (approximately 124 million tons).

Table 3 Total Tonnage by Direction

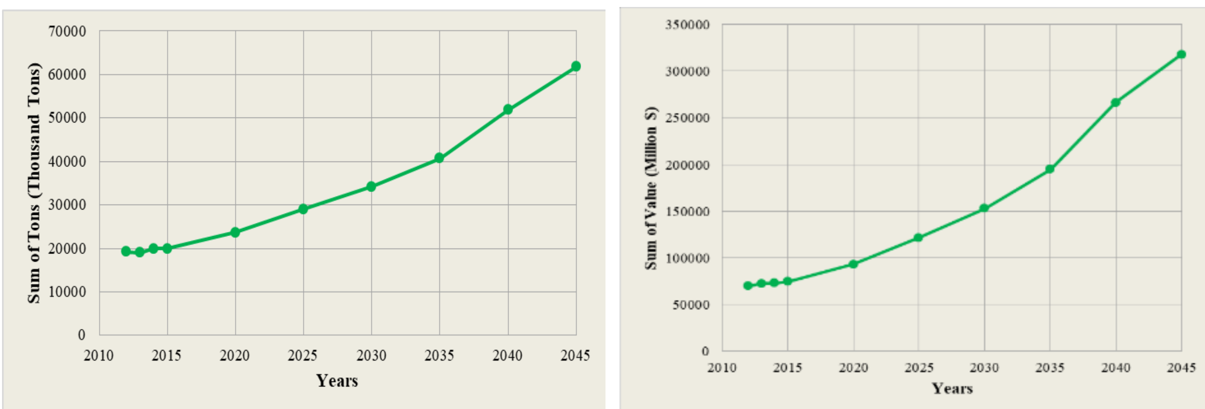
Direction	**Origin**	**Destination**	**Total Weight (million tons)**	**%**	**Total Value ($ billion)**	**%**
Domestic	Florida	Florida	468.55	66.34	296.37	32.79
	Florida	Rest of USA	46.94	6.65	136.95	15.15
	Rest of USA	Florida	123.70	17.51	253.13	28.01
Import	Foreign	Florida	39.49	5.59	71.99	7.97
Export	Florida	Foreign	19.27	2.73	69.56	7.70
Through	Outside of FL	Outside of FL	8.37	1.18	75.81	8.39
Total	---	---	**706.31**	**100.00**	**903.81**	**100.00**

Figure 4 graphically shows the total tonnages and values of goods projected until 2045. We can see that in 2045, total tonnage and value of goods are expected to increase to 873 million tons (36.55%) worth nearly $1084 billion for domestic shipments. For import, the total tonnage is expected to increase to approximately 97 million tons which is worth almost $265 billion. In case of export, the total tonnage is expected to increase to 62 million worth nearly $318 billion.

Domestic



Import



Export

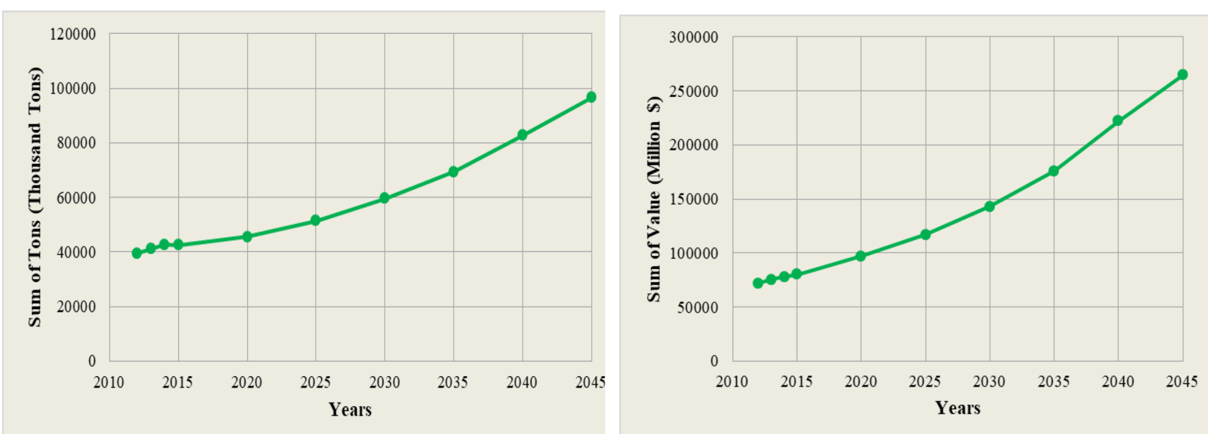


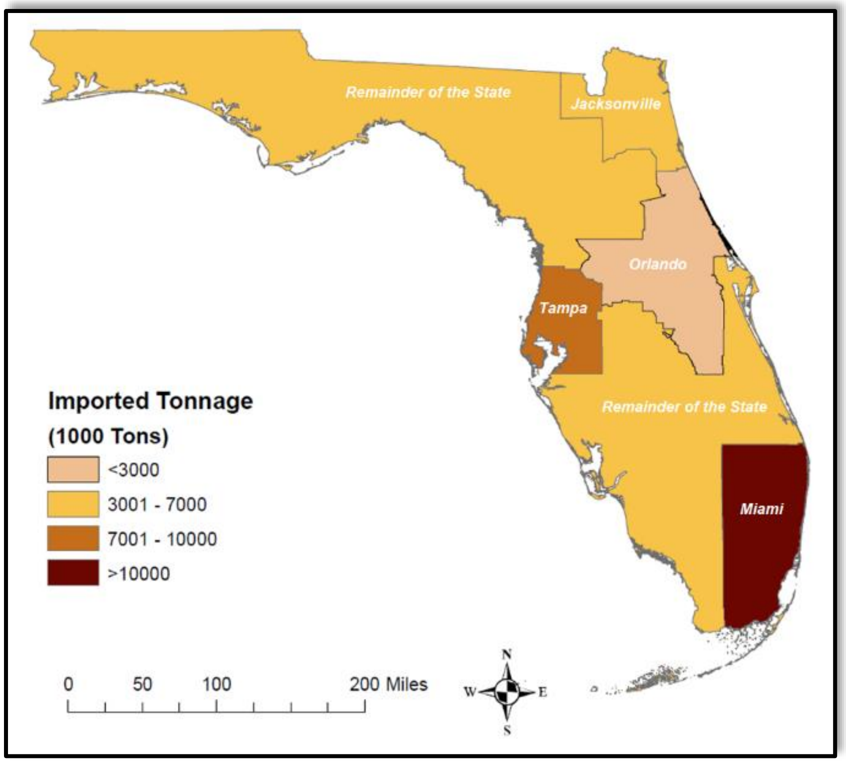
Figure 4 Predicted Tonnage (Left) and Value (Right)

3.1.2 Import (Inbound Freight)

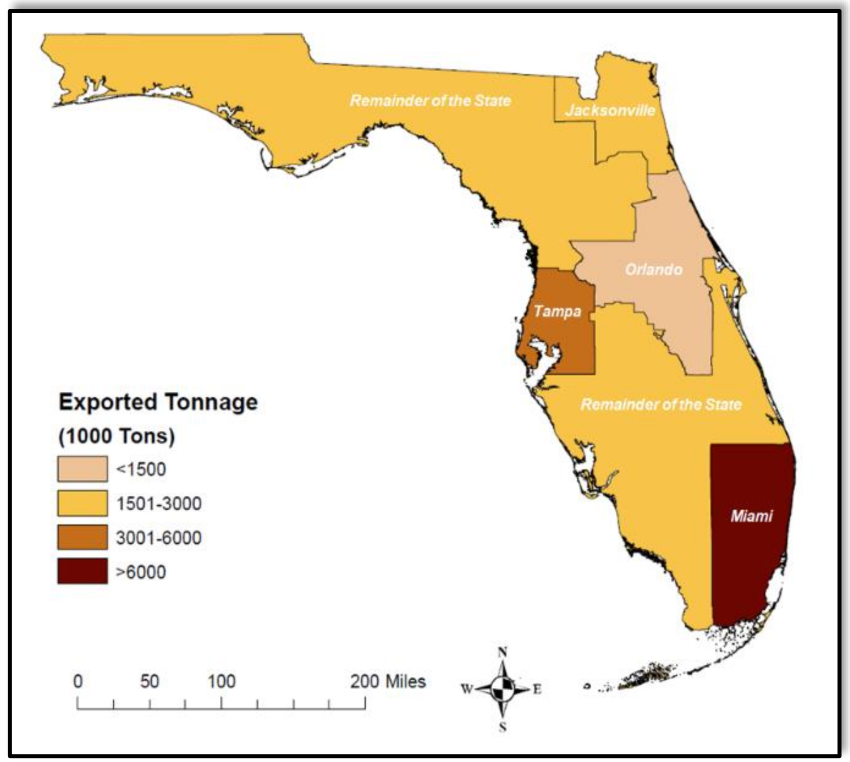
Figure 5(a) graphically represents, by region, the distribution of total inbound tonnage from foreign origins. Among the five regions, Miami is the top region receiving freight shipments (15 million tons), accounting for almost 38 percent of all imported tonnage in Florida. Tampa is next accounting for more than 22 percent (8.8 million tons) followed by remainder of the state (17.18%), Jacksonville (16.12%), and Orlando (6.24%).

3.1.3 Export (Outbound Freight)

Figure 5(b) graphically represents, by region, the distribution of total outbound tonnage from Florida to foreign regions. Of the five FAF regions, Miami accounted for 41 percent (7.9 million tons) of the total exported freight tonnages. The second highest is Tampa which exported almost 30 percent (5.8 million tons) of total exported tonnage. Orlando, Jacksonville, and remainder of the state each exported 5 to 15 percent (1 to 3 million).



(a) Imported Flow to Florida Regions



(b) Exported Flow to Florida Regions

Figure 5 Imported and Exported Flow to and from Florida Regions

3.2 Mode Share Analysis

In the state of Florida, truck is the dominant mode of freight transportation carrying more than 96 percent (448 million tons) of the total tonnage in the region. It is followed by rail accounting for approximately 3 percent (15 million tons) of the total tonnage in 2012. Figure 6 reflects the distribution of domestic freight tonnage moved within Florida by mode. Approximately 1 percent of the domestic intraregional freight travelled by water, air, pipeline and other modes. It is understandable since shipping by air is costly if it's within state while water is more time consuming than other modes.

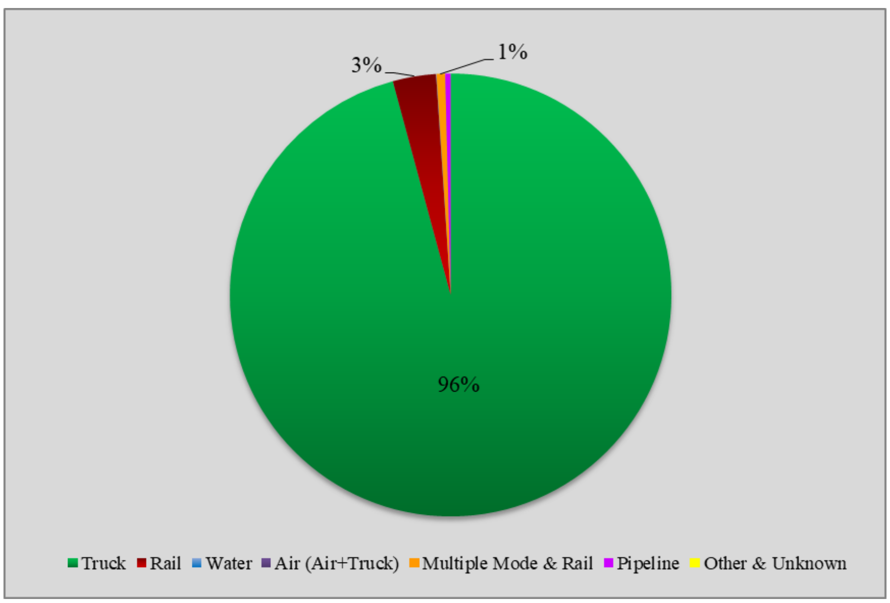


Figure 6 Mode Split by Tons – Intraregional Freight within Florida

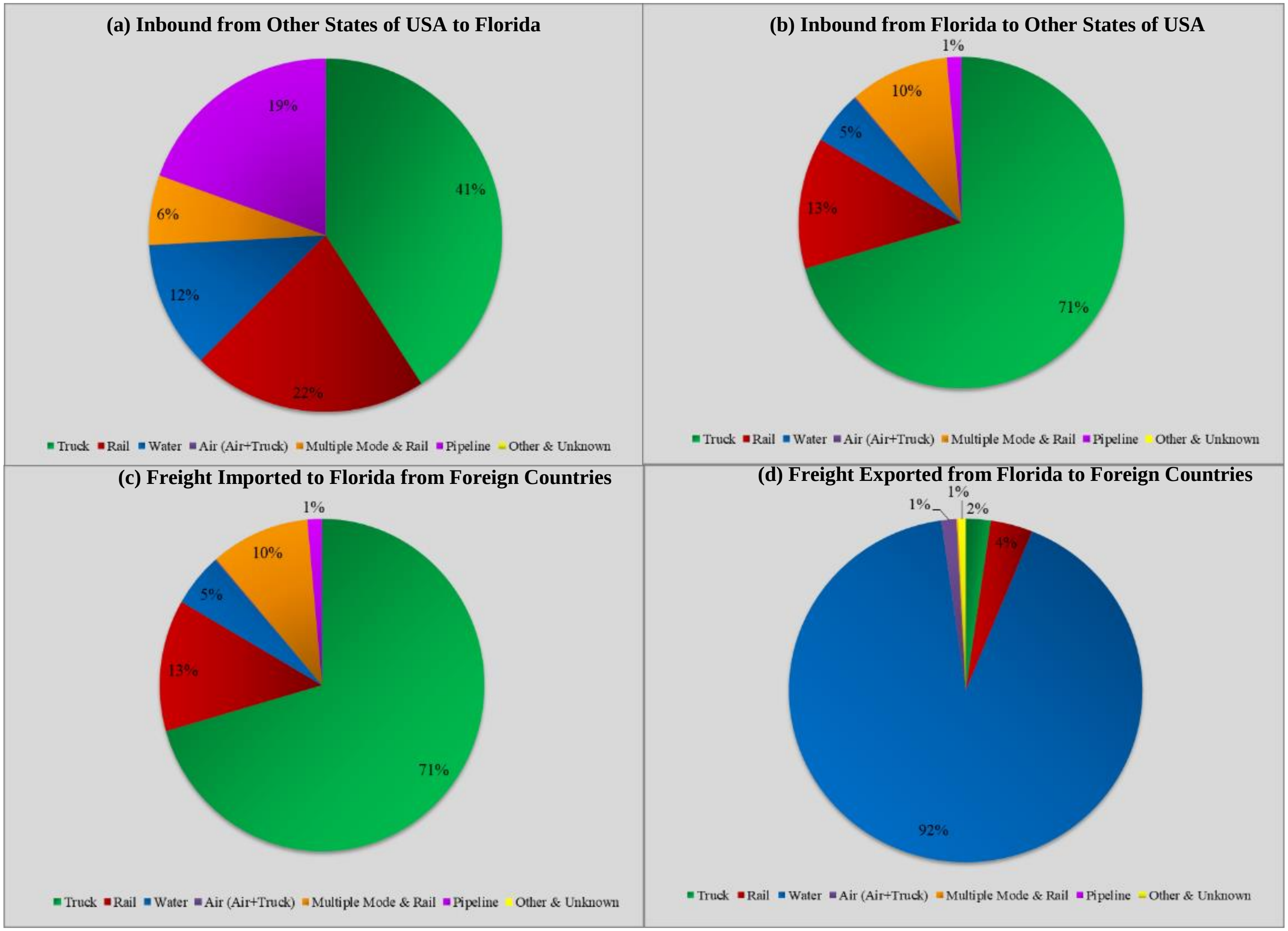


Figure 7 Modal Split by Tons

Figure 7 (a) reflects the distribution of domestic freight traveling inbound from other states of the USA to Florida by mode. Except air cargo, almost all types of mode have been used to bring freight from other states to Florida with truck (51 million tons), rail (27 million tons), pipeline (24 million tons), and water (14 million tons) being the dominant four modes.

From Figure 7 (b) Domestic outbound flows were mostly dependent on trucks. 71 percent (33 million) of total tonnage of the products was carried out of Florida to other states by Truck. The other two most common modes were Rail (13%) and multiple modes and mail (10%).

Figure 7 (c) clearly represents that majority of the commodity by tonnage imported to Florida from foreign countries was by waterways (92%). Since, Florida is surrounded by sea from three sides and has some major ports, this is expected.

Similar to import, majority of the tonnages were exported to foreign countries by waterways (83 percent or 16 million tons) while truck accounted only 10 percent of total exported weight as shown in Figure 7 (d).

3.3 Trading Partners

In addition to the analysis by mode and commodity summarized in the previous sections, it is also important to identify the state's key trading partners. By measurement of weight, most of the commodity was imported from Rest of Americas (Puerto Rico) which was greater than the weight exported to that foreign region. Compared to commodity exported to Canada and Europe, the weight of imported commodity from these two foreign regions was greater. The total tonnage of imported and exported commodity from and to South-West and Central Asia and South-East

Asia & Oceania is almost same. The weight of commodity exported to Africa was very low in 2012.

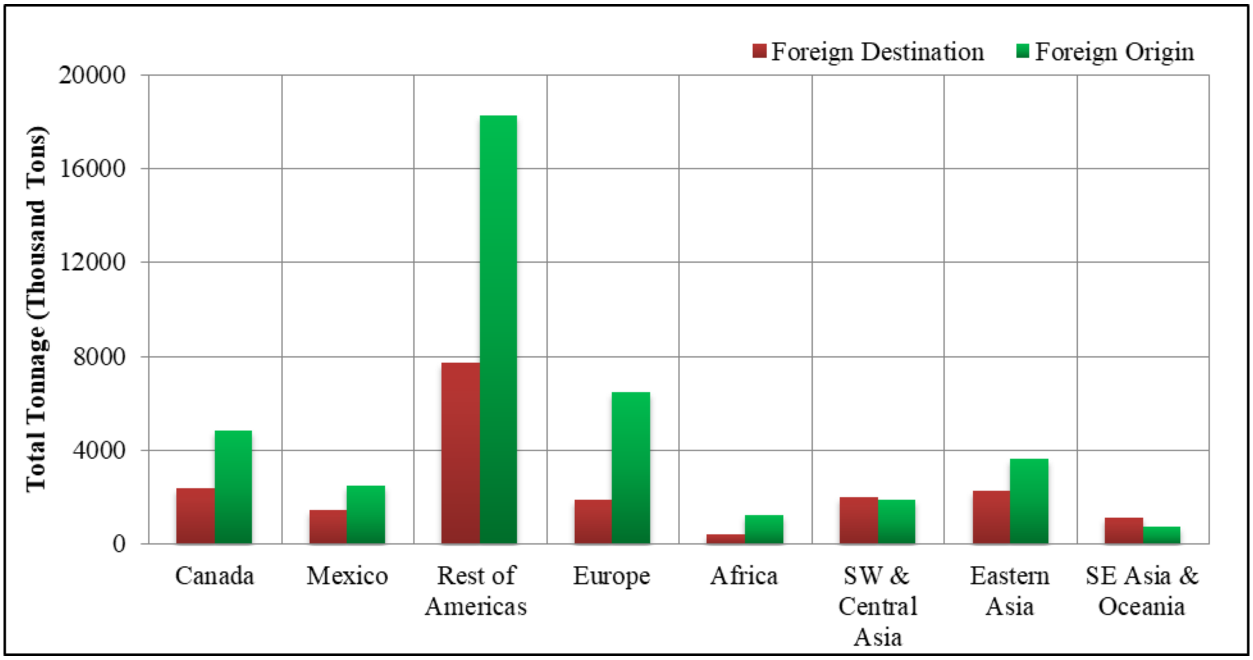


Figure 8 Total Tonnage of Commodity by Foreign Origin and Destination

3.4 Transearch Dataset

Transearch is a proprietary carrier centric comprehensive freight database owned and maintained by Global Insight Inc. It provides detailed information on commodity type (as per Standard Transportation Commodity Classification (STCC)), tonnage, value, ton-mile, origin-destination and mode used for freight movement. A Transearch domestic commodity flow database for the state of Florida was purchased from IHS/Global Insight by FDOT for the year

2011. In addition to the base year data, the database also provided projection till 2040 at a five-year interval starting from 2015.

For analysis purpose, we divided the commodity flows into four categories. These are: domestic, import, export, and through. The domestic flow is further subdivided into three groups: inbound, outbound, and within Florida. The definitions are outlined below:

- Domestic:
    - Inbound: Freight flows that originated in other states of the USA except Florida and are destined to Florida.
    - Outbound: Freight flows that originated in Florida and are destined to other states of the USA except Florida.
    - Within Florida: Freight flow that originated and terminated in the state of Florida.
- Import: Freight flows that originated in foreign countries outside of the USA and are destined to Florida.
- Export: Freight flows that originated in Florida and are destined to foreign countries outside of the USA.
- Through: All domestic and international freight flows that neither originated nor were destined to Florida but passed through the state for some leg of the journey.

This flow classification scheme is comparable with that of the FAF dataset. According to Transearch data, in the year 2011, a total of 4.5 billion tons of goods moved from, to, and within the State of Florida. Figure 9 illustrates the distribution of total domestic tonnage. We can see that

51 percent of the domestic flows occur within Florida followed by inbound flows (35%). The low share (14%) of commodity tonnage originating in Florida and terminating in the rest of the USA signifies the dominance of service industry in Florida.

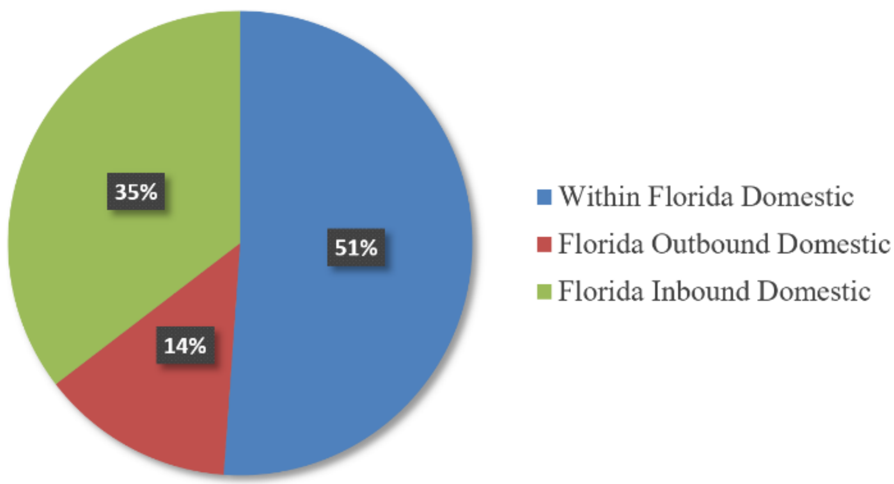


Figure 9 Tonnage Distribution of Domestic Flows

3.5 Comparison of the FAF and Transearch Databases

From the discussion in the previous sections, we now know that both FAF and Transearch report similar type of information on freight movements (tons, ton-miles, values, modes, trade type, commodity types). However, the databases differ in terms of data collection and construction methodology as well as the reporting criteria. As a result, it is very difficult to make any direct comparison of the quantities of freight movements reported. Thus, the most important part of this task of the study was a thorough comparison of these two databases. The following section presents the results of this comparison exercise.

3.5.1 Comparison of Total Weight and Value of Commodities

In the first step, we conducted the comparison between total tonnage and value of commodities at the entire state level. The results are shown in Table 4. In the overall, Transearch reports less than half the tonnage (706 million vs. 443 million) and two-thirds of the value ($904 billion vs. $635 billion) captured by FAF. Of the five trade types, the difference (both for tonnage and value) is the highest for export and import freight movements. For domestic trade types, the variation ranges between 1-2.1 times.

Table 4 Comparison of Tonnage and Value by Trade Type

Mode	Weight (Million Tons)			Value (Billion USD)		
	FAF	Transearch	Comparison (FAF/Transearch)	FAF	Transearch	Comparison (FAF/Transearch)
Truck	580.47	314.77	1.84	642.40	461.80	1.39
Rail	51.57	77.43	0.67	11.29	80.81	0.14
Water	27.85	50.96	0.55	26.05	49.94	0.52
Air	0.69	0.24	2.88	62.88	42.11	1.49
Others	45.74	0.01	4574.00	161.19	0.06	2686.57
Total	**706.31**	**443.39**	**1.59**	**903.81**	**634.72**	**1.42**

3.5.2 Comparison of Total Tonnage and Value by Trade Type and Mode

In addition, we conducted comparison of total tonnage and value by trade type and mode as well (Table 5 and Table 6). The highest variation in tonnage reporting was observed for other modes. This is expected, since Transearch doesn't cover freight movement using Pipeline mode and Pipeline represents the major share in the other mode category.

Table 5 Mode Share by Weight (Million Tons) and Trade Type (FAF and TranSEARCH)

Mode	Export		Import		Inbound Domestic		Outbound Domestic		Within Florida		Total	
	FAF	TS	FAF	TS	FAF	TS	FAF	TS	FAF	TS	FAF	TS
Truck	18.41	0.98	29.57	0.87	50.56	77.19	33.11	42.48	448.81	193.25	580.47	314.77
Rail	1.50	0.75	2.52	0.69	26.74	36.26	6.00	11.59	14.81	28.13	51.57	77.43
Water	2.03	0.66	8.91	4.37	14.32	40.40	2.44	4.46	0.15	1.07	27.85	50.96
Air	0.33	0.01	0.23	0.01	0.07	0.15	0.06	0.07	0.00	0.00	0.69	0.24
Others	1.14	0.00	2.48	0.00	32.01	0.00	5.31	0.00	4.79	0.00	45.73	0.00
Total	**23.41**	**2.39**	**43.71**	**5.95**	**123.70**	**153.99**	**46.94**	**58.61**	**468.55**	**222.46**	**706.31**	**443.34**

Table 6 Mode Share by Value (Billion USD) and Trade Type (FAF and TranSEARCH)

Mode	Export		Import		Inbound Domestic		Outbound Domestic		Within Florida		Total	
	FAF	TS	FAF	TS	FAF	TS	FAF	TS	FAF	TS	FAF	TS
Truck	69.85	3.60	61.88	3.99	155.64	184.31	79.48	75.66	275.56	194.26	642.40	461.80
Rail	1.158	0.46	1.04	0.53	5.31	46.99	2.63	16.48	1.14	16.35	11.29	80.81
Water	5.24	0.77	6.39	1.90	12.77	36.02	1.63	10.06	0.02	1.19	26.05	49.94
Air	34.14	0.68	17.38	1.18	6.45	24.14	4.74	15.73	0.16	0.38	62.88	42.11
Others	7.11	0.04	13.17	0.02	72.96	0.00	48.46	0.00	19.49	0.00	161.19	0.06
Total	**117.50**	**5.54**	**99.87**	**7.61**	**253.13**	**291.45**	**136.95**	**117.93**	**296.369**	**212.18**	**903.81**	**634.71**

CHAPTER FOUR: ECONOMETRIC DATA FUSION ALGORITHM

The primary motivation for the dissertation is the development of a fused database to realize transportation network flows at a fine spatial resolution (county level) while accommodating for production and consumption behavioral trends. Thus, we undertake disaggregation of FAF flows while augmenting with production consumption-based TS flows. Towards this end, we formulate and estimate a joint econometric model framework grounded in maximum likelihood approach to estimate county level commodity flows. The framework has two separate modules to ensure matching estimated county level flows with commodity flows in FAF and TS at the appropriate spatial resolution. A third module generates a behavioral connection between FAF and TS. In our algorithm, we connect the flows between TS and FAF by generating potential paths between the origin and destination of interest for TS flows. Note that the inherent differences in the data cannot be completely reconciled. Hence, the framework focuses on building a fused database that maximizes the match with the commodity flows in the two databases. The consideration of behavioral trends in the model framework can assist us in parameterizing TS flow relationships thus allowing us to circumvent TS for the future (if needed). This chapter describes the procedure developed for disaggregating FAF data and fusing with Transearch data that can be employed by FDOT with future data releases (e.g. FAF) and purchases (e.g. Transearch). The approach also can support FAF data disaggregation without new Transearch purchases.

4.1 <u>Insights from Data Sources</u>

Our data exploration analysis as part of Task 1 provided us with valuable insights about the datasets acquired for the study, particularly FAF and Transearch. FAF is developed by Federal

Highway Administration (FHWA) (an enhanced version of the Commodity Flow Survey (CFS)) and is a publicly available freight demand data. It is free and provides a snapshot of freight flows between and within states in the United States. Hence, it is sufficient for understanding mesoscale (the data has large spatial areas of zones comprising of multi-county urban areas, portion of states, or entire states) freight flows for policy studies. Unfortunately, the dataset does not provide adequate data about local (since, the movement information is mostly aggregated to the state and region level) or temporal trends in freight flows. In other words, the database does not have sufficient level of spatial resolution to support local, regional, or state planning and project development (Bujanda et al., 2014; Anderson et al., 2013). On the other hand, Transearch by IHS Global Insight is a proprietary data source that includes rich information on commodity flows in the form of annual tonnage, containers (for intermodal), carloads (for rail) as well as the dollar value shipped. The data has greater level of detail than FAF useful to examine logistics and modal trends. However, it is prohibitively expensive to acquire.

We have also found that major differences exist between FAF and Transearch due to their varying reporting structure. First, FAF provided information on actual flows that occurred on the road network, whereas Transearch provided demand and supply based flow information (not actual flow direction). According to FAF if 100 tons of goods were shipped between region 1 to region 2, it means that that 100 tons were directly shipped from region 1 to region 2 through the transportation network. On the other hand, according to Transearch if 100 tons of goods were shipped between county 1 to county 2, it means that there was a demand for 100 tons of goods in county 2 which was supplied by county 1. However, it doesn't tell the actual transportation route. That 100 ton might be shipped directly from county 1 to county 2 or it might have gone through some other intermediate counties while reaching county 2. Second, FAF reports flows for the 5

FAF regions in Florida while Transearch flows are reported at the Business Economic Area (BEA) level. It indicates that FAF has poorer spatial resolution than Transearch. Third, difference lies in the commodity type classification. Transearch follows the Standard Transportation Commodity Classification (STCC) code while FAF uses Standard Classification of Transportation Good (SCTG) code. As a result, FAF has 43 commodity types reported while TS has over 500 commodity types reported. Therefore, we need to develop an algorithm that reconciles these differences.

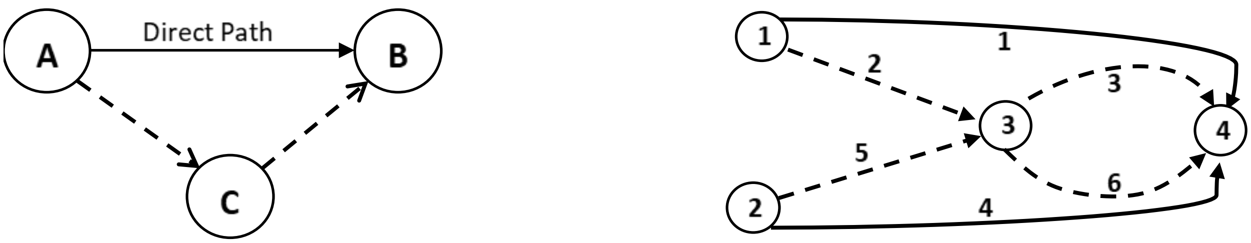
4.2 Algorithm Development

In this section, the proposed algorithm is described. Prior to discussing the algorithm details, the notations and terminology used in the algorithm are presented.

4.2.1 Network Representation

The study defines nodes, paths, and links in the usual network theoretic approach. Nodes represent county centroids. These represent either origin, destination or intermediate points. A direct connection between any two nodes is defined as a link. Paths represent a series of links that connect an origin and destination. To elaborate on the terminology, a simple representation is provided in the Figure 10. In the Figure 10(a), from origin county ‘A’, freight flow can be transferred to destination county ‘B’ in a direct path (i.e. no intermediate nodes) indicated by a solid line. The flow could also move along an indirect path. In our study, given the model is a statewide model, we assume that one intermediate node is adequate for considering all possible paths between OD pairs to ensure computational tractability of the algorithm. The path with one intermediate node is referred to as a one hop path. In the Figure 10(a), a one-hop path from county ‘A’ to county ‘B’

with an intermediate stop at county ‘C’ is shown with the dashed line. In the Figure 10(b), origin node ‘1’ and destination node ‘4’ have the following possible paths on the network. (i) ‘1’ - ‘4’ direct (link ‘1’ – say, path 1), (ii) ‘1’ - ‘3’ - ‘4’ in a one-hop path (link ‘2’ – link ‘3’ – say path 2, or link ‘2’ – link ‘6’ – say path 3). Therefore, three different paths are considered here from origin ‘1’ to destination ‘4’ that uses four different links (i.e. links ‘1’, ‘2’, ‘3’, and ‘6’).



(a) Paths between OD pairs, A, and B

(b) Links and nodes on a network

$$
\begin{array}{cc|cccccc}
 & & \multicolumn{3}{c}{O-D} & \multicolumn{3}{c}{O-D} \\
 & & \multicolumn{3}{c}{1-4} & \multicolumn{3}{c}{2-4} \\
 & Link\backslash Path & 1 & 2 & 3 & 1 & 2 & 3 \\
\hline
 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
A = & 2 & 0 & 1 & 1 & 0 & 0 & 0 \\
 & 3 & 0 & 1 & 0 & 0 & 1 & 0 \\
 & 4 & 0 & 0 & 0 & 1 & 0 & 0 \\
 & 5 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & 6 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}
$$

(c) Links- path matrix for the simple network shown on (b)

Figure 10: Paths, Links, and Nodes of a Simple Transportation Network

To represent the relationship between paths and links in our system, a link path matrix is generated. For the network in Figure 10(a) and Figure 10(b), the link-path matrix (A) is shown in Figure 10(c). The rows represent the links and the columns represent the paths between the given OD pairs (see Figure 10 for details). Each element of the matrix is a binary indicator that represents

if the link 'i' is included in the corresponding path. The variable of interest in the algorithm is the transportation network county to county flows generated by fusing TS data at the county level and FAF data at the FAF region level. Let V_{ij} represent the link flows between county pair i and j. The entire set of link flows are considered in a matrix form as V. Given the link-path matrix A, and path flow vector 'h', the link flow matrix, 'V' is given by the following equation.

$$V = A * h \tag{1}$$

4.2.2 Joint Model System

Let, y_{ij} represent the natural logarithm of the actual TS flow, and $\hat{y}_{ij}$ the estimated transearch flow. With these notations, the log-linear model takes the following form:

$$y_{ij} = \beta X_{ij} + \varepsilon_{ij} \tag{2}$$

where, X_{ij} are the independent variables for the specific OD pair $i - j$ and β represents the corresponding vector of parameters. Assuming the usual linear regression formulation, the likelihood for the estimation takes the following form:

$$LL_{TS} = \frac{\emptyset(\frac{\hat{y}_{ij} - y_{ij}}{\sigma_{TS}})}{\sigma_{TS}} \tag{3}$$

where, $\emptyset$ represent the probability density function of the standard normal distribution, and σ_{TS} is the standard deviation of ε_{ij}.

Given that TS flow is an input-output flow, the objective is to decompose these flows into actual network level link flows by considering the various paths between each OD pair. The path flows will allow us to determine the link flows. These flows are generated employing a fractional split approach. The actual path flow is unobserved; hence, a latent variable is considered, and the resulting link flows are matched with observed flows. The probability for each path is determined in a random utility approach as follows:

$$\cup_{ij}^{k} = \sum_{i,j\,\in O,D;\;k=1}^{K} \alpha_{ij}\, X_{ij}^{k} \tag{4}$$

$$P\left(X \middle| x_{ij}^{k}\right) = \frac{\exp(\cup_{ij}^{k})}{\sum_{l=1}^{K} \exp(\cup_{ij}^{l})} \tag{5}$$

Based on the path flow probability the actual flow assigned to each path is determined as follows:

$$h_{ij}^{k} = \hat{y}_{ij} * P(X|x_{ij}^{k}) \tag{6}$$

The path flow estimation leads to the estimation of the link flows V, using Equation (1). Given that these flows are available at the county level, we need to aggregate them to a coarser level to compare the flows to observed FAF flows. The aggregation is achieved as:

$$\hat{F}_{OD} = \sum_{l \in O, q \in D} V_{lq} \tag{7}$$

Let F_{OD} be the observed FAF flows. The log-likelihood for comparing the predicted FAF flows with observed FAF flows in the linear regression form is given by the following mathematical expression, where, σ_{FAF} is the standard deviation of the estimate of FAF flows.

$$LL_{FAF} = \frac{\emptyset(\frac{\hat{F}_{OD} - F_{OD}}{\sigma_{FAF}})}{\sigma_{FAF}} \tag{8}$$

Given the aggregation proposed, the contribution of the FAF log-likelihood needs to be carefully computed. While origin and destination counties have their corresponding FAF zones, the intermediate zones also have a FAF zone. Therefore, the allocation is obtained for an OD pair by apportioning the error to all FAF zones involved over the entire path set for that OD pair. For this purpose:

$$LL_{FAF}{}^{k} = \frac{\sum_{r=1}^{n} LL_{FAF}{}^{n}}{n} \tag{9}$$

where, n is the number of link in the path k = $\begin{cases} 1, & for\ direct\ path \\ 2, for\ one - hop\ paths \end{cases}$

Further, FAF zones can represent a large number of counties. To normalize for the number of counties, we employ the following equation:

$$LL_{FAF}{}^{OD,Norm} = \frac{\sum_{s=1}^{N} LL_{FAF}{}^{k}}{N_C} \tag{10}$$

where, N_c is the number of county pairs in the OD FAF region pairs. Finally, the joint log-likelihood is provided by the sum of log-likelihood for FAF and TS flow.

$$LL_{total\ i,j} = \sum\nolimits_{i,\ \ j \in TAZ} (LL_{TS_{i,j}} + LL_{FAF}{}^{OD,Norm}{}_{i,j}) \tag{11}$$

The proposed algorithm is programmed in Gauss matrix programming language (Aptech, 2015). The steps are shown in the flow chart (Figure 11).

4.3 Empirical Data

In this section, we briefly discuss the databases used in the fusion model and the data preparation procedures. Florida has five FAF regions: Jacksonville, Miami, Orlando, Tampa, and remainder of Florida (see Figure 3). On the other hand, Florida is represented as 68 zones in the TS database. In our study, we have access to the 2011 base year data for Florida that includes forecasts for 2015 through 2040 at a five-year interval.

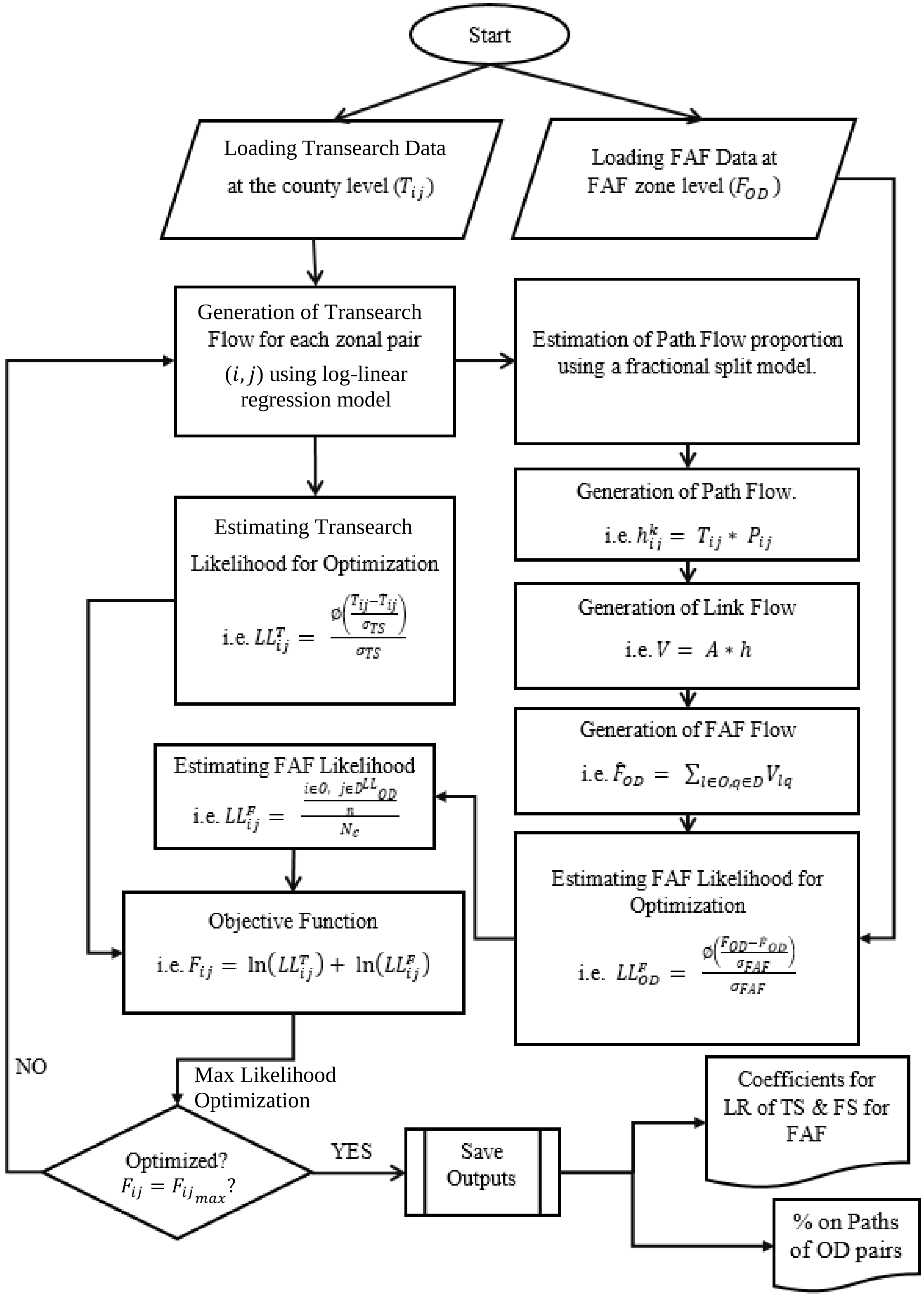


Figure 11 Flow Chart of Algorithm

4.3.1 Commodity Classification

As mentioned before, there are 43 commodity types in FAF while TS commodities are classified into 562 commodity types. To generate a comparable commodity type classification, we consolidated the different commodity types into 13 comparable commodity types in both datasets (see Viswanathan, et. al., 2008, for a similar classification of commodities). The commodity types are: agricultural products, minerals, coal, food, nondurable manufacturing, lumber, chemicals, paper, petroleum, other durable manufacturing, clay and stone, waste, miscellaneous freight (including warehousing). Table 7 provides a comparison of freight flows by consolidated commodity type within Florida. Across commodity types, we can see that the highest ratio of flows for FAF relative to TS is six for non-durable manufacturing products and chemicals. The lowest ratio is observed for miscellaneous freight and warehousing commodity type (0.28). Note that TS reports secondary flows including drayage whereas FAF does not contain any information on drayage. Thus, it is not surprising that we have a lower ratio.

Table 7 Freight Flows by Weight for Within Florida Flows in Transearch and FAF4

FCC	Transearch Flow (million tons)	FAF4 Flow (million tons)	Ratio (FAF4 flow/Transearch flow)
Agricultural Products	17.13	34.26	2.00
Minerals	51.59	191.12	3.70
Food	12.21	29.28	2.40
Nondurable Manufacturing	0.86	5.09	5.95
Lumber	5.23	19.64	3.75
Chemicals	1.72	10.28	5.99
Paper	3.04	2.80	0.92
Petroleum	13.61	59.77	4.39

FCC	Transearch Flow (million tons)	FAF4 Flow (million tons)	Ratio (FAF4 flow/Transearch flow)
Other Durable Manufacturing	5.12	12.91	2.52
Clay and Stone	24.15	39.95	1.65
Waste	7.47	29.18	3.91
Miscellaneous Freight and Warehousing	51.13	14.54	0.28
Total	**193.25**	**448.81**	**2.32**

** There is no flow for the commodity Coal in the within Florida flow

4.3.2 Independent Variable Generation

We compiled several exogenous variables for the fusion model. These are: (1) origin-destination indicator variables including Origin (or destination) is in Orlando, Tampa, Jacksonville, Miami, Remainder of Florida region, (2) socio-demographic and socio-economic indicators including population and employment, (3) transportation infrastructure indicators including road and railway line length, number of ports, airports, and intermodal facilities, and (4) several interactions of these variables. Population and employment data were collected at the county level from U.S. Census Bureau (U.S. Census Bureau, 2017a; U.S. Census Bureau, 2017b). Transportation related variables were generated using the ArcGIS platform intersecting the facility shapefiles collected from Florida Geographic Data Library (FGDL) (FGDL, 2017) with that of the county shapefile. Post-processing of the intersected files provided us the length of roadways and railways, number of seaports, airports, and intermodal facilities at the county level. Please note that these variables were compiled for the base year of 2011. Finally, for the fractional split model, we needed to generate all path choice set for every OD pair. For this purpose, we considered 1 direct path and 66 one-hop paths (that pass through another county). The paths were generated for all OD pairs with non-zero flow. The overall path matrix was quite large with number of elements

ranging from 6700 to 270000 across various commodities. For the paths created, path distances between origin and destination counties were generated as a sum of the link distances. A link distance for county pairs was determined using the shortest path procedure of ArcGIS's network OD cost tool. The highway route for the local and highways provided by the Florida Department of Transportation (FDOT) was used for this purpose.

4.4 Implementation of Data Fusion Algorithm

The proposed algorithm is implemented separately for each commodity type. For the sake of brevity, we only present the model results for two commodities: Agricultural products and Food. We discuss the results for the two commodities separately.

4.4.1 Commodity Type: Agricultural Products

In Table 8, columns 3 and 4 provide parameter coefficients and t-statistics for Agricultural products. The TS module corresponds to the overall county to county flow tonnage while the FAF module provides the fractional model estimates.

4.4.1.1 TS Module

In terms of Origin indicator variables, Jacksonville origin region is likely to have lower flow relative to other locations. On the other hand, for agricultural products Miami origin is associated with larger flows. For Destination indicator variables, Orlando variable is associated with larger flows while Miami is associated with smaller flows. The reader would note that these

indicator variables serve as region specific constants and are influenced by other exogenous variables.

For agricultural products, several destination specific attributes have significant impact on flows. The number of warehouses in the destination county is associated positively with flows to the destination county. The number of intermodal facilities in the destination county is negatively associated with flows. On the other hand, no attributes for the origin location provided significant parameters. Several interaction variables from different variable categories were also considered. The variable considering the interaction of origin county employment and destination county employment was found to be positively associated with county to county flows. The standard error of the estimate represents the standard deviation of the unobserved component in the regression model.

4.4.1.2 FAF Module

The fractional split model in the FAF module is based on a large number of alternatives. Hence, the model only allows for the estimation of generic coefficients i.e. no alternative specific effects can be estimated. The path distance variable is considered in the model. Any other origin or destination variable would require us to consider interaction with path distance. The models with such interaction variables did not provide intuitive results. Hence, we resorted to considering only the path distance variable in our FAF module. The path distance variable was negative as expected, indicating the probability for paths reduces rapidly with longer distances. The result clearly indicates a larger path flow allocation to direct paths while one-hop paths with very large distance are likely to have a very small path flow.

Table 8 Model Estimates for Agricultural Product and Food

Model	Explanatory Variables	Agricultural Product		Food	
		Estimates	*t*-stats	Estimates	*t*-stats
Transearch Module	**Intercept**	3.7763	99.4770	1.4275	9.7080
	Dummy for Origin/Destination				
	Jacksonville Origin	-0.7353	-5.4510	-	-
	Miami Origin	1.9115	10.5330	-	-
	Tampa Origin	-	-	0.8029	4.3680
	Orlando Origin	-	-	0.4654	3.0710
	Orlando Destination	0.7980	8.4710	-	-
	Miami Destination	-1.8459	-13.2560	-	-
	Destination County Attribute				
	Number of Warehouses	3.8474	20.3630	-	-
	Number of Ports	-	-	0.1649	4.0970
	Number of intermodal facilities	-0.1948	-5.4730	-	-
	Network Length (in KM)	-	-	1.2484	10.5550
	Origin County Attribute				
	Network Length (in KM)	-	-	1.6818	18.8570
	Interaction Variables				
	Number of Origin County Employment (in 10^3) * Number of Destination County Employment (in 10^3)	0.7266	9.6940	-	-
	Number of Origin County Employment (in 10^3) / Total Destination County Population (in 10^6)	-	-	-0.8736	-9.7270
	Standard Error of the Estimate for Transearch	1.8413	87.7510	2.4667	62.8660
FAF Module	**Path Distance (in 10 Miles)**	-0.0248	-1.199	-1.0858	-1.0200
	Standard Error of the Estimate for FAF	2.1615	22.591	0.8101	8.3480
Number of observations		4070		2447	
Log-Likelihood of the model		-11496.773		-6850.817	

4.4.2 Commodity Type: Food

In Table 8, columns 5 and 6 provide parameter coefficients and t-statistics for Food.

4.4.2.1 TS Module

For Food commodity, indicator variables for Tampa and Orlando origins are positively associated with flows. The magnitude of coefficient for the Tampa region is larger than the corresponding magnitude of coefficient for Orlando region. In terms of destination county attributes, number of ports and road network length in the county are associated positively with food flows. The commodity flow for Food is also influenced by origin county road network length. In terms of interaction variables, the interaction of origin county employment and destination county population was negatively associated with Food flows.

4.4.2.2 FAF Module

Similar to the model for agricultural products, we found negative relationship between the path distance and the path flow proportions in the model for food as well. The magnitude of the parameter is substantially larger for Food relative to Agricultural products. To be sure, these two parameters are not directly comparable.

4.4.3 Model Validation

To evaluate the performance of our proposed algorithm, several validation exercises were conducted. To be sure, the county to county freight flows generated from the exercise do not have an observed counterpart to validate. Hence, we resort to validation by intuition. For example, the ratio of FAF and TS for agricultural products is 2 (see Table 9). After fusing FAF and TS databases,

the ratio of the fused flows with TS flows was found to be 1.45. A similar exercise for Food yielded a value of 1.62 (relative to the original ratio of 2.40). In both cases, the results are quite reasonable.

Table 9 County Level Link Flow Prediction for Agricultural Product and Food

FCC	Description of Flow	Mean (Thousand Tons)	Std. Dev. (Thousand Tons)	Total (Million Tons)	No of Observations	FAF4 vs TS Ratio	Fused Link flows vs TS Ratio
Agricultural Products	TS County to County Flow	4.21	179.22	17.13	4070	2.00	1.45
	Estimated County Level Link Flow	5.51	22.11	24.75	4489		
Food	TS County to County Flow	4.99	35.06	12.21	2447	2.40	1.62
	Estimated County Level Link Flow	4.42	37.17	19.83	4489		

As a second step, we plot the relationship between county to county flows for TS and fused flows. The plots are created by considering proportion of statewide flows originating (or destined) to each county. Figure 12 and Figure 13 provides the plots for Agricultural Products and Food, respectively. In these figures, the plots for TS are on the left and the plots for fused flows are on the right. We can see from the figures that for Agricultural Products, both origin and destination-based plots, are quite similar. The counties in Central and South Florida regions account for larger share of the flows in TS as well as fused flows. For Food, the fused flows indicate a larger share of flows in Central and South Florida relative to the TS flows. However, the overall trends are still very similar.

As a final comparison exercise, we compare TS and fused flows originating from Miami-Dade County for the two commodities. For this purpose, we plot the tonnage of flows transferred between counties (see Figure 14). The thicker the line on the road network, the larger is the tonnage

transferred. From the figure, it is evident that we observe substantially thicker lines for fused flows. This is expected because fused flows should represent network flows whereas TS flows only represent origin destination flows. Hence, they always are likely to pass directly, whereas fused flows would be a result for multiple origin destination flows. Overall, the three validation steps provide evidence that the fusion algorithm provides outputs as expected from a joint system disaggregating FAF with production consumption trends form TS.

4.5 Summary

A major hurdle in freight demand modeling has always been lack of adequate data on commodity movements amongst different industry sectors for planning applications. The primary motivation for this study is the development of a fused database to realize transportation network flows at a fine spatial resolution (county level) while accommodating for production and consumption behavioral trends. To achieve the goal, the dissertation undertakes disaggregation of FAF flows while augmenting with production consumption-based TS flows. Towards this end, the dissertation formulates and estimates a joint econometric model framework embedded within a network flow approach grounded in maximum likelihood technique to estimate county level commodity flows. The algorithm is implemented for the commodity flow information from 2012 FAF data for five FAF zones and 2011 TS databases for 67 counties in Florida. Overall, our model system predicted well as manifested from the ratio of fused flows to observed TS flows for the two commodities for which the results are presented (Agricultural Products and Food). Moreover, the path distance coefficients are intuitive. As expected, shorter paths are allocated higher fraction of the flows compared to the longer paths.

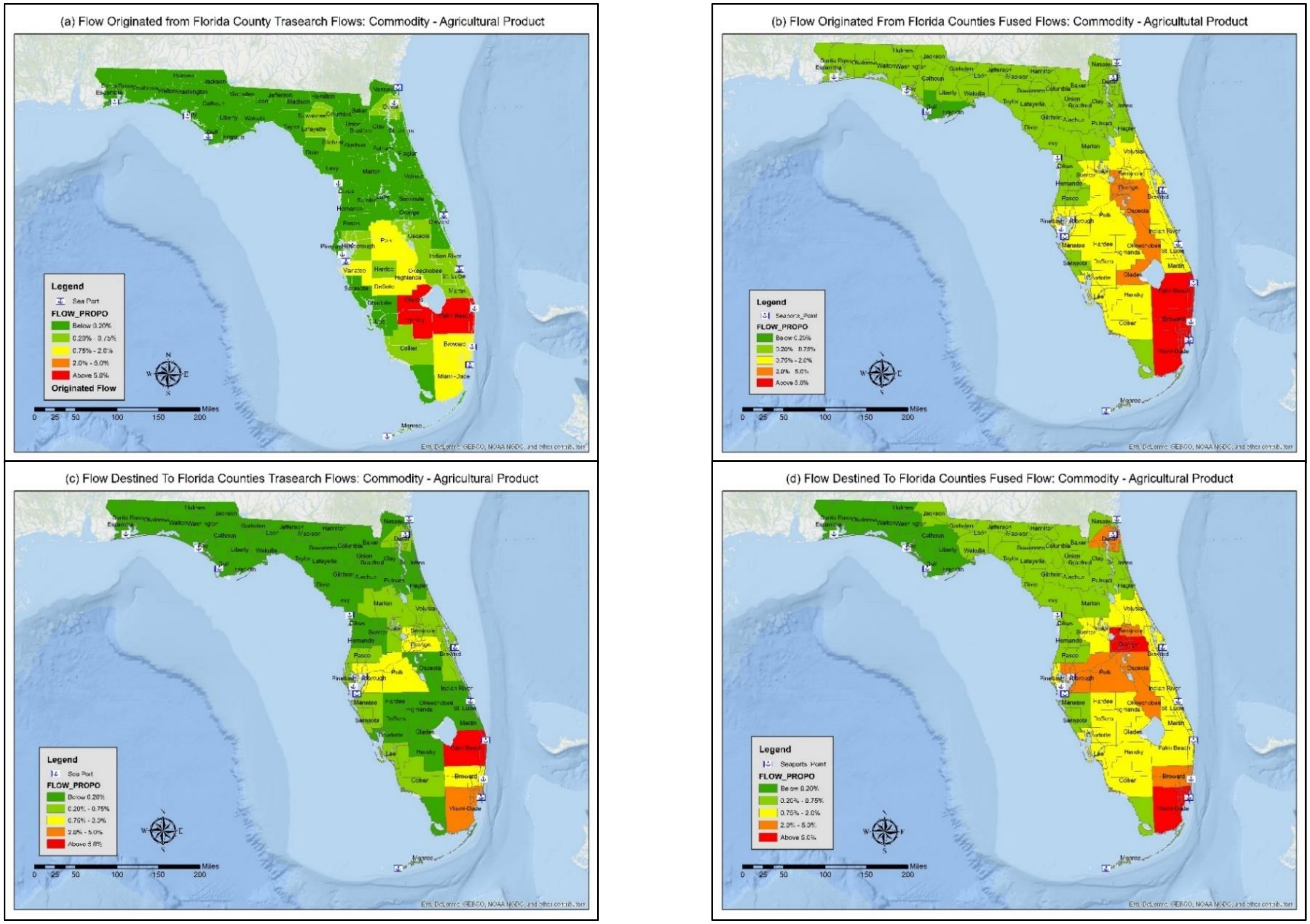


Figure 12 Transearch and fused flows within Florida counties for agricultural product

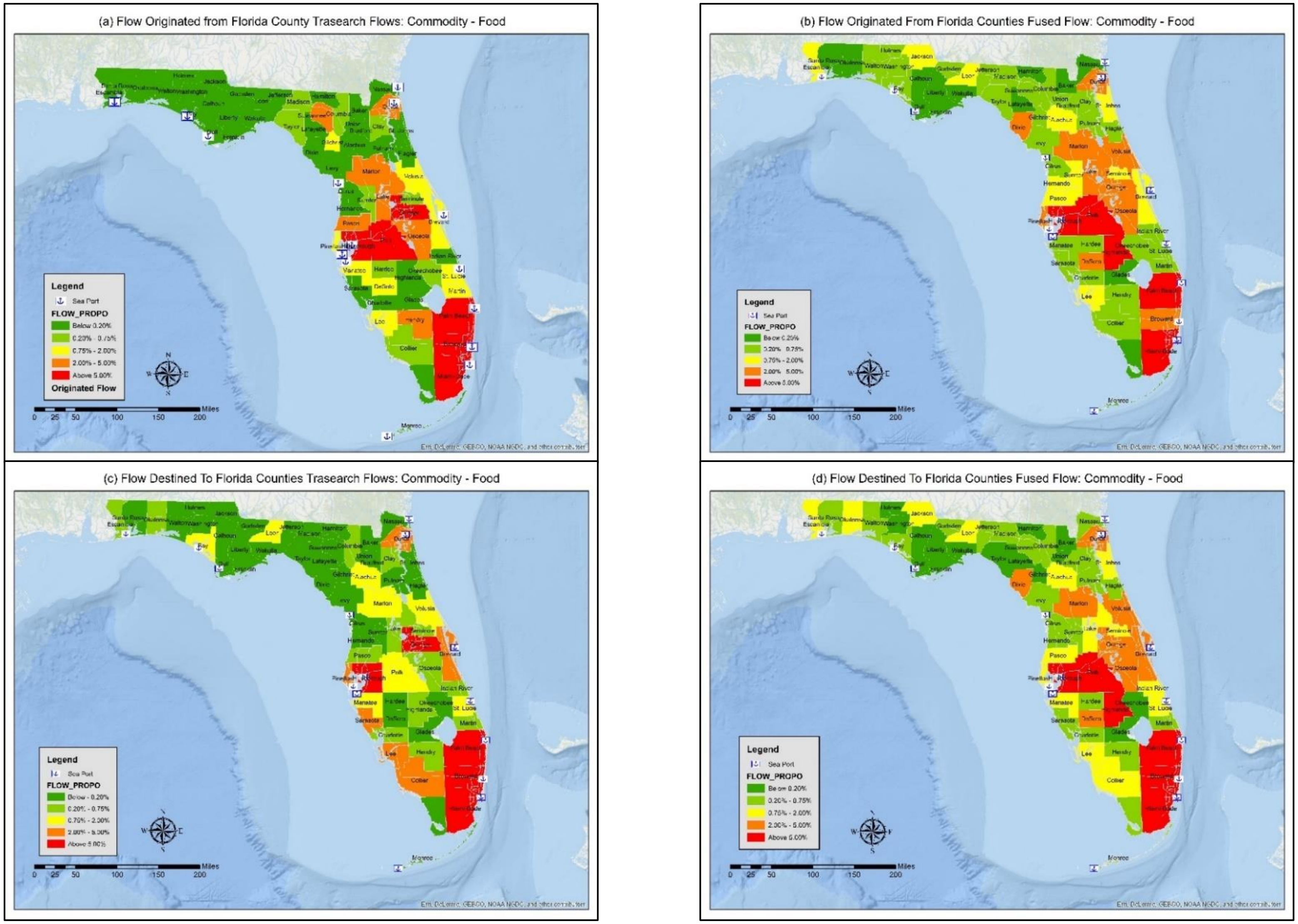


Figure 13 Transearch and fused flows within Florida counties for food

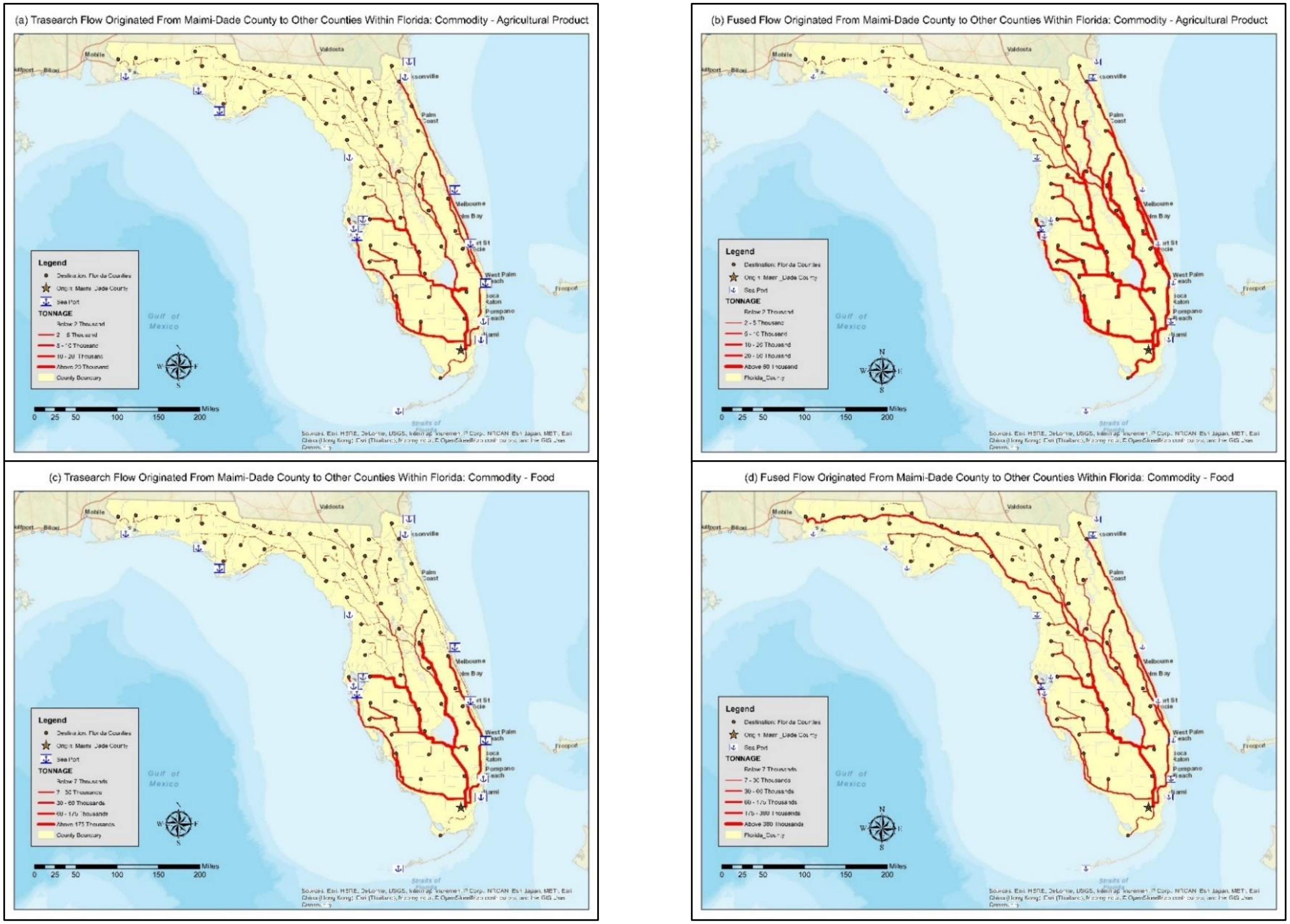


Figure 14 Transearch and fused flows from Miami-Dade County for agricultural products and food

CHAPTER FIVE: IMPLEMENTATION OF ALGORTHIM INCLUDING EXPORT AND IMPORT FLOWS AND FOR FUTURE YEARS

5.1 Introduction

In the preceding chapter, we developed a data fusion algorithm that draws from FAF4 and Transearch data. The algorithm bridges the two data sources while considering their inherent differences and similarities. More specifically, we formulated and estimated a joint econometric model framework embedded within a network flow approach and grounded in maximum likelihood (ML) technique to estimate county level commodity flows. The algorithm was developed as a proof of concept and is implemented for the base year (2011) for within Florida truck flows for two commodities. The flows within the state contribute to a small share of freight flows (about 30% of flows reported in FAF are within state flows). Considering flows entering/leaving Florida from other states in the US and other countries (export and imports) would provide a complete representation of freight flows on Florida's transportation system. With the ever-expanding global trade, imports and exports originating or destined to foreign countries is also likely to affect flows entering or leaving Florida. For example, North American Free Trade Agreement (NAFTA) has resulted in an increase in flows to and from Canada and Mexico to the US (Mendoza, et al. 1999). Various other trade deals and several legislative efforts have also resulted in increased trade with many foreign nations (SteadieSeifi, et al. 2014). Thus, it is essential for state planning agencies to consider out-of-state flows and foreign flows within the overall

freight planning platform to accommodate for potential fluctuations in these flows. The consideration is particularly relevant to Florida that has several ports with significant import and export flows.

In this current chapter, we extend the algorithm proposed in the previous chapter to consider spatial regions within Florida and outside Florida (in the United States and in foreign regions). The spatial resolution employed in the algorithm allows for multiple resolutions. Specifically, we resort to very detailed resolution inside Florida and larger aggregate zones to represent zones outside Florida. Furthermore, long-range planning of freight flows requires estimates of flows for future years. Thus, in this chapter, in addition to developing the algorithm for base year data fusion, we employ the proposed algorithm to generate future year predictions at the selected spatial resolution for various commodity types. Specifically, the current study generates estimates for freight flows from 2015 through 2040 in five-year increments. The model was implemented for thirteen commodity type classifications (more on this in data section). The generated estimates are evaluated to understand how freight flows are likely to evolve in the Florida region.

5.2 Research Methodology

In this section, we describe the proposed fusion methodology. Please note that the proposed model system is implemented separately for each commodity.

5.2.1 Model Structure

The model structure is presented in Figure 12. The network considered includes: (a) nodes representing zonal centroids to represent origins, destinations or intermediate points, (b) links representing a direct connection between any two nodes, and (c) paths representing series of links connecting origins and destinations. The relationship between paths and links in the system is generated using link-path matrix (created following the conventional network theory). Each element of the matrix is a binary indicator that represents if the link 'i' is included in the corresponding path (for more details on generating the link path matrix please see previous chapter).

In our analysis, transportation network-based zone to zone flows is to be generated by fusing Transearch and FAF data. The two datasets are available at different spatial resolutions. Hence, the algorithm computes error metrics at the appropriate spatial resolution, separately for each dataset. The outputs from Transearch module is generated at the Transearch data level and then converted to paths that provide us with estimated zonal flows. These flows are appropriately aggregated to the FAF spatial resolution for error evaluation.

The fusion process includes two components that generate log-likelihood measures as error metrics (the higher the log-likelihood the higher is the match). The first component includes a linear regression (or log-linear regression) model system that relates estimated Transearch flow to observed Transearch flow (see LL_{ij}^{T} in Figure 12). The linear regression model is estimated with independent variables related to origin and destination. Given that Transearch flow is an input-output flow (and not a transportation link flow), the objective is to decompose these flows into actual network level link flows by considering the various paths between each OD pair. The path

flows will allow us to determine the link flows. These flows are generated by employing a fractional split approach. The fractional split model is similar to the multinomial logit model structure but allows for fractional dependent variables (see Sivakumar and Bhat, 2002 and Bhowmik, et al. 2018; for examples of fractional split models). The actual path flow is unobserved; hence, a latent variable is considered, and the resulting link flows are matched with the observed flows.

These flows are generated at a finer resolution than the FAF zones. Hence, we need to appropriately aggregate them to a coarser level to compare the flows to observed FAF flows (see F_{OD} in Figure 12). The flows are again compared employing a linear regression structure at the FAF zone level. The aggregation is achieved based on the FAF zone of the origin and destination. The reader would note that the contribution of the FAF log-likelihood needs to be carefully computed (see LL_{OD} in Figure 12). While origin and destination zones have their corresponding FAF zones, the intermediate zones also have a FAF zone. Therefore, the allocation is obtained for an OD pair by apportioning the error to all FAF zones involved over the entire path set for that OD pair. Further, FAF zones can represent a large number of counties. Hence, a normalization is also employed (see LL_{ij}^{F} in Figure 12). Finally, the joint log-likelihood is provided as the sum of log-likelihood from Transearch component and FAF component (see F_{ij} in Figure 12). The proposed algorithm is programmed in Gauss matrix programming language.

Table 10 External Zones and Major Highways

Zone	States Included	Main Highways for Entering/Exiting Florida
1	South Carolina, North Carolina, District of Columbia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, New Hampshire, Vermont, Maine, Virginia, Part of Georgia	I-95
2	Indiana, Wisconsin, Illinois, Part of Georgia, Part of Kentucky	US 41
3	Tennessee, Ohio, Michigan, Part of Alabama	I-75, US 231
4	West Virginia, Part of Georgia	US 19, US 319
5	Louisiana, Texas, Part of Mississippi	US 90, US 98
6	California, New Mexico, Arizona, Part of Alabama, Part of Mississippi	I-10, US 331
7	Kansas, Colorado, Utah, Missouri, Part of Kentucky	I-75 N > I-24 W > I-57 N > I-64 W> I-70 W, US 27
8	Arkansas, Oklahoma, Nevada	I-75 N > I-20 W > 1-22 > 1-40 W > US 93
9	Nebraska, Wyoming, Oregon, Idaho	I-75 N > I-24 W > I-57 N > I-64 W > I-84
10	South Dakota, Montana, Washington	I-75 N > I-24 W > I-57 N > I-64 W > I-90 W
11	Iowa, Minnesota, North Dakota	I-75 N > I-24 W > I-57 N > I-64 W > I-70 W
12	-	Foreign (US, Canada, and Rest of Americas)

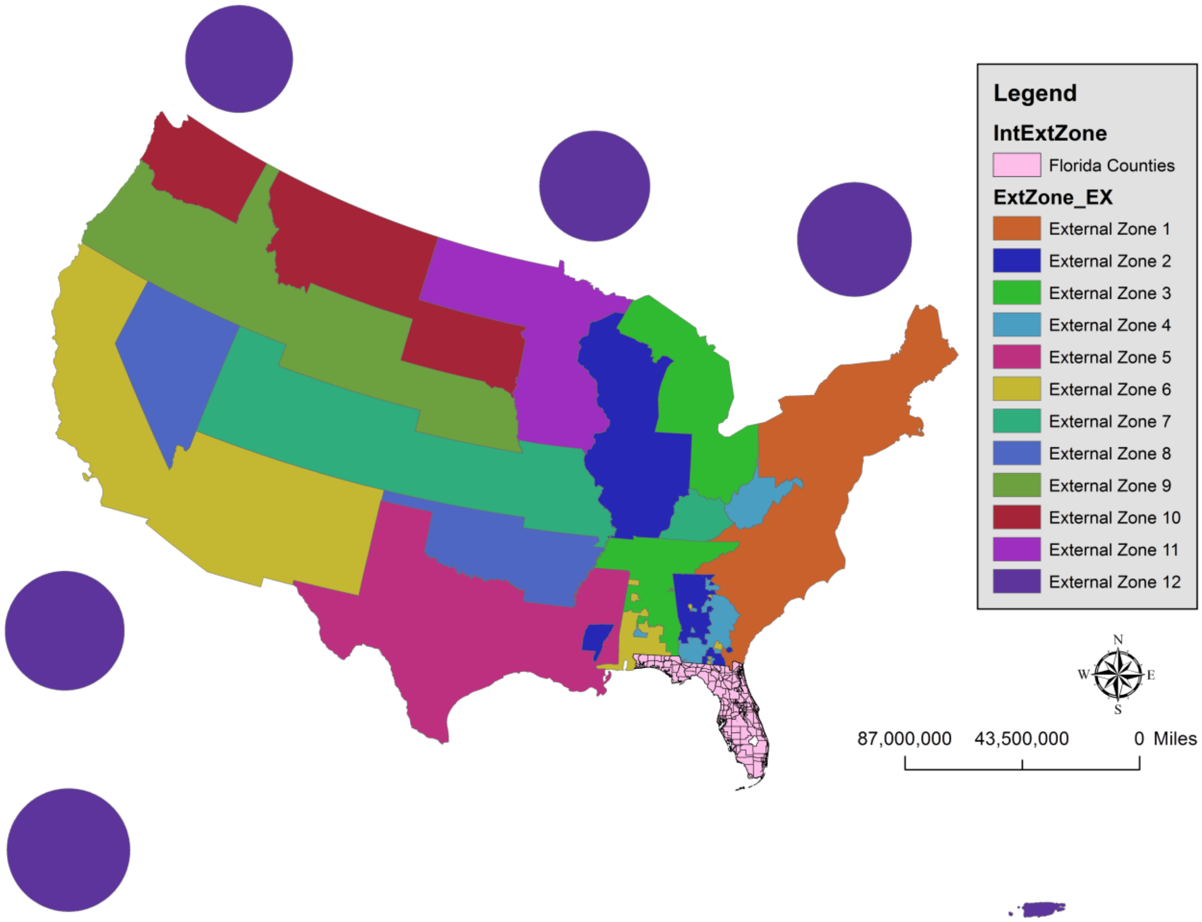


Figure 15 Spatial Zones Defined for the Algorithm

5.2.2 Implementation of the Algorithm for Internal and External Zones

In this study, a total of 12 external zones outside of Florida were considered. Of these, 11 zones were from the rest of US while one external zone included foreign regions (Canada, Mexico, and Puerto Rico). The 11 external zones within US was created based on national highways entering Florida (assuming the commodities entered/exited Florida from/to the outside regions through these major corridors). The apportioning was conducted in ArcGIS platform by visually

examining the routes entering/exiting Florida region. For states that are further away from Florida, the connections were made based on the most likely roadways entering Florida. Figure 16 shows the spatial zones considered in our study.

In the algorithm, flows between Transearch and FAF are connected by generating potential paths between the origin and destination of interest for Transearch flows. Between any origin destination pair, we allow for direct paths and one-hop paths (with one intermediate node). This leads to 6,097 (i.e. 67*67=4,489 internal + 67*12*2=1,608 external) potential O-D pairs for Transearch flows. For every OD pair, one direct path is considered. The number of indirect paths considered varies by the origin and destination zones. If both origin and destination are within the state, only zones within the state are considered as intermediate zones. Thus, for these O-D pairs a total of 65 one-hop paths exist (in addition to the one direct path). On the other hand, if one of the trip ends is external to the state, the number of one-hop paths is 66 (because all but one zone can serve as an intermediate point). Thus, our total path matrix has 6,097 direct paths (67*67+67*12*2) and 3,97,913 indirect paths (67*67*65 + 67*12*2*66). Thus, a total of 4,04,010 paths are considered in our analysis. The link flows generated are aggregated to FAF zone level. In our case, we have 145 (i.e. 5*5=25 internal and 5*12*2=120 external) potential origin-destination pairs at the FAF spatial resolution. The reader would note that consideration of higher number of external zones (as opposed to 12) or multiple intermediate stops would result in substantial computing time for the fusion algorithm.

5.3 Data Assembly Procedure

The primary data sources used in our study include FAF4 and Transearch. The base year for FAF4 data 2012 and it contains 43 commodity types. The FAF data also includes future year data from 2015 through 2040 at a five-year interval. Transearch data is available for the base year 2011 and contains 562 commodity types. Data assembly process for our research involved multiple steps as described below.

The commodity type differences between the two data sources were addressed by mapping the two data sources to a 13-commodity classification (for details, see Viswanathan, et al. 2008). These commodity types include: (1) agricultural products, (2) minerals, (3) coal, (4) food, (5) nondurable manufacturing, (6) lumber, (7) chemicals, (8) paper, (9) petroleum, (10) other durable manufacturing, (11) clay and stone, (11) waste, (12) miscellaneous freight (including warehousing), and (13) unknown.

FAF and Transearch data report foreign flows for different countries. Transearch reports flows to and from 3 foreign regions only: Canada, Mexico, and Rest of Americas (including Virgin Island and Puerto Rico). On the other hand, FAF4 reports flows to and from 8 international regions including Canada, Mexico, Rest of Americas (including Virgin Island and Puerto Rico), Europe, Africa, South-West and Central Asia, Eastern Asia, and South-East Asia and Oceania. Thus, for our data fusion exercise a compatible dataset of FAF is prepared by removing the foreign flows from those regions absent in the Transearch dataset.

Next, several independent variables were generated for model development. The socio-demographic variables include population and employment at origin and destination counties. The infrastructure and transportation facilities include number of intermodal facilities, number of

warehouses, number of ports, total length of roadway and railway in the origin or destination regions. The population data for the regions within the U.S. was drawn from U.S. Census Bureau, while the population of Canada and Mexico are drawn from Statistics Canada and from National Institute of Statistics and Geography, Mexico (INEGI), respectively. All these data are compiled for the year 2011. The National Highway Planning Network shape file was intersected with the external zone to get the total roadway length for 2010 for the external zones within Florida. For the fractional split model, the centroidal distance between the origin and destinations are estimated using ArcGIS network analysis.

Table 11 Comparison of Tonnages by Trade Type

Trade Type	Before Removing Foreign Flows			After Removing Foreign Flows		
	Flows (million tons)		Comparison (FAF4/TS)	Flows (million tons)		Comparison (FAF4/TS)
	TS	FAF4		TS	FAF4	
Export	**2.39**	**23.41**	**8.06**	**2.39**	**11.59**	**4.85**
Import	**5.96**	**43.71**	**6.63**	**5.96**	**25.57**	**4.29**
Inbound Domestic	153.98	123.7	0.8	153.98	123.7	0.8
Outbound Domestic	58.6	46.94	0.8	58.6	46.94	0.8
Within Florida	222.46	468.55	2.11	222.46	468.55	2.11
Total	443.39	697.95	1.58	443.39	676.35	1.53

5.4 Model Estimates for Base Year

The proposed algorithm is implemented separately for each commodity type. The estimated coefficients of the models for the thirteen commodities are presented in the Table 13. Please note that the model outcomes of the Transearch module demonstrate what factors affect the

overall county-to-county (production/attraction) flow while the FAF module outcomes provide the impact of path distances within a fractional model setting. For the sake of brevity, results for only two commodities: Paper and Clay and stone are described here in detail.

5.4.1 Commodity Type: Paper

The model results for Commodity type Paper are presented in the 10th column of Table 13. The t-statistics for the parameters were suppressed for space reasons. The reader would note that only parameters significant at the 90% level were retained.

5.4.1.1 TS Module

Our results indicate that destinations with increased population attract more paper flows. The impact is greater if the destination region has ports. It makes intuitive sense as paper consumption depends on the level of development and urbanisation of a geographic region. Number of intermodal facilities at origin or destination, ports at destination, and roadway coverage at origin also impact paper flow positively. The impact of road and railway coverage at destination region has negative impact on paper flows while number of ports and road length coverage at origin have positive impacts. This might be indicating that waterways are preferred mode of transport for pulp and paper products as these commodities are transported in bulk. The positive effect of intermodal facilities, both at origin and destination, imply that convenience of product storage is important for this commodity type. Interestingly however, the interaction of destination employment with destination number of warehouses deters paper commodity flow.

5.4.1.2 FAF Module

The path distances for both internal and external zones were negative as expected. This implies that the fractional freight flow from a distant production zone to a consumption zone is smaller than from a closely located production zone, everything else being equal. As paths with longer distances are less likely to be chosen for transporting freight; a larger flow will be allocated to direct paths.

5.4.2 Commodity Type: Clay and Stone

5.4.2.1 TS Module

Our model results indicate that origin population and destination employment negatively impact clay and stone commodity flow. It might seem counterintuitive at first glance. However, population and employment are surrogate indicators for regional growth and urbanization; the results are perhaps suggesting that Clay and Stone are less in demand in highly developed areas. On the other hand, the effect of origin county employment and the interaction of origin and destination employment indicates a higher commodity flow from counties with higher employment rates; a general reflection of high productivity levels associated with high employment rates. The interaction of number of warehouses at destination with employment at destination also has a positive impact on Clay and Stone flow between counties while the interaction of number of warehouses at destination with employment at origin has a negative impact.

5.4.2.2 FAF Module

Both coefficients for path distances are found negative indicating that distance is an impedance that has deterring impact on freight flows and higher proportion of the flows would be allocated to direct paths. The reader would note that while the external path distance coefficient is statistically insignificant we retained it as distance intuitively has a negative impact on allocation of flows.

5.4.3 Model Validation

To check if the fusion model outputs are providing reasonable results, for both commodities, we computed the ratio of fused link flows (FAF flow) with the observed Transearch flow and compared with the ratio of observed FAF and Transearch flow. For Paper, the observed ratio was 1.48 while the ratio with the fused flow is found to be 1.32. For Clay and Stone, the observed ratio is 1.35 while the ratio with the fused flow is 1.31. In both instances, the ratios are close to the actual ones illustrating that the fusion model outputs are reasonable.

Table 12 Estimation Results for the Statewide Joint Models

Module	Variable	Agricultural Products	Minerals	Coal	Food	Nondurable Manufacturing	Lumber	Chemicals	Paper	Petroleum	Other Durable Manufacturing	Clay and Stone	Waste	Miscellaneous Freight & Ware House
TS Module	Constant	3.043	-0.179	0.206	0.045	3.066	0.015	2.538	2.392	-0.056	0.027	0.047	0.019	0.044
	Employment (origin)	-[1]	-	-	-0.056	-	-	-6.984	-	-	-	-	-	-
	Road and rail length (origin)	-	-	-	-0.130	-	-	-	-	-0.246	-	-	0.004	-
	No of intermodal facilities (origin)	-0.300	-0.318	-	-	-	-	0.298	0.410	-	-0.033	-	0.025	-
	Road length (external origin)	-	-	-	-	-	-	-	0.803	-	0.035	-	-	-
	Population (external origin)	-	-	-	-	1.414	0.058	-	-	-	-	-	-	0.228
	Population (origin)	1.732	1.024	-	0.210	-	-	15.403	-	-	-	-	-	-
	No of ports (origin)	-	0.583	-	-	-	-	-0.543	-	-	-	-	-	-
	No of warehouse (origin)	0.780	-	-	0.294	2.637	-	-	-	1.091	-	-	-	-

[1] Variable not found significant

Module	Variable	Agricultural Products	Minerals	Coal	Food	Nondurable Manufacturing	Lumber	Chemicals	Paper	Petroleum	Other Durable Manufacturing	Clay and Stone	Waste	Miscellaneous Freight & Ware House
	Employment square (origin)	-	-	-	-	-	-	-	-	-	-	0.377	-	-0.401
	Population square (origin)	-	-	-	-	-	-	-	-	-	-	-0.484	-	0.328
	Employment (origin) * Employment (destination)	-	-	-	-	-	0.037	-	-	-	-	2.922	-	0.601
	Employment (origin) * No of warehouse (destination)	-	-	-	-	-	-	-	-	-	-	-4.879	-	0.340
	Population (external origin) * No of ports (destination)	-	-	-	-	-	-	-	-	-	-	-	-	-2.875
	Employment (destination)	-	-	-	0.383	-	0.100	-	-	-	-	-	-	-
	Employment square (destination)	-	-	-	-	-	-	-	-	-34.062	-	-2.462	-	-
	Population (external destination)	-	-	-	0.128	-	0.041	1.545	-	-	-	-	-	-
	No of intermodal facilities (destination)	-0.158	-	-	-	-	-	0.262	0.204	-	-	-	-	-

Module	Variable	Agricultural Products	Minerals	Coal	Food	Nondurable Manufacturing	Lumber	Chemicals	Paper	Petroleum	Other Durable Manufacturing	Clay and Stone	Waste	Miscellaneous Freight & Ware House
	Population square (destination)	-	-	-	-	-	-	-	-	3.466	-	-	-	-
	No of ports (destination)	0.294	-	-	-	-	-	-	0.103	-	-	-	-	-
	Road and Rail length (destination)	-	-	-	-	-	-	-	-0.422	-	-	-	-	-
	Employment (destination) * No of warehouse (destination)	-	-	-	-	-	-	-6.066	-4.039	-	-	0.549	-	-
	Employment (destination) * No of ports (origin)	-	-	-	-	-	-0.050	-	-	-	-	-	-	-
	Employment (external destination) * No of warehouse (origin)	-	-	-	-	-	-	-	-	-	3.716	-	-	-0.656
	Population (external destination) * No of ports (origin)	-	-	-	-	-	-	-	0.329	-	-	-	-	0.352
	Population (destination)	0.962	1.880	-	-	-	-	3.231	2.797	-	-	-	0.136	-

Module	Variable	Agricultural Products	Minerals	Coal	Food	Nondurable Manufacturing	Lumber	Chemicals	Paper	Petroleum	Other Durable Manufacturing	Clay and Stone	Waste	Miscellaneous Freight & Ware House
	No of warehouse (destination)	0.899	-	1.348	-	1.849	-	-	-	-	-0.164	-	-	-
	External destination	3.074	-	-	-	-	-	-	-	-	-	-	-0.017	-
	External origin	1.973	-	-	-	1.290	-	-	-	-	-	-	-0.028	-
	Destination Tampa	-	-	4.615	-	-	-	-	-	-	-	-	-	-
	Std. Err. (TS)	1.761	4.217	10.479	0.483	2.416	0.129	2.165	2.090	3.069	0.334	0.404	0.271	0.751
FAF Module	Path Distance (External Zones, KM)	-0.287	-0.291	-0.099	-0.291	-0.048	-0.004	-0.220	-0.048	-0.052	-0.364	-0.087	-0.290	-0.183
	Path Distance (Internal Zones, KM)	-0.031	-0.058	-0.350	-0.047	-0.292	-0.040	-0.293	-0.291	-0.286	-0.110	-0.014	-0.049	-0.003
	Std. Err. (FAF)	3.652	38.751	12.407	9.419	1.739	4.356	1.708	1.801	27.616	2.988	7.180	13.169	5.080

5.5 Framework for Future Year Predictions

5.5.1 Implementation for future years

The main emphasis of the paper is to develop a framework to predict future freight flows. Toward that end, the approach employs the base year model estimates obtained for each commodity to generate future year predictions. The procedure involves a series of steps. First, the prediction exercise requires us to generate independent variables for future years. For the forecasting exercise, the variables that were found to be significant in the models were projected into the future. For population, we investigated the population growth rate over years from the population data provided by the US Census Bureau. For this purpose, we collected population data from 2010 to 2015. An overall growth of 4% was observed for the 5-year interval. For the simplicity of estimation, we assumed a growth factor of 4% in each five-year interval for projecting the future year population. For external zones (outside of Florida), we assumed the same growth rate (4%). Similarly, a growth rate of 6.9% for employment was computed. Furthermore, we used GIS shapefile for The National Highway Planning Network to get the roadway length for the internal and external zones. The GIS file provided the highway route for the local and highways all over the US. Increase of roadway length is assumed to be at a rate of 2.7%, and 1.5% for the within Florida, and outside of the Florida, respectively. The total roadway length was then projected for future years in each five-year interval from 2015. For fixed facilities such as seaports, airports, and intermodal facilities, we assumed that no change in their counts occurred for the future years.

Second, the Transearch module provides the flows at the zonal level for forecast years. The appropriate model estimated (linear or log-linear) is employed for the prediction exercise. Third, the probabilities of choosing a route is obtained from the fractional split model corresponding to the FAF module. Then using the link and path relationship matrix, the county level fused link flow is predicted. These link flows provide the freight flows at a zonal level for the Florida state including county level flows in the state and zonal level flows from external regions (rest of the country and foreign flows). The same procedure was repeated across all commodity types for all future years starting from 2015 to 2040 in 5-year increments. The reader would note that, for a small number of cases, the predictions from log-linear models were much larger than values reasonable for the commodity. In that case, we converted these outliers to a meaningful value based on observed trends.

5.6 Prediction output

Table 14 presents results from the prediction exercise for (1) Paper and (2) Clay and Stone. In addition, the spatial variation of the predicted flows (color-coded categorization) across the years are plotted in Figures 17-20.

5.6.1 Prediction for Paper Commodity

We can see from Table 14 that paper commodity flows increase marginally from 2011 through 2015. From 2015 through 2025, the demand remains stable. Subsequently there is a minor drop in paper flow. This is intuitive, as the demand for paper is expected to go down, particularly

in North America with increasing digitization. Figure 17 shows the change in paper flows originating in each Floridian county across the years. The examination of the figures offers interesting results. Until 2030, the northeastern counties are expected to observe the largest increase in paper flows. After that, large increase is expected to occur in the Miami region. Figure 4 focuses on paper commodity flow attraction in Floridian counties. A steady increase in paper product consumption for the Central Florida region can be observed from the figure. Interestingly, the North Florida and South Florida regions offer contrasting trends.

Table 13 Comparison of Observed and Predicted Tonnage by Commodity

Year	Observed/Predicted Flow (Truck Mode Only)	Tonnage (million tons)	
		Paper Product	Clay and Stone
2011	Observed TranSearch Flow	7.79	33.32
	Predicted Flow	10.40	50.00
2015	Predicted Flow	10.76	53.29
2020	Predicted Flow	10.75	54.81
2025	Predicted Flow	10.76	57.45
2030	Predicted Flow	10.65	62.44
2035	Predicted Flow	10.66	74.04
2040	Predicted Flow	10.69	112.36

5.6.2 Prediction for Clay and Stone Commodity

A similar comparison exercise is also undertaken for Clay and Stone commodity flows. From Table 14, we can see a significant increase in the demand for this commodity type; an average increase of 16% was predicted. Compared to the base year (2011), the tonnage transported between counties is expected to be doubled by 2040. Similar to Paper demand, North-Eastern Florida region is expected to have a high growth rate for flow of clay and stone (see Figure 18).

The growth rates are likely to be higher for Central Florida and Miami region while a mild growth rate is expected for incoming flow to North Florida.

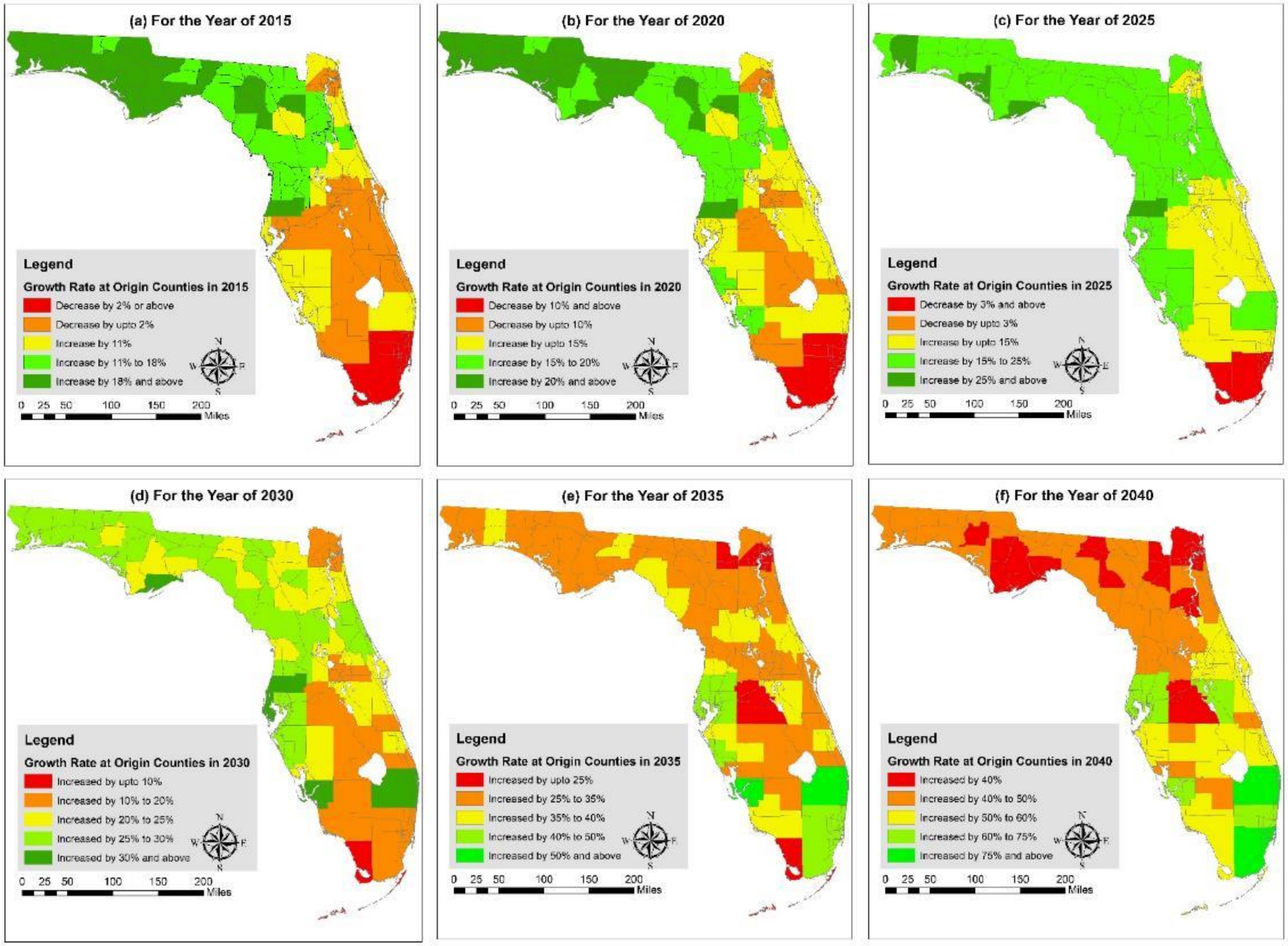


Figure 16 Growth of Predicted Link Flows at Origin Counties by Year Compared to the Base Year for Paper

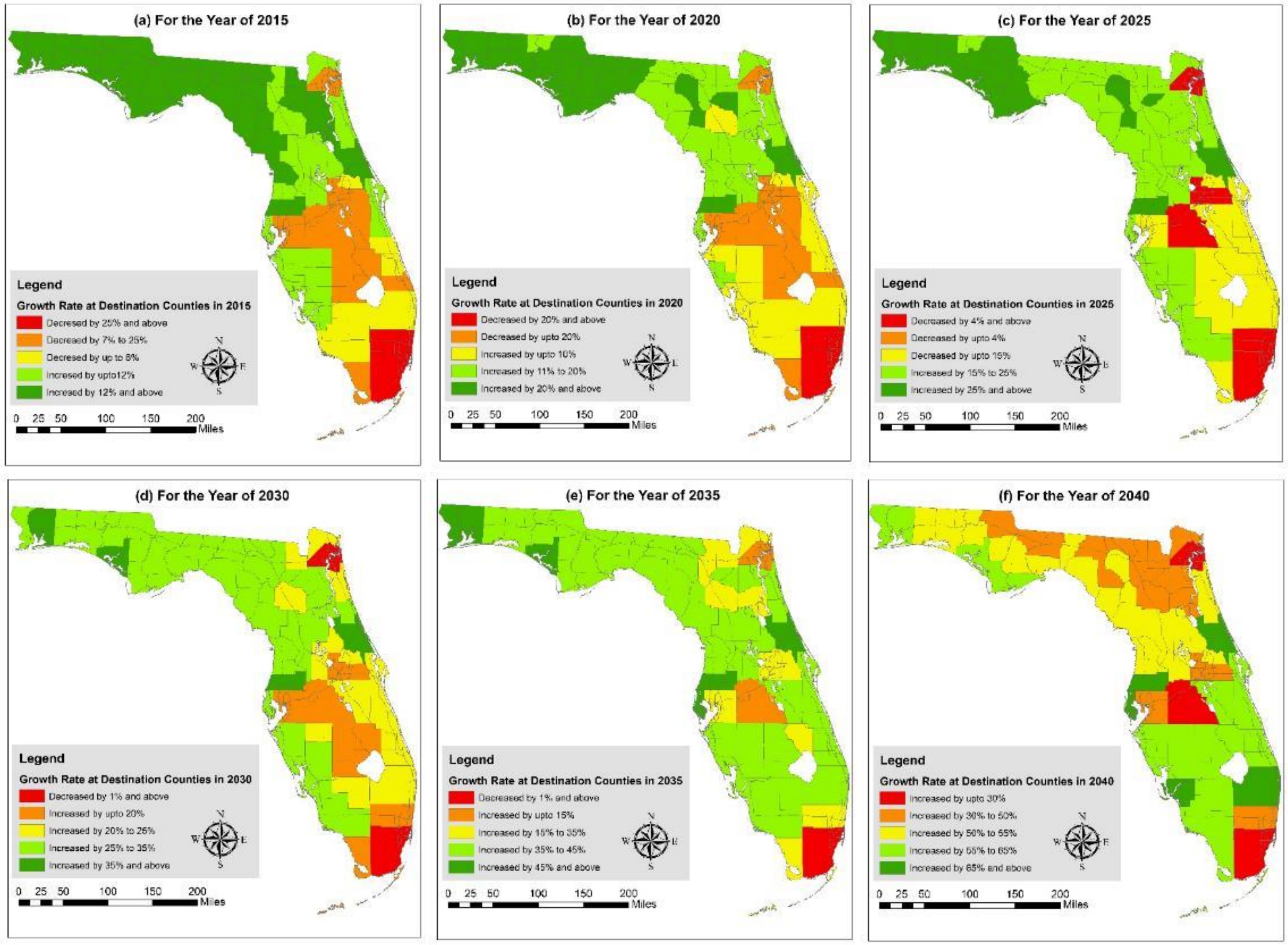


Figure 17 Growth of Predicted Link Flows at Destination Counties by Year Compared to the Base Year for Paper

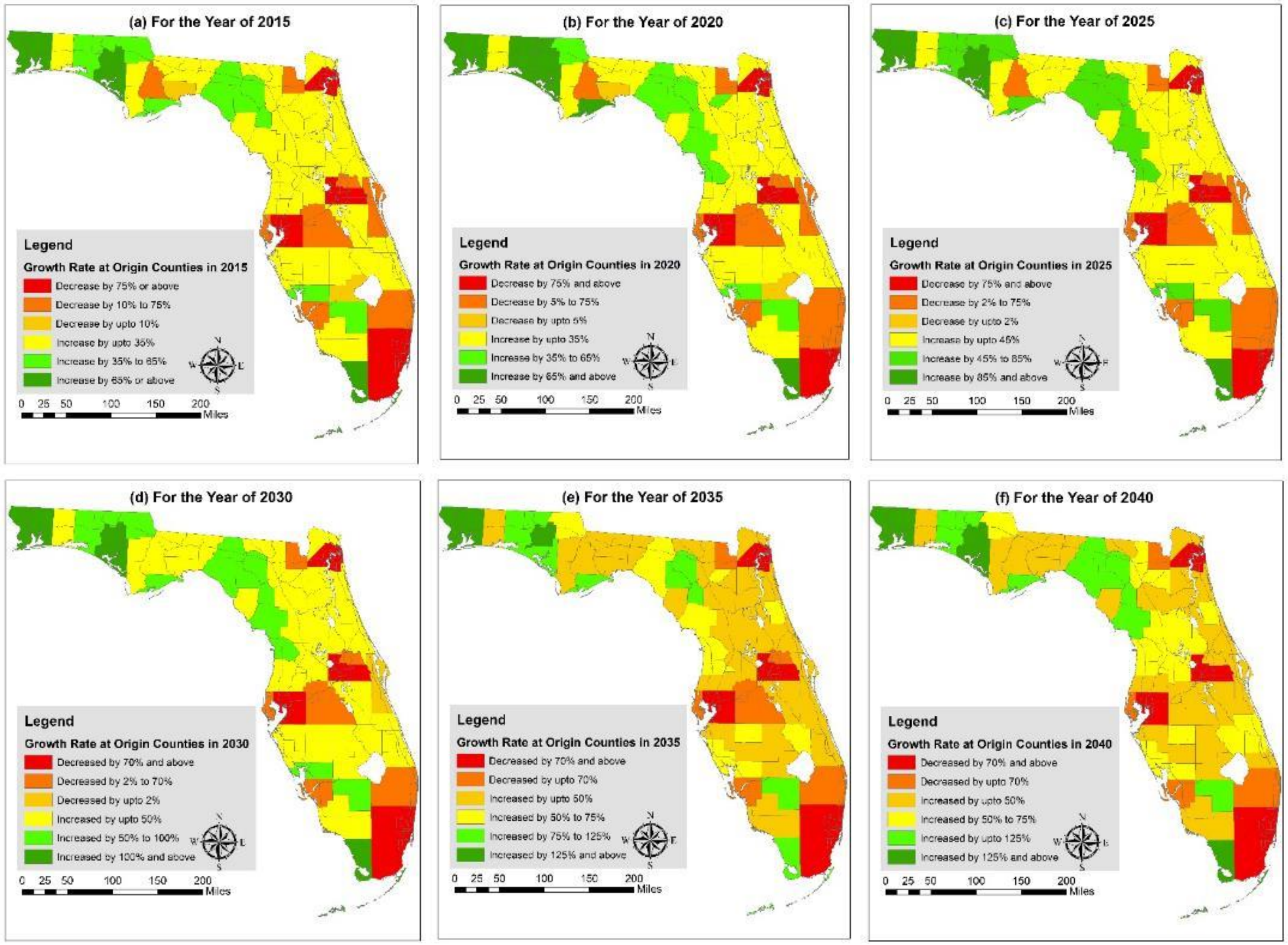


Figure 18 Growth of Predicted Link Flows at Origin Counties by Year Compared to the Base Year for Clay and Stone

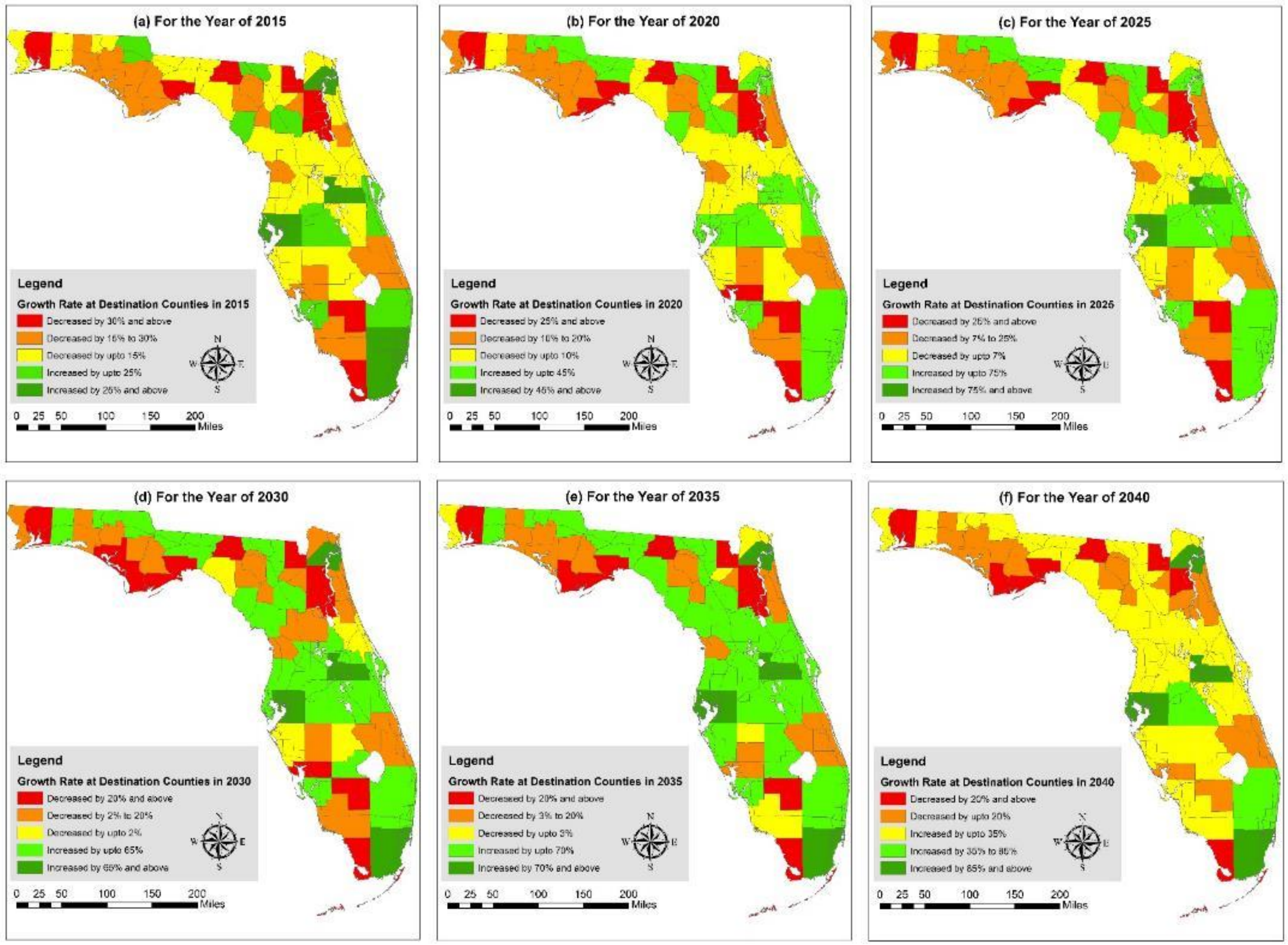


Figure 19 Growth of Predicted Link Flows at Destination Counties by Year Compared to the Base Year for Clay and Stone

5.7 Summary

At the national level, Freight Analysis Framework (FAF) data built by Federal Highway Administration (FHWA) provides detailed freight flow predictions by mode, location, commodity type for future years. However, for any statewide planning, the spatial resolution available in the FAF framework is inadequate. Purchased data from vendors such as IHS Markit (Transearch data) provides a finer spatial resolution. However, these data products are prohibitively expensive. Thus, it is beneficial to develop innovative approaches to disaggregate FAF data while also enhancing the behavioral underpinnings of the data. Along this direction, in our earlier work (6), we developed a data fusion algorithm that draws from FAF4 and Transearch data. The algorithm was developed as a proof of concept and is implemented for the base year (2011) only for within Florida truck flows for two commodities. However, for a complete representation of freight flows on Florida's transportation system, it is essential for state planning agencies to consider out-of-state and foreign flows within the overall freight planning platform. Towards this end, we extend the algorithm proposed in (6) to consider spatial region in Florida, outside Florida (in the US and foreign regions). In our study, we resort to very detailed resolution inside Florida, and larger aggregate zones to represent zones outside Florida and foreign regions. Furthermore, long-range planning of freight flows requires estimates of flows for future years. Thus, in our study, in addition to developing the algorithm for base year data fusion, we employ the proposed algorithm to undertake future year predictions at the selected spatial resolution for various commodity types. Specifically, the current study generates estimates for freight flows from 2015 through 2040 in

five-year increments. The generated estimates are evaluated to understand how freight flows are likely to evolve in the Florida region.

CHAPTER SIX: MODELING FLOWS AT A VERY FINE SPATIAL RESOLUTION

In the previous chapter freight flows for various commodity types were generated at the county level resolution. In this chapter, we document the generation of estimates at a finer statewide traffic analysis (SWTAZ) zone resolution.

6.1 Defining Spatial Resolution of Origin and Destination - SWTAZs

The state of Florida is divided into 67 internal zones based on county boundaries. The external zones are divided into 12 zones. The 11 zones in the US are defined based on the highway entering to Florida, as shown in the Figure 21. The remaining zone consists of the Alaska and the foreign zones of Canada and Mexico. To disaggregate the study used Statewide Traffic Analysis Zone (SWTAZ) defined by Florida Department of Transportation. Florida Statewide system divided Florida into 8518 TAZs. We have used these SWTAZs in our system for disaggregating the internal zones (i.e. the 67 Florida counties). For external zones, we used Bureau of Economic Analysis (BEA) Regions as the TAZs in our system. We have 312 BEA regions adopted as zones for disaggregating the 11 external zones within US. For remaining one external zone including Canada and Mexico, we consider 5 zones. The provinces in Canada are grouped into three zones as (1) Yukon, British Columbia, Alberta, Northwest Territories, Saskatchewan; (2) Nunavut, Manitoba, Ontario and (3) Quebec, Newfoundland and Labrador, Nova Scotia, New Brunswick, Prince Edward Island. In Mexico, we considered two zones for the external zones. These zones contain the states of the Mexico as the following groups (1) the states in the northern portion of

the Mexico (total 12 states); and (2) the states in the southern portion of the Mexico (total 20 states). Therefore, we have 8518 zones inside Florida, 312 TAZs in the rest of US and 5 zones outside of US (total 8835 TAZs) in our system.

6.2 Modal Share for Truck

In our disaggregation effort, we explicitly focussed on truck flows. A series of data preparation steps were employed for obtaining truck flows. First, the mode share for truck at the FAF region level is estimated from the observed FAF dataset. The mode share is estimated as a percentage of the mode with respect to the total aggregated flow for all modes. To estimate the county level mode share, we examined the origin and destination of the flows. The county level mode share is estimate then for each OD pair based on the location of the origin and destination counties within the FAF regions. Specifically, the FAF mode shares at a region level were used to expand the mode share to the entire county-to-county pair ensuring the assumption as follows. For the flows that have both origin and destination counties in same FAF region, the mode share is assumed to be the same as the corresponding FAF regions. However, if the origin and destination counties of a flow are located at two different FAF regions, the mode share at county-to-county level OD pair are assumed to be the same as the average mode share of the two FAF regions. Eventually, the link flows carried by the truck are obtained by the multiplication of the estimated mode share for truck and the link flows obtained from the joint model prediction.

6.3 Disaggregation Framework

From the joint model estimation, we estimate the county to county link flows. The link flows are obtained for each commodity separately. With post processing for truck mode share the link flows are disaggregated to obtained final disaggregation at a finer spatial resolution. A factor multiplication method is used to disaggregate the county level flow to SWTAZ flow. In this method, a two-stage factor multiplication is proposed. The factors are estimated both at origin and at destination level. The proportion of the flows from a TAZ with respect to the total flow originated from a county is considered the factor at the origin level. Similarly, the flow proportion of a county among the TAZs are the factor for that specific TAZ at the destination level. Multiplying the factors at origin and destination level, the county level origin-destination pair flows are disaggregated in to the TAZ level origin-destination pair flow among the corresponding TAZs that exist in those counties. The factors are estimated using a random utility factional split model approach. TAZ level exogenous variable are used in the fractional split model.

6.3.1 Fractional Split Model

The proportions at origin and destination level are estimated in separate models using fractional split models. Each of the county to county origin-destination pairs are further split into finer and smaller TAZs. Therefore, multiple OD pairs will be obtained from a single county to county OD pair. Considering each TAZ level OD has a proportion to share the flows in between that corresponding county to county pairs. The proportion would be in the range of from zero to one. Additionally, the summation of the proportion within a county would be one. This proportions are the dependant variable for our fractional split model.

Now, mathematically, let, $\boldsymbol{y_{qi}}$ be the proportion of originated/destined flow from a TAZ within a county. Where, q is the county of origin/destination and i is the TAZ within the county

Hence, mathematically, $\boldsymbol{0 \le y_{qi} \le 1}$, and $\sum_i \boldsymbol{y_{qi} = 1}$

If $\boldsymbol{X_{qi}}$ be the vector for the independent variables, the model structure is as follows. The expected proportion from an originated SWTAZ or to a destination SWTAZ is given by the following equation.

$$\boldsymbol{E(y_{qi}|X_{qi}) = \frac{e^{\beta X_{qi}}}{\sum_j e^{\beta X_{qj}}}} \tag{12}$$

The proportion for the disaggregated level SWTAZ within a county is maximized using a multinomial logit choice approach. The log-likelihood is given by the following equation.

$$\boldsymbol{\mathcal{LL}(\theta) = ln\left(\prod_{i=1}^{n} (\widehat{y_{qi}})^{y_{qi}}\right) = \sum_{i=1}^{n} ln(\widehat{y_{qi}})^{y_{qi}}} \tag{13}$$

The model estimates can be used to generate the future years prediction of the proportion of the total flow. To do so, we used the significant variable in the model for the base year and extrapolated the variable for the future years as described in the previous chapter.

6.3.2 Disaggregation Method

From the fractional spilt models, we get the estimated proportion for each origin/destination. Let us say we have a freight flow from an origin SWTAZ “i” in the origin county “p” to a destination SWTAZ “j” in the destination county “q”. From the fractional split model for origin SWTAZ, we can estimate the originated proportion ($\hat{y}_{pi}$) of the SWTAZ “i” with

in county “p”. Let, “F_{pq}” be the freight link flows between the county pair “p” and “q”. Thus, originated link flow from SWTAZ “i” is given by the following equation.

$$\boldsymbol{F_i = \hat{y}_{pi} * F_{pq}} \quad (14)$$

This originated flow is thus further apportioned into the SWTAZs at the destination county according the proportion estimated through the fractional split model at the destination end. Therefore, the flows from SWTAZ “i” to the destination SWTAZ “j” can be calculated using the equation below.

$$\boldsymbol{F_{ij} = \hat{y}_{qj} * F_i} \quad (15)$$

Figure 22 presents the illustration of the methodology of the data fusion and the two-stage disaggregation to SWTAZ level origin-destinations.

6.4 Data Preparation

The data are threefold in this study; 1) Input data for the disaggregation method, 2) dependent variables for the fractional split models, and 3) Independent variables for the fractional split models. The following subsection will describe the data preparation processes for these data in the study step by step.

6.4.1 Disaggregation Input Data

First, we prepared the data for the input in the disaggregation procedure. From the econometric data fusion models, we estimated coefficients for the variables influence the freight flows for the base year (2011). The regression model coefficients in the TS module are used to

estimate the Transearch like county to county freight flows. These freight flows are divided into the path flows between the origin-destination pairs using the coefficients from the fractional split model in the FAF module. In final step we converted the path flows to link flows using the Link-Path-Matrix. (Please see the previous study for details). These link flows are the county to county freight flows on the transportation network. For the system considered in the study, we have freight flows for 6097 OD pairs (67*67 internal-within Florida and 12*67*2 external-either origin or destination is in outside of Florida). These flows dataset with 6097 record are the input data for our disaggregation procedure.

6.4.2 Dependent Variable: Fractional Split

The dependent variable for the fractional split model is the proportion of freight flows originated from (or destined to) a SWTAZ within the origin (or destination) county. We have access to the detailed data for the state of Florida. We aggregate the data for each county level flows originated from the counties. The ratio of the freight flow originated from a SWTAZ by the freight flow originated from the county contains the SWTAZ provides the proportion factor for the origin SWTAZs. Similarly, for the destination SWTAZs, we estimated the proportion factors for destination. This proportion factors are the variable that we would like to produce using the fractional split models. Therefore, these variables were used as the dependent variables for the fractional split models.

6.4.3 Identifying Outliers in OD

The simplistic method of generating proportion allow all the SWTAZ to have a proportion factor. Within a county there are multiple SWATZs ranging from 5 to 809. On an average, there are 112 SWTAZs (standard deviation 144.55) in a county. Therefore, county consisting with higher SWTAZs would provide a very low probabilities for a large number of SWTAZs. With a critical exploration of the fractions, we observed many the flows are allocated in to SWTAZ level OD pairs that shares a very low amount of flows. These flows are as low as 5 kg. For a commodity annual freight flow of 5 kg is not a realistic value. Hence, we adjusted the proportion factor obtained from the fractional split model estimates. Since we are using the data fusion of FAF and Transearch data, we took advantage of Transearch disaggregated data in adjusting the proportion factors. We had access to SWTAZ level disaggregation of Transearch dataset for the state of Florida. This data was for the base year (2011) and future years (2015-2040). We studied the number of SWTAZ exists as the origin/destination, and the number of SWTAZ origin-destination pairs. Table 16 shows the number of OD pairs found in the Transearch dataset. We examine the bottom 1 to 15 percentiles of the predicted fraction for each of the counties. We removed the small fractions using a threshold value based on the percentile value within each OD pairs. For a single commodity, the threshold percentile value was fixed across all the OD pairs. However, the threshold varies from commodity to commodity. The threshold is set within the range of 98th percentile to 85th percentile for the different commodities. While selecting the threshold, we ensure that each of the internal or external zones have at least five SWTAZs with non-zero proportion. This confirms that all the zones will have flows in disaggregation process. In next step, the number of OD pairs exist after restructuring the factor is investigated for all the years of 2011,

2015, 2020, 2025, 2030, 2035, and 2040. We ensure the number of OD pairs are close enough to the one in the Transearch dataset. Eventually, the proportion of flow is re-estimated within the SWTAZ exists after the above stated preprocessing of the dataset.

For Coal commodity, the observed data was very small. Only 38 flows are recorded in the dataset. Therefore, we could not estimate the models for the commodity. We assumed the same proportion is valid for the commodity across the year. As a result, only for Coal, we used deterministic method to estimate the origin and destination fractions for the SWTAZs.

6.4.4 Independent Variable: Fractional Split

In the econometric data fusion models used detail socio-demographic, economic, data, transportation infrastructure and facility data. We considered simple demographic information for the fractional split models. We have used population and employment data for the origin (or destination) SWTAZs within Florida, and population of the origin (or destination) SWTAZs for the external zones at the origins (or destinations). We do not have access to any data that directly provides population and employment at the SWTAZ level. To generate the desired data for these variables, we used GIS preprocessing of the variables. We used GIS shapefile of census tract from tiger shapefile. Using the census tract shape file, we overlay the SWTAZ GIS file for the internal and external zones within the USA. From the intersection of the map area, we identified the area SWTAZ that belongs to the portion of the census tract zones. The portion of the census tract zones are estimated, and area of the covered census tract zones and the population density of those census tract zones are identified. If the area of a SWTAZ "i", intersects with the area of "k"th portion of

the census tract zone of "j" census tract zones (A_{jk}), mathematically the area of SWTAZ "i" is given by the following equation.

$$\boldsymbol{A_i = \sum_{j \in CT,\ k \in IZ} A_{jk}}\ ; \quad \forall\ \boldsymbol{CT = \{Set\ of\ intersecting\ Census\ Trac\};} \quad \boldsymbol{IZ\ = \{Set\ of\ intersecting\ zones\}} \tag{16}$$

The population of a particular SWTAZ is estimated by the proportion of area and the population density of the Census Tract Zones. For multiple zone involvement, the population for different area intersection is estimated and summing all the estimated population the total population for the SWTAZ is calculated. Let, "δ_j" be the population density of the census tract zone "j". Mathematically the population for the SWTAZ is defined as follows.

$$\boldsymbol{Pop_i = \sum_{j,k} \delta_j * A_{jk}} \tag{17}$$

Using the above procedure, the population of the SWTAZs are estimated and prepared for the variables in the fractional split model. Similar approach is used for the employment for the internal (within Florida) SWTAZs.

For the external zone outside of the USA, we collected the population data by state or province for the Canada and the Mexico. We used the database from the Statistics Canada and the National Institute of Statistics and Geography, Mexico (INEGI), for population data for the Canada and the Mexico, respectively. To get the population for the SWTAZs that are outside of the USA, we aggregated the population data of the states/provinces involved in the defined SWTAZs by the study for those international zones. We did not consider the employment variable for the external zones.

The base year population data and employment data are used to estimate the models. However, the future year prediction of the proportion needed the variable to predict the variable for the future year by extrapolation. We investigated population data from 2010 to 2015 from the US Census Bureau. We examine the population growth rate in the dataset. From our examination, we found a 4% growth rate for each 5 years. We used a simplistic assumption of 4% growth rate factor in each five-year interval for projecting the future year population. We used a unified growth rate factor for all the internal and external zones in our study. Similarly, a 5-year growth rate of 6.9% for employment was computed.

6.5 Outcome of Disaggregation to SWTAZ

6.5.1 Estimates for Model of Proportion for SWTAZ

24 different fractional split models were estimated for 12 different commodities both at origin and destination proportion for the SWTAZs within the counties in Florida and external zones. We estimated the models with population and employment for the SWTAZs located within the Florida counties, and population for the SWTAZs located at outside of the Florida. Table 15 presents the result of the models. Columns 2-4 represents the model estimates at origin SWTAZs for the corresponding commodity in the row. Again, columns 5-7 are the estimates for the destination SWTAZs for each commodity in the rows. From the table we found that the population and employment both are positively proportional to both the originated flow from, and the destined flow to, a SWTAZ. The population at the origin or destination SWTAZ are found to have a higher

impact for the originated and destined flow for the SWTAZs within Florida than the SWTAZ at the external zones. This effect is intuitively expected since the SWTAZ within Florida are relatively smaller compared to the SWTAZs at the external zones. Moreover, the half of the total flows are found to be the internal-internal flow (i.e. within Florida flow). As a result, it is not surprising to observe the higher impact of the population in the internal SWTAZs' proportion. Besides, all the commodity but the agricultural product, waste, and the miscellaneous freight have the similar trend for the effect of the employment. For the commodities except these three, the effect of employment is found to be higher for the proportion of the origin SWTAZs than the effect of populations. This implies that the employed population is contributing directly or indirectly to produce these commodities.

Again, for agricultural product, waste and miscellaneous freight are likely to have lower effect of employed population than the that of whole population, on the proportion originated from the SWTAZs within a county at Florida, or within an external zone. The area with higher employed population is likely to have more industry to create more jobs for the population at the location. It is intuitive to think that the firming and agricultural industry will be in less industrially developed areas. Therefore, agricultural products are likely to produce in the less employed area than the higher employed area. Hence, it makes proper sense to have such outcome of effect of employment compared to the over all population.

Again, for the destination the proportion factors are influenced more by the employed population of the SWTAZs than the overall population of the SWTAZs. This implies that, the destination SWTAZs have more industries are likely to generate more economic growth that eventually increasing the demand of all the commodities. The higher magnitude of employment

estimates indicates the developed areas are more likely to consume more commodities than the less industrially developed regions.

Table 14 Model Estimates for Generating Proportion of Originate and Destined Freight Flow

FCC	Origin Link Flows Proportion Model			Destination Link Flows Proportion Model		
	SWTAZ Population (in millions)		SWTAZ Employment (Within Florida) (in thousands)	SWTAZ Population (in millions)		SWTAZ Employment (Within Florida) (in thousands)
	Within Florida	External Zone		Within Florida	External Zone	
Agricultural products	1029.23	1.90	0.94	985.77	2.72	1.69
Minerals	1013.54	0.71	1.10	948.77	1.85	1.44
Food	901.41	2.42	1.55	822.01	2.22	1.90
Nondurable manufacturing	1339.44	2.31	1.98	809.45	-	-
Lumber	876.33	1.58	1.29	1021.77	1.59	1.56
Chemicals	600.25	2.13	1.49	802.83	2.68	1.94
Paper	885.23	1.99	1.80	871.8	1.58	2.01
Petroleum products	521.29	2.49	2.26	884.51	1.57	1.61
Other durable manufacturing	853.99	2.77	1.55	936.15	2.38	1.81
Clay and stone	922.02	2.00	1.21	972.09	1.94	1.59
Waste	1029.23	1.90	0.94	985.77	2.72	1.69
Miscellaneous freight & warehousing	1029.23	1.90	0.94	985.77	2.72	1.69

6.5.2 Disaggregation to SWTAZ

For the disaggregation exercise, we used the above stated methodology to disaggregate the flows in to SWTAZ level flows. We have tested if the disaggregated flows are matching with the county level link flows. Furthermore, we generate GIS map for the originated and destined flows for the SWTAZ flows within the Florida states to visualize the effect of the disaggregation

procedure. The figures 23 and 25 present the visual assessment of the disaggregated flows for agricultural product.

In first step, generate a color-coded spatial flow originated from both the counties and the SWTAZs separately. Figure 23 presents the comparison of the agricultural product flow for the SWTAZs within the counties of the Florida. From the graphical representation, we can observe a clear difference of the freight flow generation. The disaggregated flows to SWTAZs provides a detail representation of the freight flow generation location at a disaggregated spatial zone, from which area within the county. In northern part of Florida, the freight flow originated from county is observed as lower rates. With a disaggregated flow at SWTAZ level, the originated flow shows comparatively higher flows in most of the SWTAZs, with a very small number of SWTAZs with the similar trend found in the county level flows. The SWTAZs in the southeast Florida shows the similar trend. The flows in are relatively higher in the southeast Florida near the coasts. This area comprises with the midsize and large cities. From figure 23, the Miami area is shown in enlarged figure to have a closer look. However, the figure does not confirm any significant relation of total flows from a SWTAZ with the location of the nearby cities.

In figure 24, we generated the similar visual representation of flow destined to the Florida counties (at left of the figure) and to the SWTAZs in the Florida (at the right of the figure). In Figure 25 and 26, we examined the three important regions, with the flow density originated and destined to the SWTAZs. The figure also shows the city center location. The cities are classified in five class for the state of Florida. These are (1) Very small city: Population of below 10 thousand, (2) Small city: Population of 10-50 thousand, (3) Midsize city: Population of 50-100 thousand, (4) Large city: Population of 100-250 thousand, and (5) Very large city: Population of

more than 500 thousand. In figure 25, we observed that the location of the city increases the flow density of the nearby SWATZs within a certain buffer area. It is also observed that a larger city is likely to have a higher influence on the flow density from the nearby SWTAZs. Multiple small city or midsize city that are close enough, have similar effect of a large city. These group of cities together likely to increase the flow density in the neighbouring SWTAZs. In Orlando region, multiple cities are found with high flow density SWTAZs. Also, in Miami region, the coastal area is mostly comprising with the cities ranging from small to large cities. These coastal areas are thus found with heavy flows. In figure 26, we examine the similar aspect for the agricultural product destined to SWTAZs within Florida. The location of city and size of city shows similar influence in the freight flow density attraction to the SWTAZs.

In the final exercise, we plotted the freight flow on the roadway network. We assumed that the link flows are carried by the freight trucks from the centroidal point of a zone to another zone is through the shortest path between these two points. Therefore, we assigned the entire flow between two zones on the shortest path found in the network analysis from the GIS analysis. In figure 27, we generated the network flows both for link flows aggregated at the county level (left), and the link flows disaggregated at the SWTAZ level (right). For the comparison we selected a SWTAZ (SWTAZ ID 8045) that has a geometric centroid closer to the geometric centroid Miami-Dade County. The SWTAZ also has the highest flow among the other SWTAZs within the Miami-Dade County. The SWTAZ contains almost 33 percent of the total flow from the Miami-Dade County. For comparison, in the first map (on the left), we plotted the same amount of flow as the SWTAZ 8045 produce and assigned the flow on network from the Miami-Dade County to the other counties within the Florida. In the second map (on the right), we plotted the flows from

SWTAZ 8045 to the other SWTAZs within the Florida. Therefore, the flow shown on the first map in the figure 27 is the aggregated flow at county level, which is disaggregated in the second map in the figure. From the figure, we found that the disaggregated flow provides a detail flow information on the network with assigned flow everything on the shortest path. This implies that the disaggregated data provides a better OD trip of freight truck moving among the SWTAZs within the state of Florida. Moreover, the figure also shows that there are many local areas that are expected to have flows outside of the interstate highways. These are getting focused in the disaggregated flows. Especially at the larger city and at its vicinity, the SWTAZs are attracting flows to create more flows on the state and other highways through the surrounding regions of the cities. This will help network assignment procedure for the network assignment for the freight carrying trucks. The logistic planning will be benefited from this local flow information, as well.

Table 15 Number of Unique Export and Import Flows Compared to Flows within the Country by Year (in 100 million)

FCC	2011	2015	2020	2025	2030	2035	2040
Agricultural products	4.80	4.84	4.86	4.88	4.88	4.88	4.88
Minerals	1.31	1.31	1.32	1.31	1.31	1.31	1.31
Food	7.62	7.60	7.62	7.63	7.63	7.63	7.63
Nondurable manufacturing	10.53	10.53	10.60	10.65	10.65	10.65	10.65
Lumber	6.90	6.97	6.97	7.00	7.00	7.00	7.00
Chemicals	4.16	4.16	4.19	4.21	4.21	4.21	4.21
Paper	12.06	11.93	12.01	12.07	12.07	12.07	12.07
Petroleum products	1.38	1.39	1.40	1.42	1.42	1.42	1.42
Other durable manufacturing	28.72	28.84	29.27	29.41	29.41	29.41	29.41
Clay and stone	6.70	6.70	6.74	6.74	6.74	6.74	6.74
Waste	1.56	1.56	1.54	1.51	1.51	1.51	1.51
Miscellaneous freight & warehousing	59.15	57.36	57.44	57.49	57.49	57.49	57.49

6.6 Summary

We disaggregated the county level link flows to the Statewide Traffic Analysis Zone (SWTAZ) as the finest spatial zones. For the SWTAZ level disaggregation with the same method would led us to disaggregate the 17-zone system to an 8835-zone system. Florida Statewide system divided Florida into 8518 TAZs. We used a simplistic approach of proportion weighing method to disaggregate the county level flows. We checked back the number of disaggregated origin-destination with the one of disaggregated Transearch flows for the truck mode only.

From the joint model estimation, we estimate the county to county link flows for each commodity separately. A two-stage factor multiplication method is used to disaggregate the county level flow to SWTAZ flow. The factors are estimated both at origin and at destination level. The factors are estimated using a random utility factional split model approach. TAZ level exogenous variable are used in the fractional split model. 24 different fractional split models were estimated for 12 different commodities both at origin and destination proportion for the SWTAZs within the counties in Florida and external zones. We adjusted the proportion factor obtained from the fractional split model estimates for county consisting with a large number of SWTAZs using threshold value for the different commodities.

In the model estimates, the population and employment both are positively proportional to both the originated flow from, and the destined flow to, a SWTAZ. The population at the origin or destination SWTAZ are found to have a higher impact for the originated and destined flow for the SWTAZs within Florida than the SWTAZ at the external zones. From the model, it implies that the employed population is contributing directly or indirectly to produce these commodities. Also, the destination SWTAZs have more industries are likely to generate more economic growth that

eventually increasing the demand of all the commodities. The higher magnitude of employment estimates indicates the developed areas are more likely to consume more commodities than the less industrially developed regions. In final step, we generate GIS map for the originated and destined flows for the SWTAZ flows within the Florida states to visualize the effect of the disaggregation procedure. A color-coded spatial flow originated from and destined to both the counties and the SWTAZs provides detailed flow information.

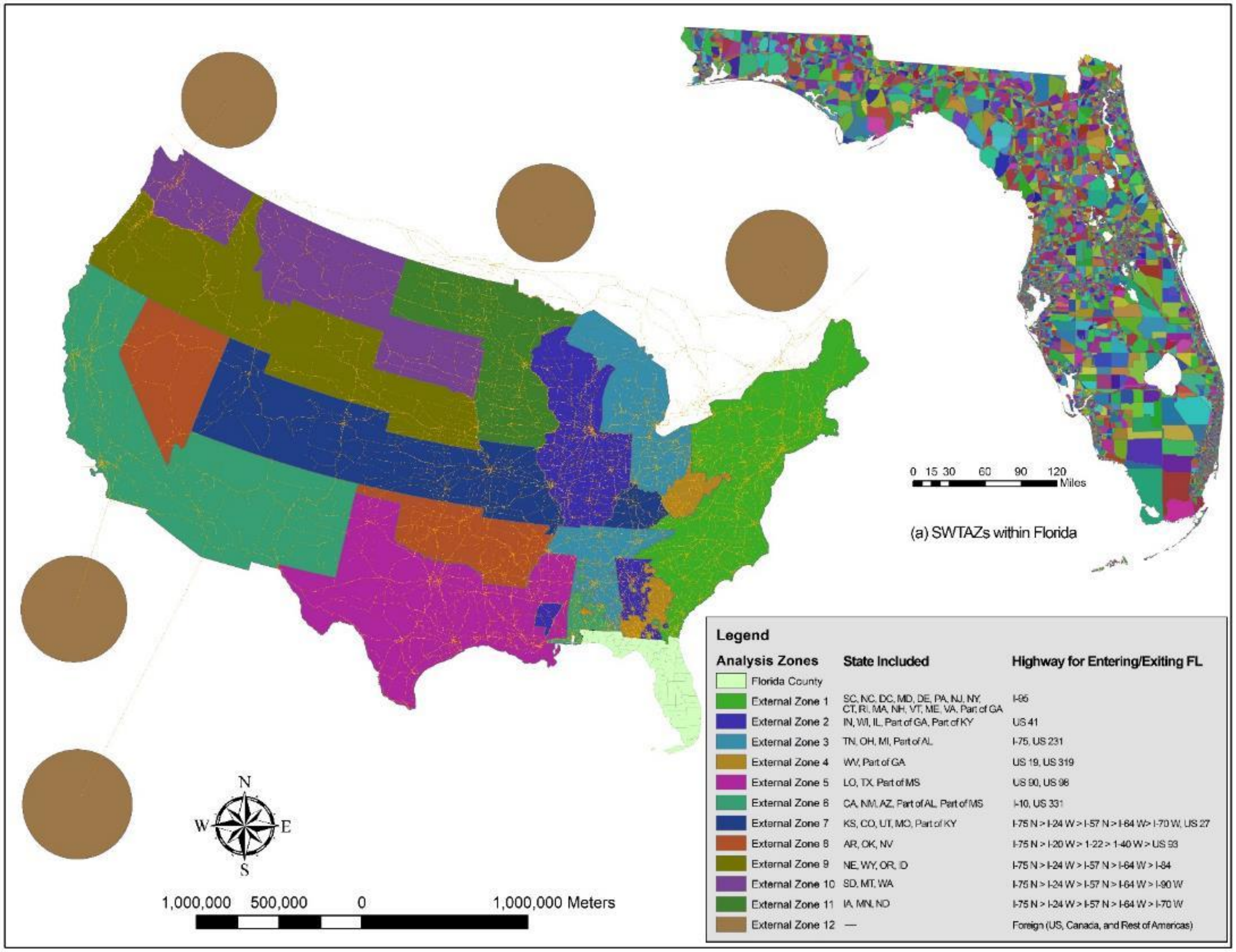


Figure 20 Spatial Zones Defined in the Study

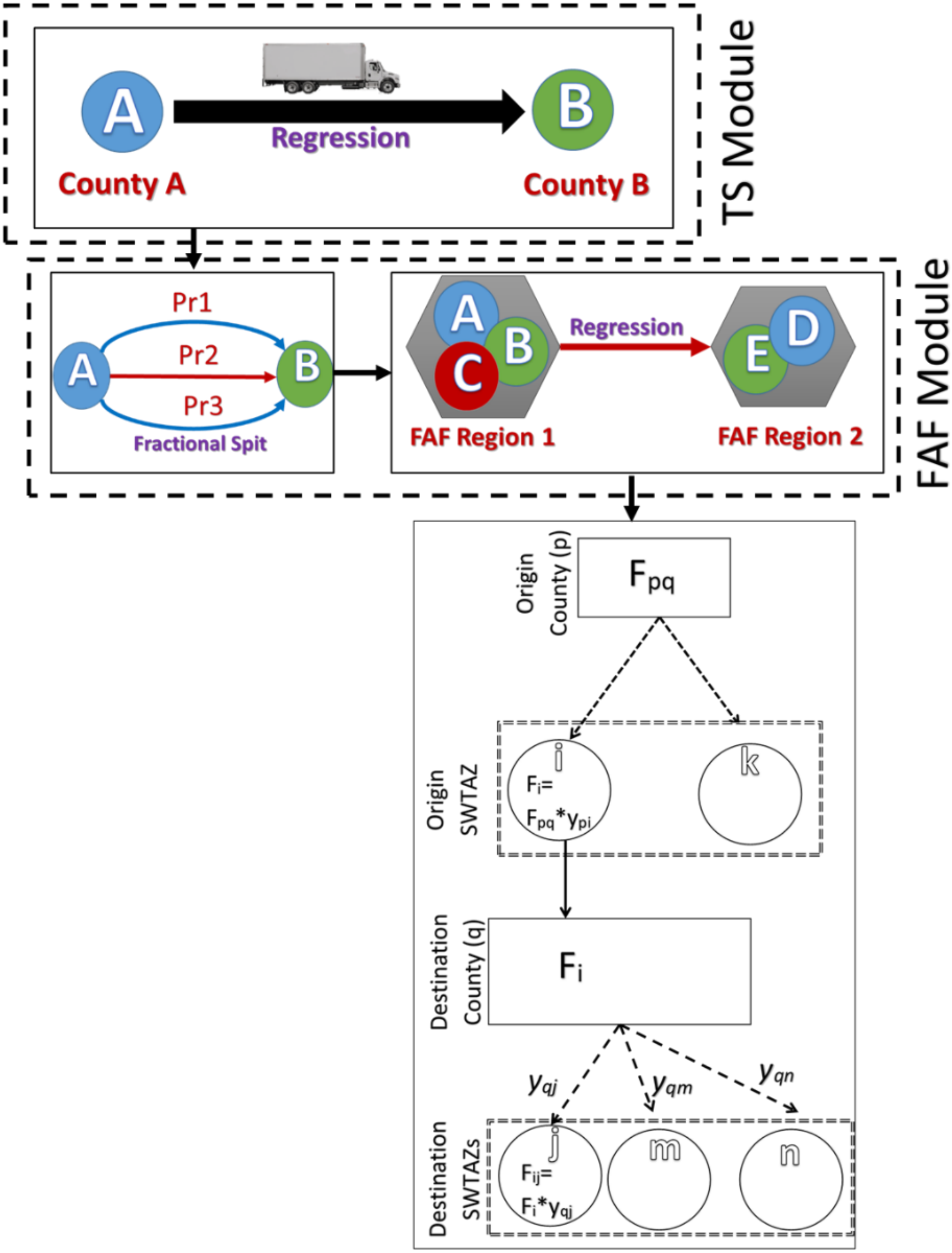


Figure 21 Data Fusion and Two-Stage Disaggregation of County Level Freight Flows

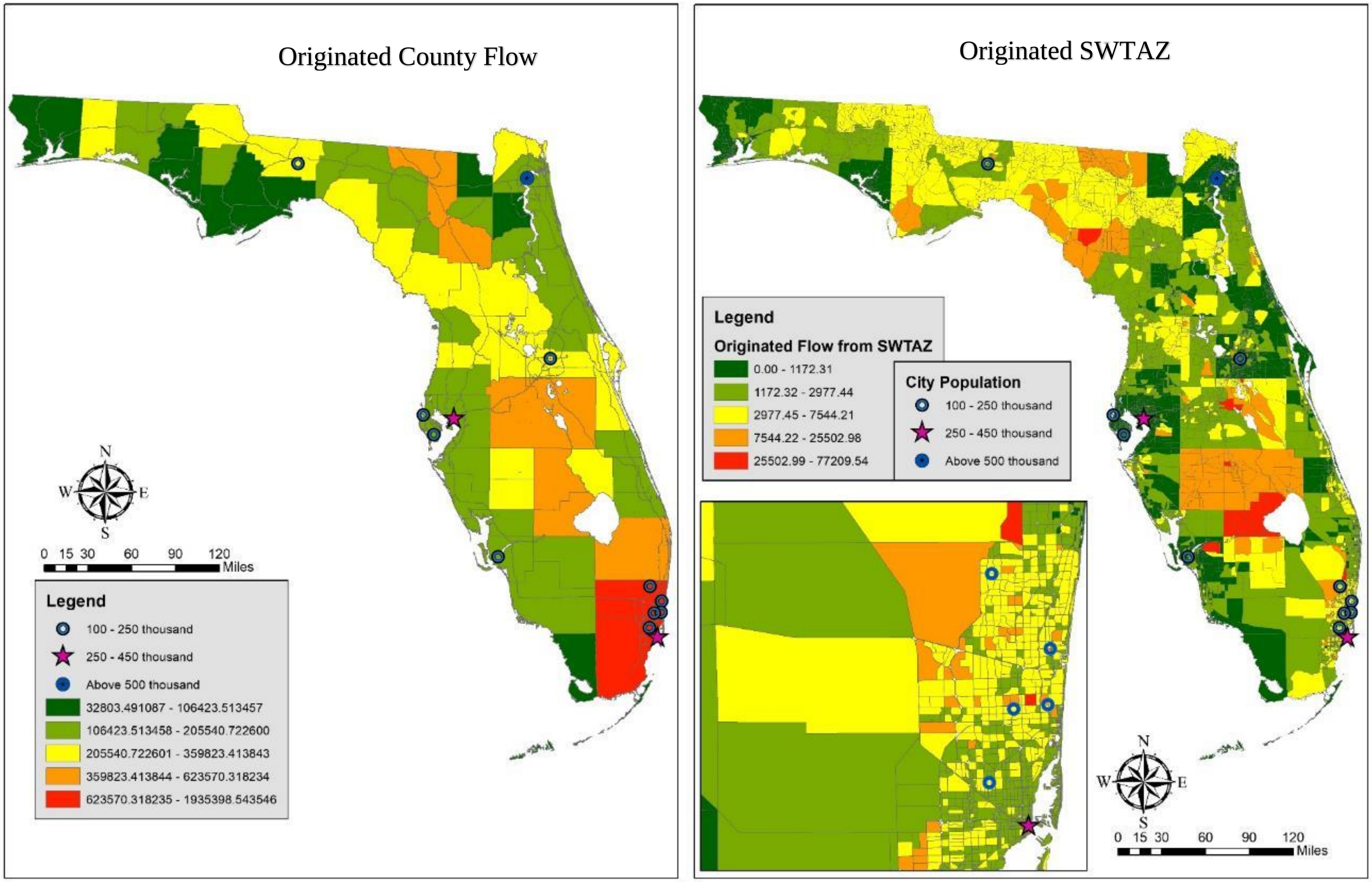


Figure 22 Disaggregation of Flow Originated from SWTAZs in Florida for Agricultural Product

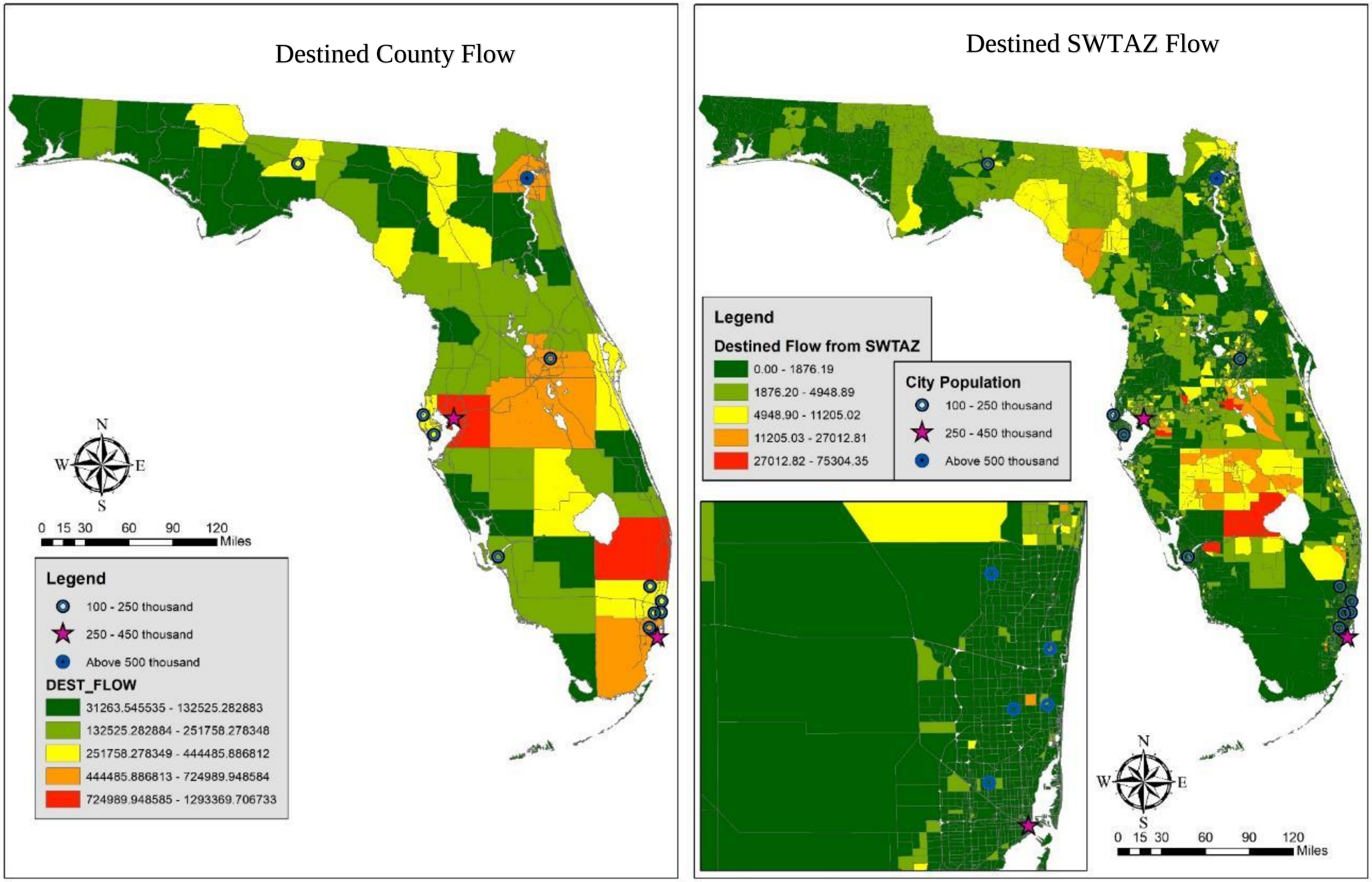


Figure 23 Disaggregation of Flow Destined to SWTAZs in Florida for Agricultural Product

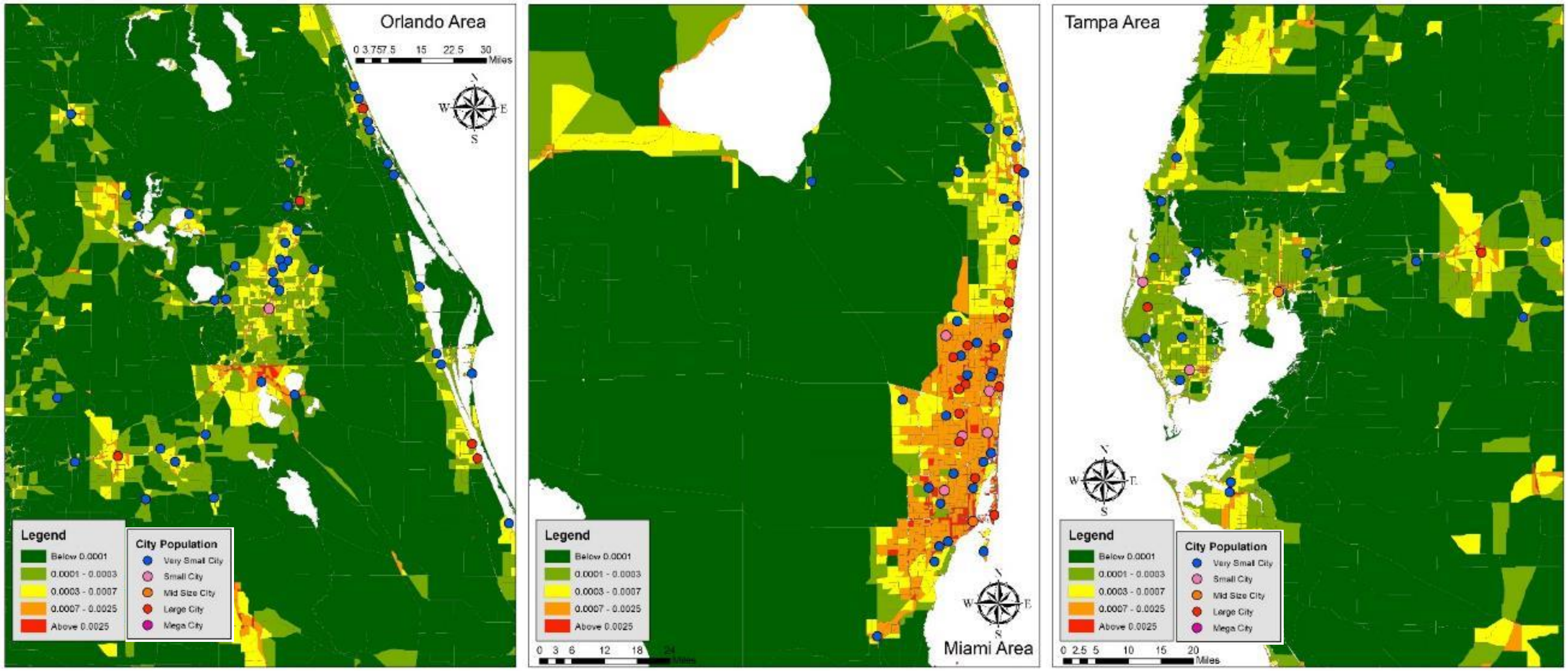


Figure 24 Originated Freight Flow Density from SWTAZs and the Cities of Different Population Size for Agricultural Product

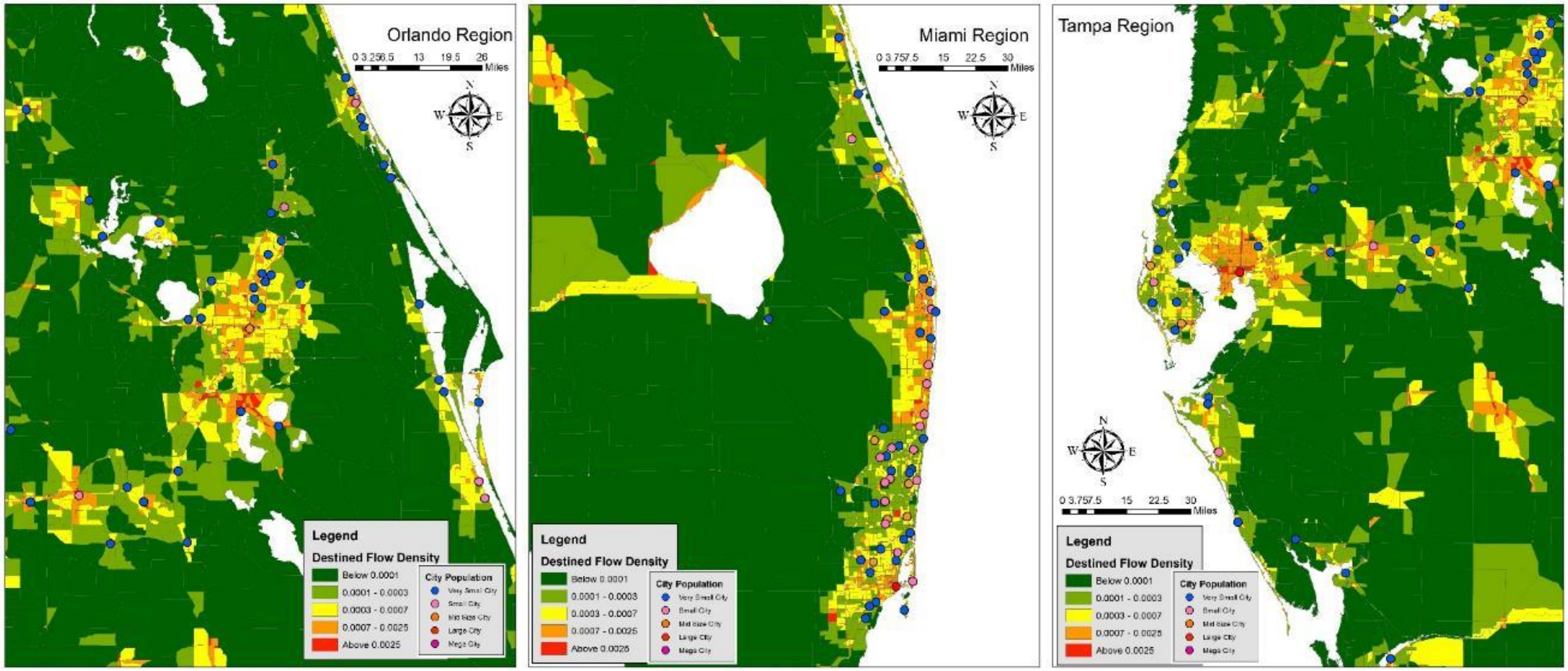


Figure 25 Destined Freight Flow Density from SWTAZs and the Cities of Different Population Size for Agricultural Product

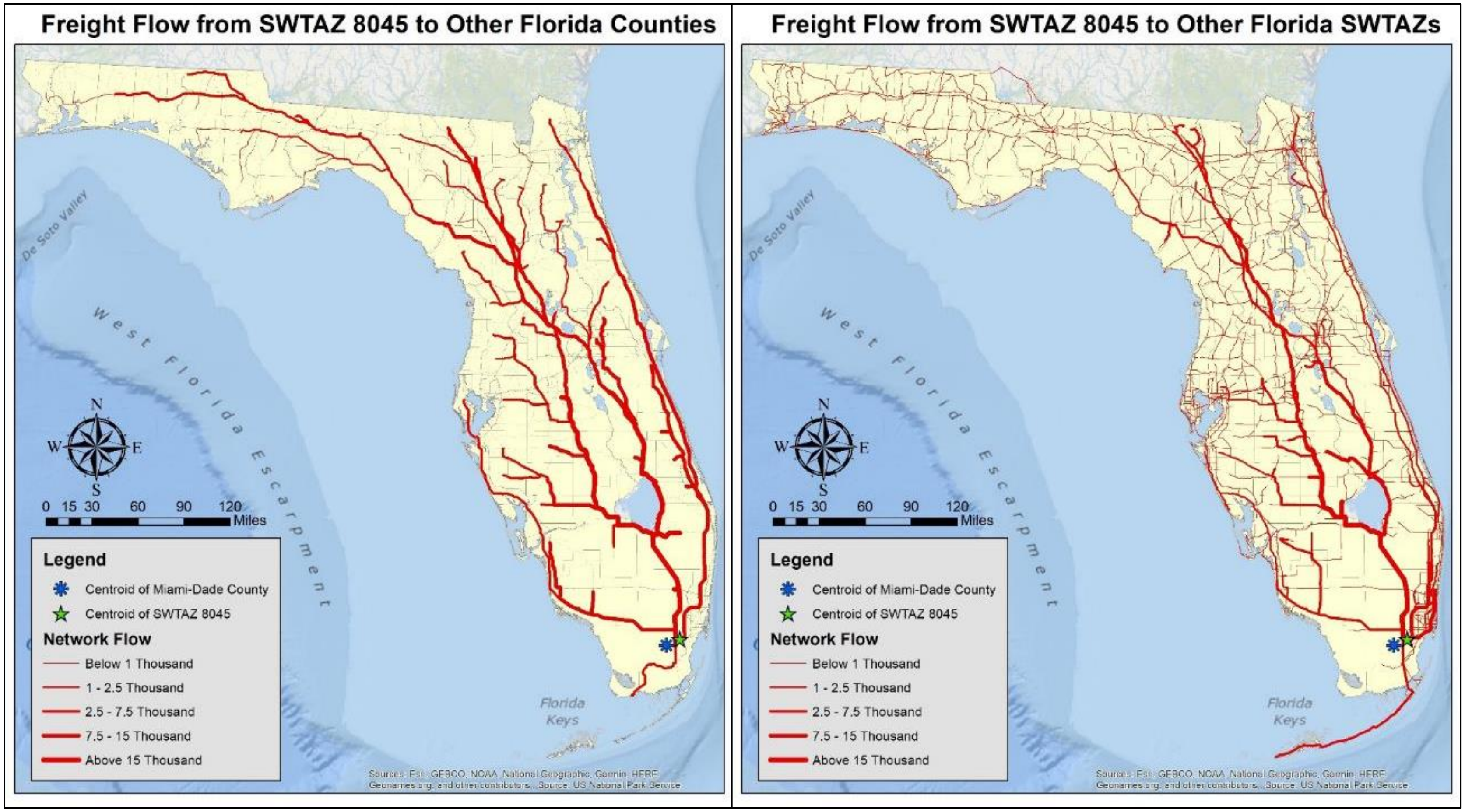


Figure 26 Originated Flow from SWTAZ 8045 to Counties and Other SWTAZs of Florida for Agricultural Product

CHAPTER SEVEN: SENSITIVITY OF FREIGHT FLOW ROUTE CHOICE

In Chapter Four, we developed an econometric algorithm for generating freight movement data at a finer spatial resolution. In our algorithm, the interconnection between FAF and TS is accomplished through the various paths and the associated route choice decision. The path probability of the FAF module is considered using a random utility based fractional split approach. The proposed algorithm is tested with only distance variable affecting path choice. However, it is possible that several other functional forms for distance and variables might influence path choice. Therefore, as part of this task, a sensitivity analysis on route choice is performed and tested to assess the improvement of the fused joint model and the predicted datasets. This chapter describes the sensitivity analysis on the route choice for the freight flow among the origins and destinations.

7.1 <u>Sensitivity of Number of Routes Available</u>

In our earlier exercise, we considered that all the possible paths are available for route choice between a specific origin-destination pair. Therefore, including the direct path and the one-hop (indirect) paths each origin-destination pair we considered 66 paths available alternative routes. However, while examining the proportions estimated for the freight flow on a path (route), we found that the shorter paths between a specific origin-destination likely to have higher proportion of the flow. It is also observed that the percentile value for the proportion are relatively high which led us to assume a hypothesis that there are a certain number of routes are feasible as the available alternative paths between an origin-destination pair. Aiming to test this hypothesis,

we conducted a sensitivity analysis for the number of alternative routes or paths. We created three different scenarios to test the sensitivity. Those are (i) 5 paths are available as alternatives, (ii) 10 paths are available as alternatives, and (iii) 20 paths are available as alternatives.

7.2 Sensitivity of Number of Stops Enroute

Number of intermediate stops enroute is one of the significant factors to determine the total trip distance. A long-haul trip for freight flow may possibly designed to stop one or more stops depending upon the logistics model of the shipper. If the intermediate stops are on the way to the destination, the stops are more likely to be incorporated in the route alternatives. The number of possible routes when all the counties are considered as the intermediate stops are shown in the last column of the Table 17. We can see that with the increase of stops, total possible route construction increases exponentially. Therefore, it would be beneficial to examine the different number of stops between the origin-destination. This will provide an idea about the effect and the threshold number of stops to be considered for the data fusion model. Besides the stops that are not on the route are not feasible. For example, let us consider a freight flow from Orlando (Orange County) to Miami (Miami-Dade County). The flow that are not on the way to Miami from Orlando such as Duval county in Jacksonville region and Leon County in Tallahassee region. In two-hop paths one of the possible routes would be Orange – Duval – Leon – Miami. It is obvious that the route would provide an unrealistic time and cost optimization for freight flow to be carried from Orlando – Miami direct freight flow. Again, these adds up the total number of possible route construction theoretically. It would be thus almost impossible to estimate the models with such a large number of alternatives while most of these will not be feasible. Moreover, the increase of intermediate

stops eventually increases the time to construct the link-path matrix and increase the size of the matrix significantly. For the dissertation, we limited the scope of the study to confine the sensitivity analysis up to three-hop paths. Hence, the scenarios we considered are (i) one-hop paths, (ii) two-hop paths, and (iii) three-hop paths.

7.3 Scenarios for Sensitivity Analysis

As stated above, we considered three different case scenarios each for investigating effect of the number of paths, and number of intermediate stops. As a result, we obtained total nine (i.e.3*3) scenarios for sensitivity analysis. The nine scenarios are presented in the following table. Please note that, the sensitivity of number of stops are investigated in conjunction to the number of paths available as alternatives. This provides a benefit in reducing the run-time of the models even if the total number of routes constructed increased exponentially with addition of just a single stop.

Table 16 Different Scenarios for Sensitivity Analysis

Scenario	Number of Stops Enroute	Number of Available Alternative Routes	Number of Total Possible Alternative Routes
S-0	1	66	66
S-1	1	5	66
S-2		10	66
S-3		20	66
S-4	2	5	4226
S-5		10	4226
S-6		20	4226
S-7	3	5	266306
S-8		10	266306
S-9		20	266306

7.4 Model Framework for Sensitivity Analysis

We tested two different aspects for the sensitivity analysis as stated above. This section describes the change in model framework to investigate the sensitivity of the freight flow route choice on the number of available paths and the number of intermediate stops enroute.

7.4.1 Reconstruction of Link-Path Matrix

Link-Path matrix is the key relationship matrix for aggregate the path flow into the link flow at the same spatial level. Since the matrix provides the information of existence of a link in a path using a dummy indicator, the matrix needed to be reconstructed for two-hop and three-hop paths. The new matrices provide the previous paths as well as the additional paths for two-hop and three-hop paths. Hence, the one-hop path matrix includes direct paths and indirect paths consisting only one-hop paths. Besides, the two-hop paths include direct paths and indirect paths consisting one-hop and two-hop paths. Similarly, three-hop paths matrix comprised of direct and indirect paths with one, two, and three, intermediate stops in a route. Finally, the matrices kept the top five, ten, or twenty paths only based on the shortest paths among the OD pairs. To be sure, the shortest path between two points are the direct paths between them. Therefore, in all the cases we obtained direct paths and 19 shortest indirect paths between an OD pair.

7.4.2 Modification in Log-Likelihood in the Model Structure

In our algorithm, we considered the normalization of the log-likelihood by the number of intermediate stops on the corresponding paths. Along with the change in the link-path matrix, the key change in the model framework is identifying the appropriate log-likelihood for the FAF

module of the model and normalize with the proper number of the intermediate stops. In previous chapter, the model used only direct paths and on-hop indirect paths. Therefore, the loglikelihood is normalized by a variable "n"; where n = 1 when the route is a direct path, and n = 2 when the route is an indirect one-hop path. In current sensitivity analysis the value of n become from 1 to 4. Mathematically n is given by as follows.

$$\boldsymbol{n} = \begin{cases} \mathbf{1};\ \boldsymbol{when\ direct\ path} \\ \mathbf{2};\ \boldsymbol{when\ one-hop\ path} \\ \mathbf{3};\ \boldsymbol{when\ two-hop\ path} \\ \mathbf{4};\ \boldsymbol{when\ three-hop\ path} \end{cases} \tag{18}$$

7.5 Model Outcome

We estimated nine different models for nine different scenarios for commodity paper. We extensively look into the commodity to explore the sensitivity of the different functional form of the models. The model estimates are shown in the Table 18. Please note that the variables in the earlier model are used where the model structure was built to consider all possible one-hop paths (e.g. 66 paths) for each origin-destination pair. It can be observed that, all the variable found significant in earlier model are found significant in all the nine models estimated for this specific goal. The estimates are found of same sign in the models except for external road length variable. This variable is associated with the external zones and shows different sign in the case of smaller number of alternative available for route choice. Hence, we can conclude that the effect is found following the similar trend while considering the more intermediate stops or while reducing the number of paths to a reasonable number of alternatives. However, the magnitude of the estimates

changes slightly from scenario to scenarios. It can be noted here that, with the reasonable number of alternatives (e.g. 10 or more) the model estimates stabilize for the data fusion.

For sensitivity analysis, we are more interested in the FAF module, since this module affects the route choice probabilities. From the table it is obvious that, for internal county as an origin or destination stop, the estimates change is not significant enough to distinguish.

We generated GIS maps to visualize the originated and destined flow for the Florida counties. In Figure 28 and 29, originated and destined flow for the commodity paper for nine different scenarios that we considered are presented. It is observed that for one-hop path, with the increase of alternative paths, the originated flows from the county are increasing for several counties.

Table 17 Model Estimates for Paper Commodity in Different Scenarios

Module	Parameters	S-0	S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8	S-9
		Estimates									
TS Module	Constant	2.392	2.230	2.367	2.405	2.316	2.351	2.381	2.316	2.368	2.339
	Population (destination)	2.797	4.043	3.664	3.251	3.887	3.889	3.576	3.919	3.841	3.921
	No of intermodal facilities (destination)	0.204	0.125	0.175	0.170	0.188	0.178	0.176	0.189	0.184	0.168
	No of ports (destination)	0.103	0.003	0.120	0.123	0.122	0.111	0.111	0.120	0.115	0.092
	Road length (external origin)	0.803	-0.211	0.189	0.343	-0.026	0.004	0.289	-0.031	0.057	-0.030
	No of intermodal facilities (origin)	0.410	0.507	0.499	0.470	0.510	0.505	0.484	0.508	0.490	0.494
	Road and Rail length (destination)	-0.422	-0.505	-0.454	-0.415	-0.447	-0.462	-0.409	-0.446	-0.452	-0.460
	Employment (destination) * No of warehouse (destination)	-4.039	-6.648	-5.915	-5.169	-6.277	-6.38	-5.870	-6.377	-6.329	-6.430
	Population (external destination) * No of ports (origin)	0.329	1.832	3.087	2.984	1.859	1.737	2.355	1.718	1.705	1.712
	Std. Err. (TS)	2.090	2.165	2.143	2.114	2.174	2.142	2.117	2.174	2.142	2.114
FAF Module	Path Distance (Internal Zones, KM)	-0.291	-0.159	-0.291	-0.291	-0.291	-0.290	-0.291	-0.291	-0.292	-0.289
	Path Distance (External Zones, KM)	-0.048	-0.012	-1.040	-0.152	-1.041	-0.995	-1.041	-0.986	-1.041	-0.994
	Std. Err. (FAF)	1.801	1.160	1.811	1.748	1.825	1.716	1.622	1.839	1.700	1.800

** S-0: 66-No. 1-hop paths; S-1: 5-No. 1-hop paths; S-2: 10-No. 1-hop paths; S-3: 20-No. 1-hop paths; S-4: 5-No. 2-hop paths; S-5: 10-No. 2-hop paths; S-6: 20-No. 2-hop paths; S-7: 5-No. 3-hop paths; S-8: 10-No. 3-hop paths; S-9: 20-No. 3-hop paths.*

7.5.1 Sensitivity of Number of Stops Enroute and Routes Available

While the model estimates were found reasonable for the scenarios, we further examine the probabilities. We looked at the probability changes in the different scenarios. Table 19 shows the Cumulative probability distribution on the direct path and indirect paths.

Table 18 Estimated Probabilities for on Routes Among the Origin-Destination Pairs

Number of Stop	Number of Paths	Mean Cumulative Probability (and mean S.D) of	
		Direct Path	Indirect Paths
1 Hop	5 Paths	0.7186 (0.180)	0.286 (0.188)
	10 Paths	0.789 (0.205)	0.207 (0.214)
	20 Paths	0.833 (0.218)	0.172 (0.228)
2 Hop	5 Paths	0.485 (0.288)	0.518 (0.29)
	10 Paths	0.516 (0.313)	0.487 (0.314)
	20 Paths	0.527 (0.322)	0.477 (0.324)
3 Hop	5 Paths	0.485 (0.288)	0.518 (0.29)
	10 Paths	0.517 (0.314)	0.486 (0.316)
	20 Paths	0.530 (0.326)	0.473 (0.327)

From the table, we found that, the flows are most likely to be carried on the direct paths. With the increase of alternatives, the probability of choosing direct paths are increasing. This implies that, the increase of alternatives increases more paths that offer more constrains (e.g. longer travel time) than the direct paths. Therefore, the probability of choosing the direct path become higher. As the intermediate stops increases from one to two, the probability of the cumulative probability of indirect paths increases. While considering the two-hop paths we are allowing more stops on the way to destination. Therefore, for an origin and destination with a longer distance we are allowing more two-hop paths that are potentially shorter in distance compared to other one-

hop paths with intermediate stops away from the route. As a result, we are more likely to have shorter distance alternative to compete with the direct path.

7.6 Summary

In our algorithm, the interconnection between FAF and TS is accomplished through the various paths and the associated route choice decision. The path probability of the FAF module is considered using a random utility based fractional split approach. It is possible that several other functional forms for distance and variables might influence path choice. Therefore, as part of this task, a sensitivity analysis on route choice is performed and tested to assess the improvement of the fused joint model and the predicted datasets. We conducted a sensitivity analysis for the number of alternative routes or paths and for the number of intermediate stops enroute. This will provide an idea about the effect and the threshold number of stops to be considered for the data fusion model. For the dissertation, we limited the scope of the study to confine the sensitivity analysis up to three-hop paths. The sensitivity of number of stops are investigated in conjunction to the number of paths available as alternatives. This provides a benefit in reducing the run-time of the models even if the total number of routes constructed increased exponentially with addition of just a single stop.

We updated link-path matrices and the log-likelihood of the model structure based on the scenarios to estimate nine different models for nine different scenarios for commodity paper. From the model, we observed that the effect of the attributes in models are following the similar trend while considering the more intermediate stops or while reducing the number of paths to a reasonable number of alternatives. However, the magnitude of the estimates changes slightly from

scenario to scenarios. It can be noted here that, with the reasonable number of alternatives (e.g. 10 or more) the model estimates are providing reasonable estimates for the data fusion.

CHAPTER EIGHT: CONCLUSION

Freight movement is a defining aspect of a region's economic viability and livability. Detailed data on freight movements would provide a greater understanding of freight patterns and its impacts on the transportation network. However, a major hurdle in freight demand modeling has always been the lack of adequate data on freight movements for different industry sectors for planning applications. The available movement data comes in many different forms, from many different sources (public or proprietary), with varying temporal and spatial resolutions, with substantial differences in the sampling and/or data collection methods. Therefore, instead of relying on a single source of data for modeling and other applications, an efficient approach would be to take build on the strengths of the multiple data sources. To achieve this, we can develop data fusion techniques to create a fused data that expands the scope of information and employ it for planning and forecasting purposes. The major goal of the dissertation is to address the above stated challenge of the absence of comprehensive data for freight transportation planning by developing an econometric data fusion algorithm using multiple datasets. The dissertation aims to develop a first of its kind fusion algorithm employing Freight Analysis Framework (FAF) and Transearch (TS) data to realize transportation network flows at a fine spatial resolution (county level and statewide traffic analysis zone) while accommodating for production and consumption behavioral trends (provided by TS). Towards this end, the dissertation formulates and estimates a joint econometric model framework grounded in maximum likelihood approach to estimate county level commodity flows. The algorithm is implemented for the commodity flow information from 2012 FAF data and 2011 TS databases. The data fusion process considers several exogenous variables including origin-destination indicator variables, socio-demographic and socio-economic

indicators, and transportation infrastructure indicators. In the first part of the dissertation, the proposed algorithm is initially employed to generate the fused freight database for the state of Florida at the county level resolution for the base year 2012 for FAF data and year 2011 for the TS databases. This task was successful to achieve the first objective of the dissertation.

With the ever-expanding global trade, imports and exports originating or destined to foreign countries is also likely to affect flows entering or leaving Florida. Therefore, we extend the algorithm proposed in the previous chapter to consider spatial regions within Florida and outside Florida (in the United States and in foreign regions). We resort to very detailed resolution inside Florida and larger aggregate zones to represent zones outside Florida. Furthermore, long-range planning of freight flows requires estimates of flows for future years. Specifically, we generated estimates for freight flows from 2015 through 2040 in five-year increments. The generated estimates are evaluated to understand how freight flows are likely to evolve in the Florida region. These tasks attain the specific objective II of the dissertation.

In the third part of the dissertation, we extend the algorithm to propose disaggregation methodology for the fused dataset. We disaggregated the county level link flows to the Statewide Traffic Analysis Zone (SWTAZ) as the finest spatial zones. For the SWTAZ level disaggregation with the same method would lead us to disaggregate the 17-zone system to an 8835-zone system. Florida Statewide system divided Florida into 8518 TAZs. Using a two-stage factor multiplication method is used to disaggregate the county level flow to SWTAZ flow. The factors are estimated both at origin and at destination level. The factors are estimated using a random utility factional split model approach. The proposed disaggregation is extended to the future year flows obtained

for years 2015 through 2040. This part of the dissertation accomplishes the objective III of the dissertation.

In the last part of the dissertation, , we consider travel distance for each path as an independent variable affecting route selection probability. We conducted sensitivity analysis of the parameterization by evaluating other alternative functional forms for the model structure and variables. In our algorithm, the interconnection between FAF and TS is accomplished through the various paths and the associated route choice decision. The proposed algorithm is tested with several other functional forms for distance and number of available alternative choices. The sensitivity analysis on route choice is performed and tested to assess the improvement of the fused joint model and the predicted datasets. This final task accomplished the final goal of the dissertation (objective IV).

8.1 Dissertation Research Impact

High resolution spatial freight data is not accessible due to financial and business confidentiality challenges. This dissertation developed an algorithm to generate high resolution spatial freight data by fusing multiple datasets. Specifically, we develop a data fusion exercise for Freight Analysis Framework (FAF) and Transearch data. The data fusion algorithm is used to generate freight flows at the county level and at the statewide Traffic analysis zone level for the base year and series of future years up to 2040. The data generated by commodity type and mode, as part of the dissertation, provides meaningful insights about the freight movement. Using data visualization, the dissertation also provides insights into the behavioral change of freight demand for different areas. The data visualization highlights varying freight flows trends across commodity

types for flows originating in Florida and/or flows destined to Florida. The data will assist us for the identifying future trends in freight transportation demand allowing us to incorporate freight demand and freight flows within infrastructure planning for county and urban regional planning. Finally, the proposed research also develops a framework to synthesize standalone Transearch like data using customized local data (in lieu of purchasing expensive data).

8.2 Limitation of the Research and Future Avenue for Research

The dissertation successfully achieved the objective of developing an algorithm to generate disaggregated freight data. However, the dissertation limits its scope to focus on the primary objective of the study. Due to unavailability of transearch data for the area outside of the state of Florida, the method is limited to only Florida. When the transearch data is available for the other state, the method can be tested for spatial transferability of models. Furthermore, the link flow data at the disaggregated level is not an observed value yet. Hence, we used institutional statistics and other aggregated methods including visualization technique for validating the freight flows at the disaggregated link level. With an observed value the model calibration is one of the potential directions of the future extension of the current research.

APPENDIX: JOINT MODELS BY COMMODITY

Table 19 Joint Models by Commodity

Variable	FCC1	FCC2	FCC3	FCC4	FCC5	FCC6	FCC7	FCC8	FCC9	FCC10	FCC11	FCC12	FCC15
	Agricultural Products	Minerals	Coal	Food	Nondurable Manufacturing	Lumber	Chemicals	Paper	Petroleum	Other Durable Manufacturing	Clay and Stone	Waste	Miscellaneous Freight & Ware House
Constant	3.043	-0.179	0.206	0.045	3.066	0.015	2.538	2.392	-0.056	0.027	0.047	0.019	0.044
Employment at Origin County	-	-	-	-0.056	-	-	-6.984	-	-	-	-	-	-
Road and Rail network length at Origin County	-	-	-	-0.130	-	-	-	-	-0.246	-	-	0.0035	-
Number of Intermodal Facilities at Origin County	-0.300	-0.318	-	-	-	-	0.298	0.410	-	-0.033	-	0.0250	-
Road network length at Origin Zone outside of Florida	-	-	-	-	-	-	-	0.803	-	0.035	-	-	-
Population at Origin Zone outside of Florida	-	-	-	-	1.414	0.058	-	-	-	-	-	-	0.228
Population at Origin County	1.732	1.024	-	0.210	-	-	15.403	-	-	-	-	-	-
Number of Ports at Origin County	-	0.583	-	-	-	-	-0.543	-	-	-	-	-	-
Number of Warehouse at Origin County	0.780	-	-	0.294	2.637	-	-	-	1.091	-	-	-	-
Employment at Origin County Square	-	-	-	-	-	-	-	-	-	-	0.377	-	-0.401
Population at Origin County square	-	-	-	-	-	-	-	-	-	-	-0.484	-	0.328
Employment at Origin County * Employment at Destination County	-	-	-	-	-	0.037	-	-	-	-	2.922	-	0.601
Employment at Origin County * Number of Warehouse at Destination County	-	-	-	-	-	-	-	-	-	-	-4.879	-	0.340

Variable	FCC1 Agricultural Products	FCC2 Minerals	FCC3 Coal	FCC4 Food	FCC5 Nondurable Manufacturing	FCC6 Lumber	FCC7 Chemicals	FCC8 Paper	FCC9 Petroleum	FCC10 Other Durable Manufacturing	FCC11 Clay and Stone	FCC12 Waste	FCC15 Miscellaneous Freight & Ware House
Population at Origin Zone outside of Florida * Number of Ports at Destination County	-	-	-	-	-	-	-	-	-	-	-	-	-2.875
Employment at Destination County	-	-	-	0.383	-	0.100	-	-	-	-	-	-	-
Employment at Destination County Square	-	-	-	-	-	-	-	-	-34.062	-	-2.462	-	-
Population at Destination Zone outside of Florida	-	-	-	0.128	-	0.041	1.545	-	-	-	-	-	-
Number of Intermodal Facilities at Destination County	-0.158	-	-	-	-	-	0.262	0.204	-	-	-	-	-
Population at Destination County square	-	-	-	-	-	-	-	-	3.466	-	-	-	-
Number of Ports at Destination County	0.294	-	-	-	-	-	-	0.103	-	-	-	-	-
Road and Rail network length at Destination County	-	-	-	-	-	-	-	-0.422	-	-	-	-	-
Employment at Destination County * Number of Warehouse at Destination County	-	-	-	-	-	-	-6.066	-4.039	-	-	0.549	-	-
Employment at Destination County * Number of Ports at Origin County	-	-	-	-	-	-0.050	-	-	-	-	-	-	-
Employment at Destination Zone outside of Florida * Number of Warehouse at Origin County	-	-	-	-	-	-	-	-	-	3.716	-	-	-0.656
Population at Destination Zone outside of Florida *	-	-	-	-	-	-	-	0.329	-	-	-	-	0.352

Variable	FCC1 Agricultural Products	FCC2 Minerals	FCC3 Coal	FCC4 Food	FCC5 Nondurable Manufacturing	FCC6 Lumber	FCC7 Chemicals	FCC8 Paper	FCC9 Petroleum	FCC10 Other Durable Manufacturing	FCC11 Clay and Stone	FCC12 Waste	FCC15 Miscellaneous Freight & Ware House
Number of Ports at Origin County													
Population at Destination County	0.962	1.880	-	-	-	-	3.231	2.797	-	-	-	0.136	-
Number of Warehouse at Destination County	0.899	-	1.348	-	1.849	-	-	-	-	-0.164	-	-	-
Destination External Zone	3.074	-	-	-	-	-	-	-	-	-	-	-0.017	-
Origin External Zone	1.973	-	-	-	1.290	-	-	-	-	-	-	-0.028	-
Destination Tampa	-	-	4.615	-	-	-	-	-	-	-	-	-	-
Std. Err. (TS)	1.761	4.217	10.479	0.483	2.416	0.129	2.165	2.090	3.069	0.334	0.404	0.271	0.751
Path Distance (External Zones, KM)	-0.287	-0.291	-0.099	-0.291	-0.048	-0.004	-0.220	-0.048	-0.052	-0.364	-0.087	-0.290	-0.183
Path Distance (Internal Zones, KM)	-0.031	-0.058	-0.350	-0.047	-0.292	-0.040	-0.293	-0.291	-0.286	-0.110	-0.014	-0.049	-0.003
Std. Err. (FAF)	3.652	38.751	12.407	9.419	1.739	4.356	1.708	1.801	27.616	2.988	7.180	13.169	5.080

REFERENCES

Aly, S., & Regan, A. (2014). Disaggregating FAF 2 Data for California. *Journal of Behavioural Economics, Finance, Entrepreneurship, Accounting and Transport, 2*, 47-57.

Aptech. (2015). *Home - Aptech* (Version 14.0). Retrieved from http://www.aptech.com/

Anderson, M., A. Youngblood, and H. Gregory. (2007) A Statewide Freight Flow Model for Alabama. Presented at 86th Annual Meeting Transportation Research Board, Washington, D.C.

Aultman-Hall, L., B. Johnson, and B. Aldridge. (2000) Assessing Potential for Modal Substitution from Statewide Freight Commodity Flow Data. Transportation Research Record: Journal of the Transportation Research Board. 1719: 10–16.

Bernardin, V. L., J. Avner, J. Short, L. Brown, R. Nunnally, and S. Smith. (2011) Using Large Sample GPS Data to Develop an Improved Truck Trip Table for The Indiana Statewide Model. Presented at 4th Innovations in Travel Modeling Conference, Tampa, FL.

Ben-Akiva, M., and G. de Jong (2013) The Aggregate-Disaggregate-Aggregate (ADA) Freight Model System. In Freight Transport Modelling (M. Ben-Akiva, H. Meersman, E. V. Voorde, ed.), Emerald Group Publishing Limited, Bingley, U.K., pp. 69-90.

Bhowmik, T., S. Yasmin, and N. Eluru. (2018) A Joint Econometric Approach for Modeling Crash Counts by Collision Type. Analytic Methods in Accident Research. 19: 16–32.

Bujanda, A., Villa, J., & Williams, J. (2014). Development of Statewide Freight Flows Assignment Using the Freight Analysis Framework (Faf3). *Journal of Behavioural Economics, Finance, Entrepreneurship, Accounting and Transport, 2*, 47-57.

de Jong, G., I. Vierth, L. Tavasszy, and M. Ben-Akiva. (2013). Recent Developments in National and International Freight Transport Models Within Europe. Transportation. 40(2): 347–371.

Eatough, C. J., S. C. Brich, and M. J. Demetsky. (2000) A Statewide Intermodal Freight Transportation Planning Methodology. Transportation Quarterly. 54(1): 145-155.

FHWA. Intermodal Surface Transportation Efficiency Act of 1991. (1991). US Department of Transportation, 1991.

Federal Highway Administration (FHWA) & Bureau of Transportation Statistics (BTS). (2012). *Freight Analysis Framework (FAF4).* Retrieved from http://faf.ornl.gov/fafweb/Default.aspx

Giaimo, G. T., and R. Schiffer. (2005). Transportation Research Circular Number E-C075: Statewide Travel Demand Modeling. Transportation Research Board. Washington D.C.

Giuliano, G., Gordon, P., Pan, Q., Park, J., & Wang, L. (2010). Estimating Freight Flows for Metropolitan Area Highway Networks Using Secondary Data Sources. *Networks and Spatial Economics, 10*, 73-91.

Harris, G. A., Anderson, M. D., Farrington, P. A., Schoening, N. C., Swain, J. J., & Sharma, N. S. (2010). Developing Freight Analysis Zones at a State Level: A Cluster Analysis Approach. *Journal of the Transportation Research Forum*, *49,* 59-68.

Holguín-Veras, J., List, G. F., Meyburg, A. H., Ozbay, K., Passwell, R. E., & Yahalom, S. (2001). An Assessment of Methodological Alternatives for A Regional Freight Model in The NYMTC Region an Assessment of Methodological Alternatives for a Regional Freight

Model in The NYMTC Region. New York: New York Metropolitan Transportation Council (NYMTC).

Huang, W.-J., and R. Smith. (1999). Using Commodity Flow Survey Data to Develop a Truck Travel-Demand Model for Wisconsin. Transportation Research Record: Journal of the Transportation Research Board. 1685: 1–6.

Jones, E. G., and A. Sharma. (2002) Development of Statewide Freight Trip Forecasting Model for Nebraska. Presented at 81st Annual Meeting, Transportation Research Board, Washington, D.C.

Jin, T. G., M. Saito, and G. G. Schultz. (2012) Development of a Statewide Commodity Flow Distribution Model Using Composite Friction Factors. Presented at 8th International Conference on Traffic and Transportation Studies, Changsha, China.

Krishnan, V., and K. Hancock. (1998). Highway Freight Flow Assignment in Massachusetts Using Geographic Information Systems. Transportation Research Record: Journal of the Transportation Research Board. 1625: 156–164.

Mendoza, A., C. Gil, and J. Trejo. (1999). Multiproduct Network Analysis of Freight Land Transport Between Mexico and the United States. Transportation Research Record: Journal of the Transportation Research Board. 1653: 69–78.

Mitra, S., & Tolliver, D. (2009). Framework for Modeling Statewide Freight Movement Using Publicly Available Data. *Journal of the Transportation Research Forum, 48*, 83-102.

National Institute of Statistics and Geography, Mexico (INEGI). (2017, December 05). *Population by Regions*. [Data file]. Retrieved from http://en.www.inegi.org.mx/default.html

Oliveira-Neto, F., Chin, S. M., Hwang, H., Peterson, B., & Moraes. (2012). *Methodology for estimating ton-miles of goods movements for U.S. freight multimodal network system*. Paper presented at the 91st TRB Annual Meeting, Washington DC.

Opie, K., Rowinski, J., & Spasovic, L. (2009). Commodity-Specific Disaggregation of 2002 Freight Analysis Framework Data to County Level in New Jersey. *Transportation Research Record: Journal of the Transportation Research Board, 2121*, 128-134.

Pendyala, R., V. Shankar, and R. McCullough. (2000). Freight Travel Demand Modeling: Synthesis of Approaches and Development of a Framework. Transportation Research Record: Journal of the Transportation Research Board. 1725: 9–16.

Ranaiefar, F., Chow, J. Y., McNally, M. G., & Ritchie, S. G. (2014). *A Structural Direct Demand Model for Inter-Regional Commodity Flow Forecasting*. Paper presented at the 93rd TRB Annual Meeting, Washington D.C.

Ranaiefar Joseph Chow, F. Y., & Rodriguez-Roman Pedro Camargo Stephen G Ritchie, D. V. (2013). *UCI-ITS-WP-12-4 Geographic Scalability and Supply Chain Elasticity of a Structural Commodity Generation Model Using Public Data*. Paper presented at the 92nd TRB Annual Meeting, Washinton D.C.

Rodriguez-Roman, D., Masoud, N., Jeong, K., & Ritchie, S. (2014). Goal Programming Approach to Allocate Freight Analysis Framework Mode Flow Data. *Transportation Research Record: Journal of the Transportation Research Board, 2411*, 82-89.

Ross, C., Kumar, A., Wang, F., & Hylton, P. (2016). 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). Atlanta, GA: Georgia Tech Research Corporation.

Rowinski, J., Opie, K., & Spasovic, L. N. (2008). *Development of Method to Disaggregate 2002 FAF2 Data Down to County Level for New Jersey*. Paper presented at the 87th TRB Annual Meeting, Washington DC.

Ruan, M., & Lin, J. (2010). Synthesis Framework for Generating County-Level Freight Data Using Public Sources for Spatial Autocorrelation Analysis. *Transportation Research Record: Journal of the Transportation Research Board, 2160*, 151-161.

Sivakumar, A., and C. Bhat. (2002). Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis. Transportation Research Record: Journal of the Transportation Research Board. 1790: 80–88.

Statistics Canada. (2017, December 05). *Population and demography*. [Data file]. Retrieved from http://www.statcan.gc.ca

SteadieSeifi, M., N. P. Dellaert, W. Nuijten, T. Van Woensel, and R. Raoufi. (2014) Multimodal Freight Transportation Planning: A Literature Review. European Journal of Operational Research, 2014. 233(1): 1–15.

The Florida Geographic Data Library (FGDL). (2017, July 15). *FGDL Search/ Download Data*. Retrieved from http://www.fgdl.org/metadataexplorer/explorer.jsp

U.S. Census Bureau. (2017a, July 15). *Population Clock*. [Data file]. Retrieved from http://www.census.gov/popclock/

U.S. Census Bureau. (2017b, July 15). *Employment County Data*. [Data file]. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Tavasszy, L. A., K. Ruijgrok, and I. Davydenko. (2002). Incorporating Logistics in Freight Transport Demand Models: State-of-the-Art and Research Opportunities. Transport Reviews. 32(2): 203–219.

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

Xiong, D., F. Zhao, L.-F. Chow, and S. Chung. (2007). Integrating Data and Models for Analysis of Freight Movements on Multimodal Transportation Systems for Florida. Florida Department of Transportation. http://www.fdot.gov/research/completed_proj/summary_pl/fdot_bd015_13_rpt.pdf. Accessed April 30, 2018.