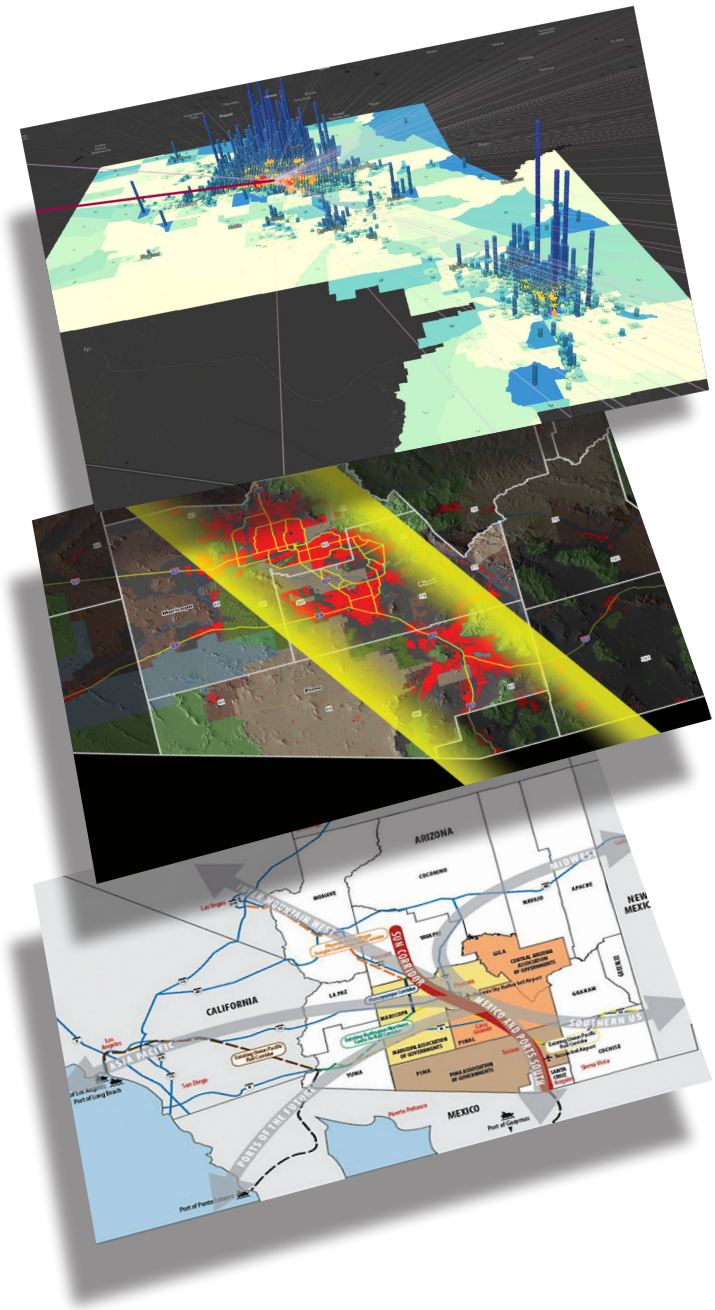
Strategic Highway Research Program 2 (SHRP2)
Freight Demand Modeling and Data Improvement (C20)
Implementation Assistance Program (IAP)



Mega-Regional Multi-Modal Agent-Based Behavioral Freight Model

MAG Next Generation Freight Demand Model

Final Report

Prepared with assistance from
Federal Highway Administration (FHWA)
SHRP2 C20 IAP Grant

U.S. Department of Transportation
Federal Highway Administration

April 5th, 2018 Update

MARICOPA ASSOCIATION of GOVERNMENTS

ADOT

PAG
Pima Association of Governments

Disclaimer

This report and/or data was funded in part through grant[s] from the Federal Highway Administration and/or Federal Transit Administration, U.S. Department of Transportation. The contents of this report reflect the views and opinions of the author(s) who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily state or reflect the official views or policies of the U.S. Department of Transportation, the Arizona Department of Transportation, or any other State or Federal Agency. This report does not constitute a standard, specification or regulation.

The information in this publication is provided on an "as is" basis, and there are no warranties, express or implied, including, but not limited to, any warranties of merchantability or fitness for a particular purpose. In no event shall MAG be liable for any damages resulting from the use of the information. MAG provides the information in good faith and has endeavored to create and maintain accurate data. The users of this report and/or data are advised to use the information with caution and to independently verify accuracy.

Best Data Fusion Application Award

From among a pool of more than 20 applications, MAG SHRP2 C20 Project was awarded "Best Data Fusion Application" in TRB Innovations in Freight Data Workshop hosted in Irvine, California, May 17-18, 2017. The workshop planning committee recognized this project "for incorporating traditional and underutilized data sources to model freight flows by industry, addressing a common and important data gap". Workshop materials are available at http://www.trb.org/conferences/freightdata2017.aspx. The presentation by Project Manager Vladimir Livshits during the Award Winners Panel on May 17, 2017 is accessible at http://onlinepubs.trb.org/onlinepubs/conferences/2017/FreightData/LivshitsPPT.pdf.

Project Contributors -

Maricopa Association of Governments

Arizona Department of Transportation

American Transportation Research Institute, Inc.

Cambridge Systematics, Inc.

CDM Smith, Inc.

Federal Highway Administration

IHS Global Insight

InfoGroup

Pima Association of Governments

RS&H, Inc.

StreetLight Data, Inc.

The Tioga Group, Inc.

Walls and Associates, Inc.

List of Project Contributors was updated on May 19, 2017

April 5th, 2018 Update

Acknowledgement

This work was sponsored by the Federal Highway Administration and Maricopa Association of Governments. It was conducted as part of the Round Three of Strategic Highway Research Program 2 (SHRP2) C20 Implementation and Technical Assistance Program (IAP) Grant. Authors of the report acknowledge guidance and direction throughout duration of the project from Mr. Vidya Mysore (FHWA). Mr. Birat Pandey provided valuable advice for selected tasks of the project.

The research reported was performed by Maricopa Association of Governments and a number of consulting teams listed in the report. Dr. Vladimir Livshits served as the Maricopa Association of Governments Project Manager. A list of contributors is provided on the next page and at the beginning of the corresponding chapters. The project team acknowledges contribution and guidance from the Project Technical Advisory Group (membership is listed below). The research team also acknowledges project management contributions from Mr. Nicholas Kehoe, Mr. Bob Sanchez (from Toxel and Leidos correspondingly). Toxel and Leidos served as FHWA consultants during the course of this project. The project team presented results of the work at the Transportation Research Board (TRB) 96th Annual Meeting in Washington D.C. in January, 2017. Special appreciation is extended to all contributors of the presentations and the papers.

Technical Advisory Group Members

- Mr. Romare Truly (FHWA)
- Mr. Keith Killough (ADOT)
- Dr. Aichong Sun (PAG)
- Mr. Tim Strow (MAG)
- Dr. Vladimir Livshits (MAG), Project Manager

Authors

Name, Organizational Affiliation during Project's Tenure

Agencies Staff:

Vladimir Livshits, Maricopa Association of Governments, Project Manager

Daehyun You, Maricopa Association of Governments

Haidong Zhu, Maricopa Association of Governments

Kyunghwi Jeon, Maricopa Association of Governments

Lavanya Vallabhaneni, Maricopa Association of Governments, Deputy Project Manager

Pedro Camargo, Formerly with MAG, Deputy Project Manager, currently with Veitch Lister Consulting

Shuyao Hong, Maricopa Association of Governments

Sreevatsa Nippani, Maricopa Association of Governments, Deputy Project Manager

Aichong Sun, Pima Association of Governments

Hyunsoo Noh, Pima Association of Governments

James Tokishi, Pima Association of Governments

Consultant Team:

Arun Kuppam, Cambridge Systematics

Brent Selby, Cambridge Systematics

Dan Beagan, Cambridge Systematics

Dan Murray, American Transportation Research Institute

Daniel S. Smith, The Tioga Group

Jason Lemp, Cambridge Systematics

Jeff Short, American Transportation Research Institute

Krishnan Viswanathan, CDM Smith, Cambridge Systematics

Mark Ehlen, Mark Ehlen Consulting

Monique Stinson, RS&H, Massachusetts Institute of Technology

Monisha Khurana, CDM Smith, AECOM

Paul Bingham, CDM Smith, EDRG

Rob Bostrom, CDM Smith

Seth Golub, Streetlight

Gary A. Carlin, INRIX

Srinath Ravulaparthy, Cambridge Systematics

Zahra Pourabdollahi, RS&H

Table of Contents

Executive Summary 1
1 Background 1-1
1.1 Motivation 1-1
1.2 Historical Outline of Main Relevant Studies in Arizona 1-2
1.3 Executive-level Support 1-3
1.4 Overall Goals for the Project 1-4
1.5 Main Methodological Approaches 1-5
1.6 Expected Deliverables 1-6
2 Model Design 2-1
2.1 Background 2-1
2.2 Supply Chain – Tour-based Freight Modeling Framework 2-3
2.3 Regional Behavior-Based Freight Modeling 2-6
2.3.1 Modeling Framework 2-9
3 Review of Freight Data Sources for the Development of a Behavior-Based Freight Model 3-1
3.1 Introduction 3-1
3.2 Data Types and Sources 3-5
3.2.1 Data for Forecasting 3-8
3.3 Firm Synthesis 3-8
3.4 Logistics Chain Models 3-18
3.5 Transportation Chain Data 3-26
3.6 Truck Touring Models 3-30
3.7 Other Data Sources 3-32
3.8 Summary 3-33
3.8.1 Geography Gaps 3-33
3.8.2 Model Needs Gaps 3-33
3.8.3 Mode Gaps 3-33
4 Analysis of Datasets used in the Model Development 4-1
4.1 National Establishment Time Series (NETS) Data 4-1
4.2 American Transportation Research Institute (ATRI) Truck GPS data 4-3
4.2.1 Overview 4-3
4.2.2 Data Collection and Coverage 4-3
4.2.3 Data Transformations and Processing 4-5
4.3 StreetLight Truck GPS data 4-10
4.3.1 Overview 4-10
4.3.2 Data Statistics and Coverage 4-10

4.3.3 Data Processing ... 4-12
4.4 IHS Global Insight's TRANSEARCH Data ... 4-15
4.4.1 Overview ... 4-15
4.4.2 Commodity Groupings ... 4-16
4.4.3 STB Confidential Carload Waybill Sample ... 4-17
4.4.4 Data Capture Exclusions ... 4-17
4.4.5 TRANSEARCH Data Acquisition by MAG ... 4-17
4.4.6 Data Analysis Summary from TRANSEARCH ... 4-18
4.5 Freight Analysis Framework (FAF 4.1) data ... 4-30
5 Firm Synthesis Model ... 5-1
5.1 NETS Data Description ... 5-1
5.1.1 Firm Events Analysis ... 5-2
5.2 Model Specification and Estimation Results ... 5-9
5.2.1 Firm Dissolution and Relocation ... 5-10
5.2.2 Firm Growth ... 5-10
5.2.3 Firm Location Choice ... 5-11
5.2.4 Model Estimation Results ... 5-11
5.3 Simulation Validation Results ... 5-19
5.3.1 Application on Validation Sample ... 5-19
6 Supply Chain Model ... 6-1
6.1 Model Framework ... 6-1
6.1.1 Geographical Coverage and Zone System ... 6-3
6.1.2 Decision Making Agents ... 6-5
6.1.3 Supply Chain Markets ... 6-7
6.2 Supplier Selection and Evaluation of Commodity Flows ... 6-8
6.2.1 Data Sources ... 6-9
6.2.2 Foundational Principles of the Model ... 6-10
6.2.3 Market Clearing Mechanisms ... 6-10
6.2.4 Modeling Approach ... 6-12
6.2.5 Model Design ... 6-14
6.2.6 Model Application and Preliminary Results ... 6-15
6.3 Transport, Mode, and Path Choice Model ... 6-20
6.3.1 Overview and Data Sources ... 6-20
6.3.2 CFS Microdata: Descriptive Statistics ... 6-23
6.3.3 Path Choice Model Estimation ... 6-42
6.3.4 Model Application and Preliminary Results ... 6-46
6.4 Calibration Process and Application Results ... 6-47

6.4.1 Overview and Data Sources ... 6-47
6.4.2 Calibration Process and Application Results ... 6-48
6.4.3 Model Procedures ... 6-52
6.5 Additional References: ... 6-59
7 Tour-Based Truck Model ... 7-1
7.1 ATRI Truck GPS Data ... 7-1
7.1.1 2011 ATRI Data ... 7-1
7.1.2 2014 ATRI Data ... 7-2
7.1.3 Comparison of 2014 ATRI Data versus 2011 ATRI Data ... 7-6
7.1.4 Analyses with 2014 ATRI Data ... 7-7
7.1.5 Conclusions ... 7-12
7.2 Tour-Based Methodology for Trucks ... 7-12
7.3 Light Trucks ... 7-13
7.3.1 Model Estimation ... 7-13
7.3.2 Tour Generation ... 7-14
7.3.3 Stop Frequency Model ... 7-14
7.3.4 Tour Completion Model ... 7-16
7.3.5 Stop Purpose Model ... 7-17
7.3.6 Stop Location Choice Model ... 7-19
7.3.7 Stop Time-of-Day Choice Model ... 7-21
7.3.8 Comparison to Survey ... 7-24
7.4 Medium Trucks ... 7-27
7.4.1 Model Estimation ... 7-27
7.4.2 Tour Generation ... 7-28
7.4.3 Stop Frequency Model ... 7-28
7.4.4 Tour Completion Model ... 7-29
7.4.5 Stop Purpose Model ... 7-30
7.4.6 Stop Location Choice Model ... 7-35
7.4.7 Stop Time-of-Day Choice Model ... 7-37
7.4.8 Comparison to Survey ... 7-39
7.5 Heavy Trucks ... 7-42
7.5.1 Model Estimation ... 7-42
7.5.2 Tour Generation ... 7-43
7.5.3 Stop Frequency Model ... 7-43
7.5.4 Tour Completion Model ... 7-45
7.5.5 Stop Purpose Model ... 7-46
7.5.6 Stop Location Choice Model ... 7-50

7.5.7 Stop Time-of-Day Choice Model 7-52
7.5.8 Comparison to Survey 7-54
8 Integration of Supply-Chain and Tour-Based Models 8-1
8.1 Integration Methodology 8-1
8.1.1 Supply Chain Model Output Records 8-3
8.1.2 Vehicle Choice 8-6
8.1.3 Conversion from Annual Tons to Daily trucks 8-7
8.1.4 Integration Outputs 8-7
8.2 Vehicle Type Choice Model 8-8
8.2.1 Model Estimation 8-8
8.2.2 Model Validation 8-9
8.3 ODME of Processed Truck GPS Data 8-10
8.3.1 Methodology 8-10
8.3.2 Results 8-11
9 Model Implementation and Deployment 9-1
9.1 Programming Language 9-1
9.2 Software and Hardware Requirements 9-1
9.2.1 Software Requirements 9-1
9.2.2 Hardware System Requirements 9-1
9.3 Directory Structure and Model Run 9-2
9.3.1 Working Directory and Global Variable Set Up 9-2
9.3.2 Model Run and Runtime 9-4
9.4 Model Input and Output Files 9-4
9.4.1 Firm Synthesis Model Inputs 9-5
9.4.2 Firm Synthesis Model Outputs 9-9
9.4.3 Supply Chain Model Inputs 9-10
9.4.4 Supply Chain Model Outputs 9-13
9.4.5 Truck Tour Model Inputs 9-18
9.4.6 Truck Tour Model Outputs 9-19
9.4.7 Program Directory 9-21
9.5 How to Set Up the Supplier Selection Model Workspace and Project Folder in Eclipse 9-25
9.6 Eclipse Settings for Building JAR File 9-28
10 Overall Freight Model Validation 10-1
10.1 Incorporation of Freight Models to MAG Travel Demand Model 10-1
10.2 MAG Validation 10-5
10.3 Incorporation of Freight Models to PAG Travel Demand Model 10-6
10.4 PAG Validation 10-7

11 Visualization 11-1
11.1 Freight Model Visualization 11-1
11.1.1 Firm Synthesis Model 11-2
11.1.2 Commodity Flow from Supply Chain Model 11-4
11.1.3 Commodity Flow from TRANSEARCH 11-4
11.1.4 Truck Tour Model 11-5
11.1.5 Workflow and Tools 11-8
11.2 GPS Data Visualizations 11-9
11.2.1 Heat Map 11-9
11.2.2 Vehicle Trajectory Animation 11-10
11.2.3 Summary 11-11
Appendix A. CFS Microdata Megaregions
Appendix B. Nationwide Network Development

List of Tables

Table 3-1 Different Data Types for Freight Modeling....3-5
Table 3-2 Data Sources by Data Types....3-6
Table 3-3 Data Sources by Modeling Needs....3-7
Table 3-4 BEA Input/Output Table Metadata....3-9
Table 3-5 County Business Patterns Metadata....3-10
Table 3-6 National Establishment Time-Series (NETS) Metadata....3-11
Table 3-7 Longitudinal Business Dynamics (LBD) Metadata....3-12
Table 3-8 Annual Survey of Manufacturers (ASM) Metadata....3-13
Table 3-9 Business Dynamics Statistics Metadata....3-14
Table 3-10 Business Employment Dynamics Metadata....3-15
Table 3-11 Statistics of US Businesses Metadata....3-16
Table 3-12 Non Employer Statistics Metadata....3-17
Table 3-13 Commodity Flow Survey (CFS) Metadata....3-19
Table 3-14 Freight Analysis Framework (FAF) Metadata....3-20
Table 3-15 TRANSEARCH Metadata....3-21
Table 3-16 Surface Transportation Board (STB) Carload Waybill Sample Metadata....3-22
Table 3-17 Air Carrier Statistics Metadata....3-23
Table 3-18 North American Trans-border Freight Database Metadata....3-24
Table 3-19 Port Import/Export Reporting Service (PIERS) Metadata....3-25
Table 3-20 National Highway Planning Network (NHPN) Metadata....3-26
Table 3-21 ORNL Railroad Network Metadata....3-27
Table 3-22 Vehicle Inventory and Use Survey (VIUS) Metadata....3-28
Table 3-23 Vehicle Travel Information System (VTRIS) Metadata....3-29
Table 3-24 ATRI Truck GPS Data Metadata....3-30
Table 3-25 National Performance Management Research Dataset (NPMRDS) Metadata....3-31
Table 4-1 StreetLight Data Dictionary....4-10
Table 4-2 StreetLight Data Statistics....4-11
Table 4-3 StreetLight Truck Tour Statistics....4-15
Table 4-4 Summary of Tour Purposes and Stops per Tour based on StreetLight Data....4-15
Table 4-5 Maricopa County Commodity Flow....4-20
Table 4-6 Pinal County Commodity Flow Summary....4-20
Table 4-7 Pima County Commodity Flow Summary....4-21
Table 4-8 2013 County-to-County Annual Commodity Flow in Tons....4-21
Table 4-9 2013 County-to-County Annual Commodity Flow in Dollars....4-21
Table 4-10 2045 County-to-County Annual Commodity Flow in Tons....4-22
Table 4-11 2045 County-to-County Annual Commodity Flow in Dollars indexed to 2013....4-22
Table 4-12 Internal-Internal Commodity Flow Growth by Rail for Maricopa County....4-22
Table 4-13 Internal-External Commodity Flow Growth by Rail for Maricopa County....4-23
Table 4-14 External-Internal Commodity Flow Growth by Rail for Maricopa County....4-23
Table 4-15 2013 External-External Truck Flow for Mega-Region....4-24
Table 4-16 2013 Arizona – NAFTA Annual Commodity Flow Summary....4-25
Table 4-17 2013 MAG/PAG Mega-Region and NAFTA Annual Commodity Flow Summary....4-25
Table 4-18 2013 Mega Region Inter-County Flows by Truck by SCTG2 (in tons)....4-26

Table 4-19 2013 Mega Region Inter-County Flows by Truck by SCTG2 (in dollars) 4-27
Table 4-20 2013 Mega Region Inter-County Flows by Rail by SCTG2 (in tons) 4-28
Table 4-21 2013 Mega Region Inter-County Flows by Rail by SCTG2 (in dollars) 4-29
Table 4-22 FAF Modes 4-32
Table 4-23 2015 Mega-Region Commodity Flow Summary by SCTG2 (in tons) using FAF4 Data 4-33
Table 4-24 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in millions of dollars) using FAF4 Data 4-34
Table 5-1 Evolutionary Transition of Firm Population in MAG and PAG from 1990 to 2012 5-3
Table 5-2 Distribution of Firm Births by Industry Types from 1991 to 2012 5-5
Table 5-3 Distribution of Firm Deaths by Industry Types from 1991 to 2012 5-6
Table 5-4 Distribution of Firm Relocations by Industry Types from 1991 to 2011 5-7
Table 5-5 Firm Growth Model Estimation Results (Calibrated) 5-12
Table 5-6 Firm Dissolution Model Estimation Results (Calibrated) 5-14
Table 5-7 Firm Relocation Model Estimation Results (Calibrated) 5-15
Table 5-8 Firm Move-In Location Choice Model Estimation Results (Calibrated) 5-16
Table 5-9 Firm Move-Within Location Choice Model Estimation Results (Calibrated) 5-17
Table 5-10 Firm Birth Rate Estimates by Industry Type 5-18
Table 5-11 Summary of 20 Percent Sample Validation Results 5-19
Table 6-1 Zone System and Zone Numbers in the Model 6-3
Table 6-2 MAG/PAG Total Number of Business Establishments Enumerated in the Model 6-7
Table 6-3 SCTG Classes Considered in the Model 6-8
Table 6-4 Average Importance Weights of Supplier Evaluation Criteria 6-14
Table 6-5 Transportation costs used in Path Choice model development 6-21
Table 6-6 Summary of Cost Assumptions 6-21
Table 6-7 Summary of Travel Time Assumptions 6-22
Table 6-8 Arizona-Based Shipment Records in the CFS 2012 Microdata Sample 6-23
Table 6-9 Shipment Weights and Values by Mode: Unweighted (Sample) and Weighted (Expanded) 6-26
Table 6-10 Origin and Destination Regions of Shipments in the Arizona-Based Shipment Dataset 6-27
Table 6-11 Sample Size: Arizona-Based Shipments by Mode and Size Category 6-30
Table 6-12 Sampled Export Shipments by Alternative and Destination 6-30
Table 6-13 Sampled Export Shipments by Alternative and Commodity Category 6-31
Table 6-14 Value per Pound Categories Used in Model Estimation 6-31
Table 6-15 Sample Shipments: Average Value per Pound by SCTG Category 6-33
Table 6-16 Commodity Categories by SCTG Code 6-35
Table 6-17 Number of Shipments by SCTG and Mode/Shipment Size Category (Sample) 6-36
Table 6-18 Number of Shipments by SCTG and Mode/Shipment Size Category (Sample) 6-37
Table 6-19 Number of Shipments by SCTG and Mode/Shipment Size Category (Sample) 6-38
Table 6-20 Value per Pound and Great Circle Distance 6-41
Table 6-21 Mode and Shipment Size Categories Used in Path Choice model 6-42
Table 6-22 Path Choice model: summary statistics 6-43
Table 6-23 Path Choice Model Parameter Estimates 6-45
Table 6-24 Values-of-Time for Freight Shipments in a New York-Based Stated Preference Survey 6-46
Table 6-25 MAG/PAG Modal Split by Kiloton in FAF4.1 6-48
Table 6-26 Modeled vs. Observed Total Directional Flows 6-50
Table 6-27 Modeled vs. Observed Inbound Flows to MAG/PAG Region 6-51
Table 6-28 Modeled vs. Observed Outbound Flows from MAG/PAG Region 6-51

Table 6-29 Modeled vs. Observed Modal Split 6-52
Table 6-30 Commodity Classification Based on Unit Values....6-56
Table 6-31 Commodity Classification Based on Type of Commodity 6-56
Table 6-32 Transportation Costs Calculation Assumptions by Mode 6-57
Table 6-33 Transportation Time Assumptions by Mode....6-57
Table 7-1 2011 ATRI Truck GPS Data 7-1
Table 7-2 Summary of Light Truck Trips....7-14
Table 7-3 Stop Frequency Model – Light Trucks 7-15
Table 7-4 Tour Completion Model – Light Trucks 7-16
Table 7-5 Stop Purpose Model – Light Trucks 7-18
Table 7-6 Stop Location Choice Model – Light Trucks....7-20
Table 7-7 Time-of-Day Choice Model – First Trip – Light Trucks 7-22
Table 7-8 Time-of-Day Choice Model – Subsequent Trips – Light Trucks 7-24
Table 7-9 Summary of Medium Truck Trips 7-27
Table 7-10 Stop Frequency Model - Medium Trucks....7-29
Table 7-11 Tour Completion Model – Medium Trucks....7-30
Table 7-12 Stop Purpose Model - Medium Trucks....7-32
Table 7-13 Stop Location Choice Model - Medium Trucks 7-36
Table 7-14 Time-of-Day Choice Model - First Trip - Medium Trucks 7-38
Table 7-15 Time-of-Day Choice Model - Subsequent Trips - Medium Trucks 7-39
Table 7-16 Summary of Medium Truck Trips 7-42
Table 7-17 Stop Frequency Model - Heavy Trucks 7-44
Table 7-18 Tour Completion Model - Heavy Trucks....7-45
Table 7-19 Stop Purpose Model - Heavy Trucks 7-46
Table 7-20 Stop Location Choice Model - Heavy Trucks 7-50
Table 7-21 Time-of-Day Choice Model – First Trip – Heavy Trucks....7-53
Table 7-22 Time-of-Day Choice Model – Subsequent Trips – Heavy Trucks 7-54
Table 8-1 Tour Type and Tour Purposes in Supply Chain and Truck Tour Models 8-3
Table 8-2 NAICS2 Industry Codes 8-4
Table 8-3 Tons per Truck by SCTG and Truck Size....8-5
Table 8-4 Integration Model Outputs 8-8
Table 8-5 Model Estimation Results 8-9
Table 9-1 Programming Language for Model Implementation 9-1
Table 9-2 Sub Directories in the Main Working Directory 9-3
Table 9-3 Model Runtime 9-4
Table 9-4 Input Files of the Firm Synthesis Model 9-6
Table 9-5 Firm Seed Data Structure....9-7
Table 9-6 Base Year Socioeconomic Data Structure 9-8
Table 9-7 Firm Synthesis Outputs....9-10
Table 9-8 Zone system 9-10
Table 9-9 Supply Chain Model Inputs 9-12
Table 9-10 CommodityFlows_firmlevel.csv....9-14
Table 9-11 firm_all.csv attributes....9-15
Table 9-12 io_make_2007.csv attributes....9-15
Table 9-13 io_use_2007.csv attributes 9-15
Table 9-14 BuyerFirms.csv data table attributes....9-15

Table 9-15 SupplierFirms data table attributes 9-16
Table 9-16 zonesupply.csv data table attributes 9-16
Table 9-17 zonedemand.csv data table attributes 9-16
Table 9-18 industrysupply.csv data table attributes 9-16
Table 9-19 industrydemand.csv data table attributes 9-17
Table 9-20 totalsupply.csv data table attributes 9-17
Table 9-21 totaldemand.csv data table attributes 9-17
Table 9-22 annualtonnage_firmlevel.csv data table attributes 9-17
Table 9-23 Truck Tour Model Inputs 9-19
Table 9-24 Truck Tour Model Outputs 9-21
Table 10-1 Screenline validation summary 10-5
Table 10-2 Truck volumes vs. counts 10-6
Table 10-3 All volumes vs. counts 10-6
Table 10-4 Volume and Count Statistical Analysis of Freeway and Arterial 10-7
Table 10-5 Statistical Analysis by Area Type 10-9

List of Figures

Figure 2-1 Total Truck Tonnage by Direction 2-2
Figure 2-2 Supply Chain - Tour-Based Freight Modeling Framework 2-5
Figure 2-3 CMAP Freight Model Components 2-8
Figure 2-4 Individual Components of the Proposed Behavior-Based Freight Model 2-11
Figure 2-5 Example of Study Area 2-13
Figure 2-6 Firm Synthesis 2-14
Figure 3-1 Supply Chain Network 3-2
Figure 3-2 Hybrid Model Layout 3-3
Figure 3-3 Interaction between Various Networks 3-4
Figure 4-1 Mega-regional Modeling Area (green) and ATRI Data Study Area (green + blue) 4-4
Figure 4-2 Temporal Coverage of ATRI Data 4-4
Figure 4-3 Heat Map of GPS Pings in ATRI Data 4-5
Figure 4-4 Land Use Shapefile in the Modeling Area 4-7
Figure 4-5 Distribution of Truck Trip Lengths (ATRI Data) 4-8
Figure 4-6 Distribution of Truck Stop Duration (ATRI Data) 4-8
Figure 4-7 Distribution of Stops per Truck based on ATRI Data 4-9
Figure 4-8 Stops per Truck by Tour Purpose based on ATRI Data 4-9
Figure 4-9 Heat Map of GPS Pings in StreetLight Data 4-11
Figure 4-10 Time of Day Profile: Truck GPS Data vs Traffic Counts 4-12
Figure 4-11 Temporal Coverage of StreetLight Data 4-12
Figure 4-12 Distribution of Truck Stop Duration (StreetLight Data) 4-13
Figure 4-13 Distribution of Number of Stops for Medium Truck (per tour) (StreetLight Data) 4-14
Figure 4-14 Distribution of Number of Stops for Light Trucks (per tour) (StreetLight Data) 4-14
Figure 4-15 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in kilo tons) 4-35
Figure 4-16 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in Millions of dollars) 4-36
Figure 5-1 Distribution of Birth and Death Rates of Business Establishments from 1990 to 2012 5-4
Figure 5-2 Composition of Births by Industry Types from 1991 to 2012 5-8
Figure 5-3 Composition of Deaths by Industry Types from 1991 to 2012 5-8
Figure 5-4 Composition of Relocations by Industry Types from 1991 to 2011 5-9
Figure 5-5 Overview of Firm Evolution 5-9
Figure 5-6 Average of Yearly Change in Number of Firms 5-20
Figure 5-7 Average of Yearly Change in Employment 5-20
Figure 5-8 Average of Rate of Death of Establishments (Per Year) 5-21
Figure 5-9 Average of Rate of Birth of Establishments (Per Year) 5-21
Figure 5-10 Average of Rate of Choice to Move (Per Year) 5-22
Figure 5-11 Aggregate Level Validation - InfoGroup - Firm Counts 5-23
Figure 5-12 Aggregate Level Validation - InfoGroup - Employment Size 5-23
Figure 5-13 Aggregate Level Validation - Maricopa County Employer Database - Firm Counts 5-24
Figure 5-14 Aggregate Level Validation - Maricopa County Employer Database - Employment Size 5-24
Figure 5-15 Aggregate Level Validation - NETS Database - Firm Counts (Maricopa County) 5-25
Figure 5-16 Aggregate Level Validation - NETS Database - Employment Size (Maricopa County) 5-25
Figure 5-17 Aggregate Level Validation - NETS Database - Firm Counts (Pinal County) 5-26
Figure 5-18 Aggregate Level Validation - NETS Database - Employment Size (Pinal County) 5-26

Figure 5-19 Aggregate Level Validation - NETS Database - Firm Counts (Pima County) 5-27
Figure 5-20 Aggregate Level Validation - NETS Database - Employment Size (Pima County) 5-27
Figure 6-1 MAG Behavior-Based Supply Chain and Freight Transportation Framework 6-2
Figure 6-2 Sun Corridor Megaregion 6-4
Figure 6-3 Model Zone System 6-5
Figure 6-4 Firm Level Production Estimation Procedure 6-17
Figure 6-5 Firm Level Consumption Estimation Procedure 6-18
Figure 6-6 Number of Shipments by Size Categories 6-28
Figure 6-7 Percentages of Shipments by Mode Across Chosen Size Categories 6-29
Figure 6-8 Costs per Ton-Mile by Shipment Size 6-29
Figure 6-9 Mode and Shipment Size Share for Bulk Goods 6-39
Figure 6-10 Mode and Shipment Size Share for Fuel and Fertilizer 6-39
Figure 6-11 Mode and Shipment Size Share for Intermediate Products, Foodstuffs and Misc. 6-40
Figure 6-12 Mode and Shipment Size Share for Manufactured Goods 6-41
Figure 6-13 Nesting Structure of Path Choice Model 6-43
Figure 6-14 Modal Split in Preliminary Results: Modeled VS. FAF3.5 6-47
Figure 6-15 MAG/PAG Commodity Flow by SCTG Type Based on FAF4.1 6-48
Figure 6-16 Modeled vs. Observed Inbound Flows to MAG Region for SCTG 1 Before Calibration 6-49
Figure 6-17 Modeled vs. Observed Inbound Flows by SCTG After Calibration 6-50
Figure 7-1 ATRI GPS Truck Events in April 2011 7-2
Figure 7-2 ATRI GPS Truck Events in 2014 7-3
Figure 7-3 Distribution of Stop Purposes by Tour Purpose 7-4
Figure 7-4 Trip Length Frequency Distributions - 2014 versus 2011 ATRI Data 7-6
Figure 7-5 Stops per Truck - 2014 versus 2011 ATRI Data 7-6
Figure 7-6 Distribution of Tours by Stop Duration (2014 ATRI Data) 7-7
Figure 7-7 Distribution of Tours by Stops per Truck (2014 ATRI Data) 7-7
Figure 7-8 Stops per Truck by Tour Purpose (2014 ATRI Data) 7-8
Figure 7-9 Logistic Supply Chains 7-9
Figure 7-10 LU to LU – MAG and PAG 7-10
Figure 7-11 Distribution of Truck Tours by Purpose (x-axis) and by Average Travel Time (y-axis) 7-11
Figure 7-12 Distribution of Departure Time-of-Day (y-axis) by Purpose (x-axis) 7-12
Figure 7-13 Tour-Based Truck Model 7-13
Figure 7-14 Stop Frequency Comparison - Tour Model Outputs versus Survey Data 7-25
Figure 7-15 Stop Purpose Frequency - Model Outputs versus Survey Data 7-26
Figure 7-16 Time-of-Day of First Stop - Model Outputs versus Survey Data 7-26
Figure 7-17 Time-of-Day of Subsequent Stops - Model Outputs versus Survey Data 7-27
Figure 7-18 Tour-Based Truck Model 7-28
Figure 7-19 Stop Frequency Comparison - Tour Model Outputs versus Survey Data – Medium Trucks 7-40
Figure 7-20 Stop Purpose Frequency – Model Outputs versus Survey Data – Medium Trucks 7-41
Figure 7-21 Time-of-Day of First Stop - Model Outputs versus Survey Data – Medium Trucks 7-41
Figure 7-22 Time-of-Day of Subsequent Stops - Model Outputs versus Survey Data – Medium Trucks 7-42
Figure 7-23 Stop Frequency Comparison – Tour Model Outputs versus Survey Data – Heavy Trucks 7-55
Figure 7-24 Tour Completion versus Stop Frequency - Model Outputs versus Survey Data - Heavy Trucks 7-55
Figure 7-25 Stop Purpose Frequency - Model Outputs versus Survey Data - Heavy Trucks 7-56
Figure 7-26 Time-of-Day of First Stop - Model Outputs versus Survey Data – Heavy Trucks 7-56
Figure 7-27 Time-of-Day of Subsequent Stops - Model Outputs versus Survey Data – Heavy Trucks 7-57

Figure 7-28 Duration of Subsequent Stops (in hours) - Model Outputs versus Survey Data – Heavy Trucks.....7-57
Figure 8-1 Percentage Heavy Trucks by Industry.....8-9
Figure 8-2 Shares of Heavy Trucks by Total Industry.....8-10
Figure 8-3 Medium Truck Count Locations.....8-12
Figure 8-4 Commercial Light Truck Count Locations.....8-13
Figure 8-5 Assignment of ODME – Medium Trucks.....8-14
Figure 8-6 Assignment of ODME – Commercial Light Trucks.....8-15
Figure 8-7 Comparison between Counts and Assigned ODME Flow of Medium Trucks.....8-16
Figure 8-8 Comparison between Counts and Assigned ODME Flow of Commercial Light Trucks.....8-16
Figure 8-9 ATRI Data Points.....8-17
Figure 8-10 Count Locations in Highway Network.....8-18
Figure 8-11 Assignment of ODME Heavy Trucks.....8-19
Figure 8-12 Comparison between Counts and Assigned ODME Flow of Heavy Trucks.....8-20
Figure 9-1 Working Directory Set Up.....9-3
Figure 9-2 Model Run in Command Prompt.....9-4
Figure 9-3 Overall Model Structure, Inputs, and Outputs.....9-5
Figure 9-4 Firm Synthesis Model Input Directory.....9-6
Figure 9-5 Firm Synthesis Model Output Directory.....9-9
Figure 9-6 Supply Chain Model Input Directory.....9-11
Figure 9-7 Supply Chain Output Directory.....9-13
Figure 9-8 Truck tour model input directory.....9-18
Figure 9-9 Truck Tour Model Output Directory.....9-20
Figure 9-10 Model Scripts and Properties Files in Programs Directory.....9-22
Figure 9-11 Firm Synthesis Model Scripts (fsm_xxxxxxxxxx.R).....9-23
Figure 9-12 Supply chain model scripts (scm_xxxxxxxxxx.R, ssm.jar, ssm_xxxxxx.properties).....9-24
Figure 9-13 Truck tour model scripts (ttm_xxxxxxxxxx.R).....9-25
Figure 10-1 External stations for MAG network.....10-2
Figure 10-2 External stations for Mega-regional network.....10-3
Figure 10-3 External zone allocation for PAG area.....10-4
Figure 10-4 Volume-Count Statistical Result for Freeway and Arterial.....10-8
Figure 11-1 Freight Model Visualization Interface at Launch.....11-2
Figure 11-2 Firm Synthesis Model Visualization in 2D.....11-2
Figure 11-3 Firm Synthesis Model Visualization in 3D.....11-3
Figure 11-4 Firm Synthesis Model Visualization on a National Scale.....11-3
Figure 11-5 Commodity Flow Visualization (Supply Chain Model).....11-4
Figure 11-6 Commodity Flow Visualization (TRANSEARCH).....11-5
Figure 11-7 Truck Tour Model Visualization - Trip Matrix.....11-6
Figure 11-8 Truck Tour Model Visualization - Individual Tour.....11-7
Figure 11-9 Truck Tour Model Visualization - Truck Volume Comparison.....11-7
Figure 11-10 Freight Model Visualization Workflow.....11-8
Figure 11-11 Truck Heat Map Animation Screenshot.....11-9
Figure 11-12 Vehicle Trajectory Animation using Leaflet.....11-10
Figure 11-13 Vehicle Trajectory Animation using CityPhi.....11-11

Executive Summary

The Maricopa Association of Governments (MAG), Arizona Department of Transportation (ADOT) and Pima Association of Governments (PAG) jointly developed a successful proposal for the Round 3 of Strategic Highway Research Program 2 (SHRP2) C20 Implementation and Technical Assistance Program (IAP) Grant. The Grant funds were designated for a pilot proof of concept study aimed at developing behavior-based freight model. MAG, ADOT and PAG proposal scope exceeded the original Grant offering and called for development of an operational mega-regional multimodal agent-based behavioral freight model based on the guidelines identified in the SHRP2 C20 findings. The model was to be developed in accordance with the agencies' travel forecasting and planning needs.

In order to achieve the above overall goal the following main project objectives were identified:

- Improve and expand the knowledge base
- Develop modeling methods to reflect actual supply chain management practices
- Develop modeling methods based on sound economic principles
- Maximize use of freight tools by public sector for planning and programming
- Improve availability and visibility of data between public and private sectors

This report describes the process in developing regional behavior-based freight transportation model for the Arizona Sun Corridor Megaregion based on above objectives. The Arizona Sun Corridor Megaregion is among the fastest growing megaregions in the country and a freight gateway to the international market which stress the significant importance of a freight policy analysis tool that is consistent with MAG freight flow forecasting needs to the megaregion's decision makers. The study tries to demonstrate the use of cutting-edge, behavior-based modeling approaches for evaluating freight policy impacts at the regional scale.

The structure of the report and a brief description of discussions in each chapter is explained in the following.

Chapter 1. Background

Chapter 1 outlines the organizational interest for SHRP2 project, regional leadership in developing freight demand modeling, executive support to adopt model, overall goals for the project, main methodological approaches, and expected deliverables from this project. It also explains the application process for FHWA SHRP2 C20 IAP program and grant of funds.

Chapter 2. Model Design

The model framework is expected to have 3 layers.

1) Financial/Economic layer
2) Logistics Layer

3) Transportation layer

In the economic layer each node has a certain amount of demand associated with it. In the logistics layer each node has supply associated with it with the supply chain model. The output from the logistics layer is rolled into the transportation layer via commodity flows and modes of transportation.

Chapter 3. Review of Freight Data Sources for the Development of a Behavior-Based Freight Model

This chapter discusses the model layers and the data requirements for each layer in the initial task. A list of commercially available data as well as public databases with a detailed description of its contents was compiled. Advantages and deficiencies of each of the datasets, sources and methods utilized in compiling the datasets (if available), aggregate statistics of databases (where available), including number of records, levels of geography, exclusions from datasets, pricing (where applicable and available), contact information (where applicable) and periodicity of updates were captured.

Data utilization from each source in the state of the practice and state of the art freight models was documented (whether they were candidates for estimation, calibration, and validation of models). Finally, data summary from each dataset was prepared at regional level or state level (depending upon its availability).

Chapter 4. Analysis of Datasets used in the Model Development

Based upon the data needs for each of the sub-models, there were specific datasets that were identified to develop the sub-models. MAG acquired or utilized the following datasets described in the chapter:

- National Establishment Time Series (NETS) Data
- American Transportation Research Institute (ATRI) Truck GPS data
- StreetLight Truck GPS data
- IHS Global Insight's TRANSEARCH data
- Freight Analysis Framework (FAF 4.1) data

This chapter discusses in detail the data statistics, coverage, and analysis.

Chapter 5. Firm Synthesis Model

This chapter describes the methodology for modeling firm demographics using a microsimulation approach for the Phoenix-Tucson megaregion. The model is based on principles of firmographics that predicts the location, magnitude, and size of firms in the study region. This model uses the National Establishment Time Series (NETS) database as a seed table. A series of econometric models are estimated to simulate the firm events that consider determinants such as firm internal attributes (size, age, and growth) and external attributes (market area characteristics, transportation costs, agglomeration economies).

The simulated results are validated with observed firm demographic trends along with zonal-level employment estimated using various goodness-of-fit measures.

Chapter 6. Supply Chain Model

Supplier selection model consists of two main modeling components (i) supplier selection model and evaluation of commodity flows and (ii) transport, mode and path choice model. The developed agent-based supply chain and freight transport model uses disaggregate behavior-based logistics and transportation choice models to simulate commodity flows at firm level. The model considers firms or business establishments as individual decision-making units in the freight transportation system. It assumes that logistics and supply chain decisions are made by business establishments. These logistics decisions include supplier selection, shipment size, and mode choice.

Supplier evaluation and selection process is among the most crucial logistics decisions that have been addressed extensively in supply chain management. This chapter introduces an agent-based computational economics (ACE) approach for supplier selection problem that was developed for the modeling supply chains in the MAG-PAG region.

The transportation choice models that are included in the MAG supply chain and freight transportation model are described in this chapter. A Nested Logit model is estimated for freight mode and shipment size choice. Four modes of transportation, including truck, rail, air and parcel (such as, U.S. Postal Service, UPS, and other couriers) are included in the model.

Chapter 7. Tour-Based Truck Model

The objective of truck tour model is to develop truck trip chains by industry sector by truck type. These truck trip chains are then grouped into the major linkages based on land uses the trucks make stops at and the probability of making another stop based on the number of previous stops. The tour-based model generates the number of stops by industry sector, number of stops on a tour, stop purposes, and the location and time of day of stops. This chapter describes the processing of truck GPS data from two different sources ATRI and Streetlight, and developing tour-based models by truck type using the processed GPS data.

Chapter 8. Integration of Supply-Chain and Tour-Based Models

The supply chain model outputs annual commodity flows in tons by commodity group that will need to be integrated with rest of the model system. The Truck Tour model produces tours of truck that travel from an origin of the tour to the destination of the tour with intermediate stops during the tour. This tour output is in the right time period (i.e., average weekday) and the legs of the tour, i.e., between each stop of the tour, can be unchained into a trip table, by making each leg of the tour into an individual trip between the stops.

From a highway assignment standpoint, only the truck flows from the supply chain model are used while rail, water and other modes of freight are not assigned to any networks. The trucks flows are converted to daily truck tours and trips and integrated with the highway assignment model

Chapter 9. Model Implementation and Deployment

The behavior based freight model has been implemented using R and Java script language. This chapter discusses the software and hardware requirements, directory structure, input and output files to run the integrated behavior based freight model.

Chapter 10. Overall Freight Model Validation

The output of the freight model runs are in CSV format, to assign to TransCAD highway networks the CSV format results are converted to TransCAD Matrix (.mtx) format and combined them with MAG & PAG travel demand model passenger model vehicle trip matrices. These trip matrices are separated for MAG and PAG regions and then assigned separately to the respective regions highway network.

The assignment validation is done at two different levels of geography. i) Screenlines analysis includes some of the major freeways that pass through the region and carry a large volume of trucks in the region. ii) The sum of truck volumes at all locations were compared against the observed counts by truck class and area type.

Chapter 11. Visualization

The MAG project team developed a number of interactive data visualization tools for freight model input and output data using a combination of open-source libraries and proprietary software. These interactive visualization tools facilitated quality control and assurance and ensured effective communication of the data analysis and modeling results to stakeholders. The main visualization tool combines several model components' outputs and utilizes JavaScript web mapping libraries to visualize various datasets at different levels of geography. Other innovative visualization approaches were explored to convert massive truck GPS data into information useful for freight planning and analysis. Most of the visualizations are easily accessible via a web browser. This chapter will describe the main visualization tools and discuss how they were developed.

1 Background

Vladimir Livshits

The Federal Highway Administration's Strategic Highway Research Program (SHRP2) Freight Demand Modeling and Data Improvement (C20) project aims to create a framework for innovative and improved freight datasets and freight modeling practices, to advance the state of the industry. The project includes two components – The Freight Demand Modeling component and Freight Data component, seeking breakthrough solutions for wide application. As part of the Implementation Assistance Plan (IAP) for this project, the deployment and evaluation of eleven proof-of-concept pilot projects was the top priority. Four of those pilot projects were in modeling, while seven were in obtaining local freight data, forming the basis of innovative approaches in freight demand modeling and data collection respectively.

In February 2014, Maricopa Association of Governments (MAG), the Arizona Department of Transportation (ADOT) and the Pima Association of Governments (PAG) submitted a joint application for implementation and technical assistance funds in Round 3 of the FHWA SHRP2 C20 IAP. MAG was the lead agency in the Freight Demand Modeling portion of the proposal submitted to FHWA. The biggest challenge in freight modeling that MAG has experienced so far was lack of data for model development [1]. Truck and commercial vehicle surveys proved to be very challenging and a very expensive data collection exercise [2]. In another application to FHWA as part of SHRP2 C20 IAP, MAG was the lead agency as well in the Local Data Proof of Concept, because of the potential costs involved.

In March 2014, FHWA and AASHTO announced that MAG was one of the organizations selected in Round 3 of SHRP2 implementation assistance for Freight Demand Modeling. Up to $350,000 for each applicant is available to be applied to the implementation of Behavior-Based Freight Modeling. The main purpose of the project was to develop a new generation regional freight model based on the principles identified in the SHRP2 C20 products and consistent with MAG freight flows forecasting needs.

1.1 Motivation

The Arizona Department of Transportation (ADOT) is the state transportation agency. MAG is a Council of Governments that serves as the regional planning agency for the Phoenix Metropolitan Area. PAG is a Council of Governments that serves as the regional planning agency for the Tucson Metropolitan Area.

Understanding freight flows and forecasting them are critical in determining the need for future transportation capacity on the regional highways and/or other modal infrastructure. MAG staff identified a need for the future development of the regional freight forecasting models. Expanding metropolitan regions have created a new scale of geography known as the mega-regions. Sharing environmental systems and topography, infrastructure, economic linkages, and culture and history, these regions cover thousands of square miles and are located in every part of the United States.

Arizona's Sun Corridor mega-region begins in Santa Cruz County at the southern border City of Nogales. The Sun Corridor continues north along Interstate 19 to Tucson, and then west along

[1] MAG_ADOT_PAG_Application_C20_1_IAP_Application_Model.pdf
[2] MAG_ADOT_PAG_Application_C20_1_IAP_Application_DATA.pdf

Interstate 10 to the Phoenix-Mesa metropolitan area. Prescott is at the north end of the Sun Corridor. Covering portions of five counties that include the MAG and PAG regions, the Sun Corridor is home to eight out of ten Arizonans. The Sun Corridor is a major gateway for fresh produce and manufactured goods from busy freight ports on the U.S.-Mexico border.

Driven by activity in the Sun Corridor, Arizona is expected to experience growth more than two times the national rate of growth between 2015 and 2050. Census 2010 showed that the region ranks tenth among the largest U.S. metropolitan markets by population. In addition to anticipating regional growth in freight traffic, the Sun Corridor was envisioned to be evaluated as strategic bridge between major markets within the global supply chain. Importance of freight modeling and regional truck movements forecasts have long been recognized by ADOT and MAG. Both agencies are developing and maintaining state-of-the-practice truck models. MAG began researching and maintaining truck models more than two decades ago. Given the expected growth in freight and its importance to the regional economy, MAG needs improved capabilities to analyze freight demand and in this context, behavior-based freight modeling is pursued.

Freight movement and truck movement forecasting in particular are critical for the main DOT and MPO business processes such as air quality conformity analysis, development of regional and state transportation plans and planning work on regional transportation projects. The importance of freight has grown substantially during the past few years as a result of heightened attention to regional and state-wide economic issues. The recent economic recession severely impacted Arizona and MAG region in particular. Freight is being seen as one of the key transportation elements that will contribute to economic recovery.

There are specific organizational goals in the MAG Fiscal Year 2015 Unified Planning Work Program (UPWP) and ADOT State Planning and Research Work Program that further elaborate on the interests that drive behavior-based freight modeling. These goals include development of the regional transportation plan, development of freight transportation plans and fostering transportation-related regional economic development.

Extensive experience in developing and applying regional truck models led to a realization of the importance of behavior-based facets of freight decision-making. ADOT and MAG realized that the decisions of shippers and carriers, such as tour patterns for example, significantly affect regional transportation forecasting. One of the latest developments at MAG included a truck tour-based model that resulted in a noticeable improvement in forecast validations.

1.2 Historical Outline of Main Relevant Studies in Arizona

MAG has long history of developing urban truck models. One of the first models in the Nation was developed for MAG in 1992 and became a key reference and case study in 1996 FHWA Quick Response Freight Manual. MAG urban truck model was continuously updated and improved over the years with major improvements implemented in 2008-2010 and in 2012-2013 model recalibration efforts.

In 2004, MAG completed a comprehensive "Regional Freight Assessment Study". MAG has also conducted a truck survey in 2007 and is currently working on phase two of the Freight Transportation Framework Study. ADOT and MAG have jointly purchased the 2009 IHS Global Insight TRANSEARCH database to use in all of the freight planning efforts not only in the MAG region but throughout the State of Arizona. In addition to the purchase of Global Insight data, MAG also

purchased Truck GPS data to augment the regional truck model update. MAG conducted a Truck Model and Vehicle Weight Workshop in 2011 by inviting several regional freight operators and officials from public agencies to foster a better understanding of future data needs and current modeling capabilities.

MAG's latest freight modeling effort consists of a trip based truck model developed with a variety of data, and included new data collected, acquired and purchased. The lack of statistically relevant data needed to update the internal truck model was overcome with the purchase of third-party truck GPS data. The data was purchased from commercial vendors and subsequently processed to develop robust truck modeling parameters (rates, trip lengths and trip interchanges). Light truck trip rates were asserted based upon FHWA's "Accounting for Commercial Vehicles in Urban Truck Models". New screenlines were created that encompassed all the major freeways, state highways and major arterials in the region. Several count programs were used to compile vehicle classification counts for the external stations, screenlines and for time-of-day factors for trucks. A tour-based modeling framework was also developed for heavy trucks to demonstrate the use of truck GPS data and also to move MAG's trip based truck model to a tour based platform in the coming years.

ADOT developed its first Arizona Statewide Travel Demand Model (AZTDM1) in 2009. AZTDM1 used the Quick Response Freight Manual (QRFM) for truck trip generation. The second generation model (AZTDM2), which was developed in 2010, incorporated a considerable amount of additional detail through work already completed by ADOT staff. The freight and goods movement component of AZTDM2 was based on available data from sources such as the FHWA Freight Analysis Framework (FAF3), ADOT's ATR counts, and the Bureau of Transportation Statistics (BTS) databases and MAG's short distance truck model. The 3rd generation Arizona Statewide model (AZTDM3), initiated in July 2011, focuses on transitioning the AZTDM into a 4-step model. ADOT purchased the TRANSEARCH database and used it to develop the long distance truck flows.

ADOT has been collecting vehicle classification counts manually from 2002 to 2005. After 2005 ADOT installed automatic traffic counters (ATR) to collect vehicle classification counts. Such vehicle classification counts database has been used to calibrate/validation ADOT's latest AZTDM model. In addition to the regional freight planning and modeling efforts, ADOT completed the "Multimodal Freight Analysis Study in 2008 and is planning to start the State Freight Plan in the fall of 2014.

1.3 Executive-level Support

MAG and ADOT have made freight planning and modeling a priority for both the MAG region and the State of Arizona. From an MPO perspective, the three councils of governments that represent the Arizona Sun Corridor signed a planning resolution recognizing the need to plan together to be successful as a mega-region. This planning resolution formed the Joint Planning Advisory Council (JPAC). In 2010, the JPAC launched the Freight Transportation Framework Study that was led by MAG and its JPAC partners. The main goal of the study was to identify freight-related economic development opportunities. The JPAC partnership was signed by the elected officials that represent each council of governments highlighting this as priority for the region.

The Arizona Department of Transportation formed the Transportation and Trade Corridor Alliance (TTCA) that has created partnerships with the Arizona Mexico Commission (AMC), the Arizona Commerce Authority (ACA), private sector, state and local governments, planning organizations, transportation and logistics companies and port authorities to identify and assess freight related

economic development opportunities for Arizona. In addition, the TTCA will serve as the state's freight advisory committee, as required under the recently enacted federal transportation bill. The TTCA will address defining a new approach that incorporates local initiatives into a broader state context as it pertains to transportation and trade opportunities, identifying ways to increase the value of the state's trade corridors, opportunities and challenges that impact global logistics in the state of Arizona, and Integrating rail and/or intermodal facilities for sustainable growth and job diversification in the state.

MAG, ADOT and PAG would be pivotal in understanding the geographic scope of freight activity in determining local benefits through such initiatives. Enhanced efficiency in freight movement will reflect the economic competitiveness of the region and the state. MAG and ADOT will use the results from this effort to assist with the regional and state freight planning efforts and provide crucial data to assist the policy makers that represent the Arizona Transportation and Trade Corridor Alliance and the Joint planning Advisory Council to make the right policy decisions that will continue to keep Arizona and the United States globally competitive.

A 2008 Shipper Interview conducted by ADOT offered insight about shippers appreciating the local labor availability, ability to move freight in snow-free facilities, proximity to West Coast ports, lower in-transit theft than West Coast and a very favorable representation of their business interests by Arizona Trucking Association. MAG would strive to create favorable conditions for freight transportation in the future based upon a detailed review of each market segment that offers specialized transportation services. It would also collect information on various business models for each of these market segments allowing them to remain competitive.

1.4 Overall Goals for the Project

The most important milestone of the project as envisioned by MAG, ADOT and PAG would be the implementation of the operational supply-chain and tour-based models in the state-wide and regional models respectively, besides the development of related datasets on state and regional levels. The proposed models were aimed to improve accuracy of truck movements in the regional and state-wide models, adequately demonstrated by validation results.

The specific application issues that will be resolved include identification of required infrastructure improvements, more accurate forecasting for the purpose of regional, state-wide planning, air quality conformity analysis and support of the activities aimed at economic revitalization of the state and the region. With a pilot study, MAG plans to assess the feasibility of a full behavioral freight model to replace its current state-of-the-art practice truck model.

- Provide a much-needed understanding of the discrete segments of the freight transport community, including shippers, carriers, customers, and other elements of the supply chain;
- Help public sector agencies gain a better understanding of the impacts policy decisions have on individual freight transport segments;
- Develop a well-rounded and representative understanding of freight movement that does not generalize or assume that freight movement activity is similar across different industry sectors;

- Improve consistency between passenger state-wide and regional modeling frameworks and freight models;
- Provide insight on service availability, pricing, and reliability as performance measures for different industry sectors;
- Develop improved understanding of intermodal freight movement;
- Understand how interventions and/or policies related to the multimodal freight system in Arizona could impact the location of warehouses and industries, as well as the distribution of flows in the state and the MAG region.

1.5 Main Methodological Approaches

The proposed model development is based on a few main methodological principles:

1. Development of an agent-based model that will implement a micro-simulation approach to the freight system modeling in MAG region, including synthesis of establishments, disaggregation of relevant demo-economic data in the region, formation of supply-chain and establishment interaction models and tour-based travel models.
2. Development of a multi-modal freight model at least for major commodities relevant for the MAG-PAG mega-region based on commodity-based analysis of freight flows. High concentration of commodities and relatively limited number of large trip generators in the mega-region contribute to feasibility of this approach in order to capture main freight travel patterns. By providing an in-depth look at each segment of the industry it will be possible to better understand the drivers of each industry and model them specifically.
3. Behavioral approach to the model development, including modeling of the economic behaviors of establishments, shippers and carriers in generation of travel and tour formations. Examples of a freight agent's behavior can include the reaction of supply chains to variations in fuel price, global sourcing and manufacturing decisions and technology changes. The model can assist in coordinated policy development and can be an effective tool for improving freight operations.
4. Consideration of the MAG activity-based passenger modeling framework in order to facilitate future integration of the models. Model structure at this stage is envisioned including such major components as establishment synthesis, establishment interaction model, mode choice model and tour formation model.
5. Industry Specific – The model should be based in explanatory variables specific to each industry, which allows for industry related people to have a better understanding of the model and provide directly applicable data and feedback.

These activities will be further detailed during the scoping phase of the project. This effort intends to contribute to the following strategic objectives of SHRP2-C20 project:

- Improve and expand the knowledge base.
- Develop modeling methods to reflect actual supply chain management practices.
- Develop modeling methods based on sound economic principles.

- Maximize use of freight tools by public sector for planning and programming.
- Improve availability and visibility of data between public and private sectors.

In order to achieve the above objectives, the following SHRP2 research initiatives will be addressed in this work:

1. Establish techniques and standard practices to validate freight forecasts.
2. Establish modeling approaches for "behavior-based" freight movement.
3. Establish analytical approaches that describe how elements of the freight transportation system operate, perform, and impact the larger overall transportation system.
4. Determine how economic, demographic, and other factors/conditions drive freight patterns and characteristics. Document economic and demographic changes related to freight choices.
5. Advance research to effectively integrate logistics practices (private sector) with transportation policy, planning, and programming (public sector).

It should be noticed that different research initiatives listed above will be addressed to different extent. For example, development of behavioral sub-models might be constraint by the datasets that can be made available for the model development purposes. Emphasis of this effort will be on behavioral tour-based model for a large metropolitan area. Actual freight forecasting model validation efforts might be restricted to the data available for MAG region, even though other theoretical possibilities for model validation can be investigated. Validation practices might change as new data sets and new types of data become available. Integration of private sector practices with public sector processes is a very broad area. The project will only address a portion of it applicable for the freight modeling in a regional transportation planning environment.

1.6 Expected Deliverables

The overall duration of the project is two years, from the kick-off date, while the expected deliverables from the project are listed below.

1. Detailed work plan – technical memo.
2. Data collection – final data sets to be used for the models development, data acquisition and data preparation. Data collection reports or technical memos.
3. Detailed data analysis, including commodity flow analysis. Data analysis report or technical memo.
4. Identification and list of main traffic generators and main industry sectors. The list and accompanying memo.
5. Establishments/firm synthesis models. Models, source code and models documentation.
6. Behavioral establishment interaction model/ distribution models. Models, source code and models documentation.
7. Tour formation models for different truck types/industries.

8. Multimodal network and corresponding documentation.
9. Mode choice models. Models, source code and models documentation.
10. Operational meg-regional freight model and model documentation.
11. Final report.

2 Model Design

Consultant Technical Lead – Arun Kuppam (Cambridge Systematics), Zahra Pourabdollahi (RS&H), and Monique Stinson (MIT)

Project TAG, FHWA

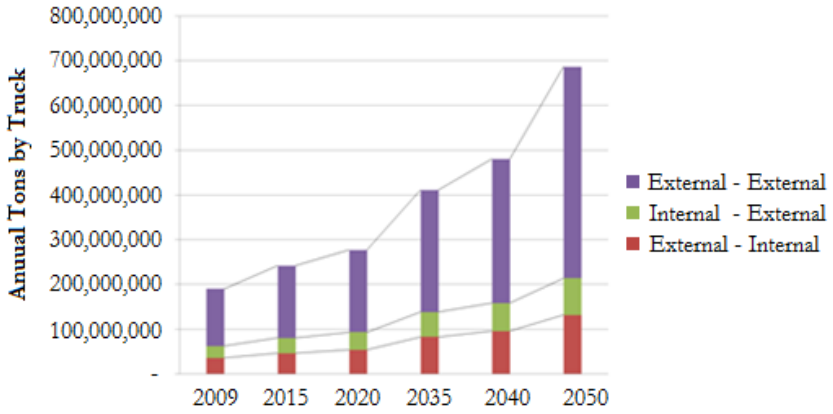
2.1 Background

Arizona has defined the Sun Corridor megaregion as covering portions of five counties that include the MAG and Pima Association of Governments (PAG) regions, and is home to eight out of ten Arizonans. The corridor begins in Santa Cruz County at the southern border of the City of Nogales, which is one of the busiest freight ports along the U.S.-Mexico border, and is a major gateway for fresh produce and manufactured goods. The corridor continues north along I-19 to Tucson, and then west along I-10 to the Phoenix-Mesa metropolitan area, and ending in the City of Prescott at the north end. Driven by activity in the corridor, Arizona is expected to experience growth more than two times the national rate of growth between 2015 and 2050. The Census 2010 showed that the region ranks 10th among the largest U.S. metropolitan markets by population.

The importance of freight modeling and regional truck movement forecasts has long been recognized by ADOT and MAG. Both agencies are developing and maintaining state-of-the-practice truck models. MAG began researching and maintaining truck models more than two decades ago. There are several reasons for these efforts and for the agency's continuous interest in improving freight modeling and taking the next steps in adopting behavior-based freight modeling.

- First and foremost, freight movement and truck movement forecasting, in particular, are critical for the main DOT and MPO business processes, such as air quality conformity analysis, development of regional and state transportation plans, and planning work on regional transportation projects.
- Second, the importance of freight has grown substantially during the past few years as a result of heightened attention to regional and statewide economic issues. The recent economic recession severely impacted Arizona and the MAG region in particular. Freight is being seen as one of the key transportation elements on the way to economic recovery.

Figure 2-1 Total Truck Tonnage by Direction [3]



The most recent update to the MAG Truck Model used the most recent Commodity Flow dataset (MAG TRANSEARCH 2009). It showed that 76 percent of the tons carried by truck that originate in Phoenix also terminated in Phoenix. It also showed that 11 percent of those freight truck tons terminated elsewhere in Arizona; 10 percent terminate in adjacent states, including Mexico; and 3 percent terminated elsewhere in the U.S. It showed that while 70 percent of the tons carried by truck that terminate in Phoenix also originated in Phoenix, 11 percent of those freight truck tons originated elsewhere in Arizona; 13 percent originated in adjacent states, including Mexico; and 10 percent originated elsewhere in the U.S. Additionally, that study showed that these tonnages are exceeded by the importance of the traffic passing through the Sun Corridor, between California and its ports, and the rest of the domestic U.S. As shown in Figure 2-1, the through (external-external) freight traffic is growing faster than freight traffic to (external-internal) or from (internal-external) the region. With the anticipation of regional growth in freight traffic, it is key to model the supply chain movements to, from, and through the Sun Corridor and its place within the global supply chain. This will lead to a better understanding of how improvements to freight supply chains can benefit freight traffic in the Sun Corridor.

[3] Source: 2009 TRANSEARCH, Cambridge Systematic, Inc.

2.2 Supply Chain – Tour-based Freight Modeling Framework

The following data has been compiled and analyzed pertaining to freight and truck travel in the MAG region:

- 2009 TRANSEARCH commodity flow data;
- FHWA's Freight Analysis Framework 3 (FAF3) commodity flow data;
- 2012 ATRI Truck GPS data;
- 2007 truck trip diary surveys and operator (establishment) surveys;
- 2002 Vehicle Use and Inventory Survey (VIUS) data;
- 2008 Department of Motor Vehicles (DMV) registration database; and
- 2010/2011 vehicle classification counts and truck screenlines.

The following model components developed as part of past contracts have also been reviewed and used for developing this model:

- TRANSEARCH commodity-flow based external truck travel model for heavy trucks;
- Truck travel survey and GPS databased internal truck travel model for medium and heavy trucks;
- Fleet size-based light truck commercial vehicle model;
- ATRI truck GPS databased tour-based model for heavy trucks; and
- Integrated truck travel model (all truck types) with MAG's passenger travel model.

All of the above have been considered while developing MAG's new behavior-based freight model. This new model follows the framework as depicted in Figure 2-2.

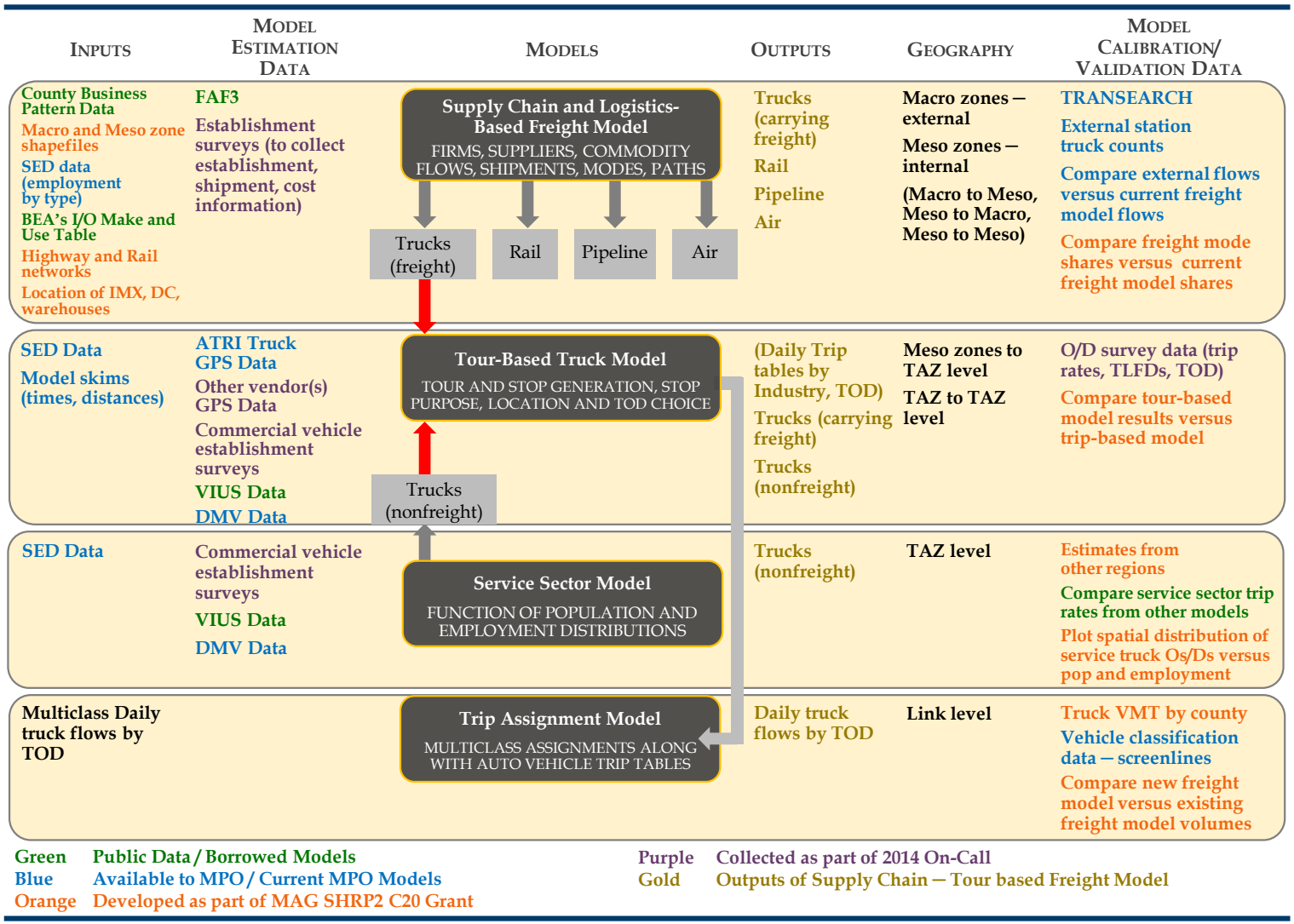
This figure provides the following information:

- **Inputs**. This includes all the data inputs necessary to apply the individual modeling components in the new model system. The key data elements include socioeconomic data, network data, county business patterns, and the U.S. Bureau of Economic Analysis (BEA) make-use tables. The flowchart also depicts the status of these data items; that is: 1) if the data can be borrowed; 2) is available to the MPO; 3) need to be developed as part of this project; or 4) to be collected as part of the Travel Survey – Data Application On-Call.
- **Model estimation data**. This includes data necessary to estimate various components of the supply chain and tour-based models. The data includes commodity flow data, establishment surveys, truck GPS data, and any existing or new commercial vehicle surveys.
- **Models**. These include supply chain models, such as firm synthesis; supplier selection model; apportionment model; transport and path choice model; vehicle type choice model; tour-based

modeling components, such as stop generation, purpose, location and time-of-day choice model; and a service sector model.

- **Outputs**. This lists the key outputs that are produced from each modeling component, such as tons by mode, trip tables by truck type, and daily truck flows.
- **Geography**. Each modeling component is applied and produces outputs at different geographic levels – macro, meso, and micro. The exact definitions of these are subject to change based on the objectives and desires of MAG's planning needs.
- **Model calibration/validation data**. This includes data that will be used to calibrate each component and validate the new model system. TRANSEARCH data, external station counts, vehicle class counts, origin-destination (O-D) survey data, and observed (Highway Performance Monitoring System (HPMS)) vehicle miles traveled (VMT) for trucks are some of the most essential data for calibration and validation. This behavior-based freight modeling framework ultimately will enable the simulation of freight network performance at a detailed level, as anticipated by SHRP2 C20 objectives. It will simulate individual vehicle movements, allowing MAG to understand the impacts of transportation infrastructure investments and operations improvements at a very detailed level.

Figure 2-2 Supply Chain - Tour-Based Freight Modeling Framework [4]



[4] Source: Cambridge Systematic, Inc.

2.3 Regional Behavior-Based Freight Modeling

Among MAG Consultants expertise was the development of an innovative supply chain and logistics-based freight modeling framework for the Los Angeles County in 2005. [5] This framework consisted of three layers for evaluating freight movements going to, from, and through the Los Angeles region. These are:

1. **Economic Layer**, which captures the macroeconomic factors that influence freight demand and commodity supply/demand relationships.
2. **Logistics Layer**, which captures the supply chain and logistics decisions that firms make in response to supply and demand conditions (including choice of distribution channels, modes, etc.).
3. **Transport Layer**, which captures the transportation system costs and performance that are linked to logistics decisions, and ultimately determine the traffic/vehicle flows on multimodal networks

These **logistics chain models** work best when the goods transported move in large lot shipments, with little mixing of the commodities, often in single shipment transport loads, or in multiple shipment loads delivered in a "descending load, return empty" pattern. Additionally, the logistics chains should be relatively simple with few choices of distribution channel options. For domestic freight movements, tour-based modeling approaches work best. The **tour-based model** components of the proposed modeling framework track the activity of vehicles. Since these components will operate at the vehicle level, they will only generate estimates of a single mode. Vehicles are associated with establishments, and vehicle activity is seen as a function of the type of activity that occurs at that establishment. The tour-based components will operate with zones, as do passenger models; and the activity estimates will be aggregated for all of the establishments in a zone. For example, the number of truck trips in a zone could be a function of the number of manufacturing facilities, distribution centers, truck terminals, and even population. The tour-based components of the proposed framework will generate the number of stops that have to be made in each zone for a particular type of truck (e.g., for a parcel pickup and delivery operation), and then string these trips together into tours. The number of stops on a tour, the type of stops, and the location of stops are all estimated from the model based on the type of truck making the tour, the activities conducted by the truck, the characteristics of the stops, and the traffic conditions in the network.

Although the underlying behaviors in logistics chain modeling and tour-based modeling are different, the basic structure of the models is similar, which makes their **integration** very easy. Each type of model represents an activity chain. In each step of the chain, choices are made concerning what activity will occur next, the type of facility to be used and the location of the facility. These choices lead to other choices about the modes, or types of equipment to be used and the loads the equipment will carry. These choices will be modeled using a discrete choice simulation model incorporating various forms of logit and nested logit choice models. Each choice in the logit model is a function of the location of attractions in the chain, the magnitude of these attractions, the characteristics of each facility or stop option, and previous choices made in the activity chain. Although the underlying behavior of each type of model has

[5] Fischer, M. J., et al., Cambridge Systematics, An Innovative Framework for Modeling Freight Transportation in Los Angeles County, presented at the 84th TRB Annual Meeting, January 2005.

noticeable differences, the basic mathematical structure is similar, making the formulation of an integrated modeling approach mathematically easier to accomplish. The common mathematical structure of the two models also suggests the possibility of using common software for the integrated model. The tour-based model developed for MAG using the ATRI GPS data [6] was envisioned to be integrated with the logistics chain models in this project.

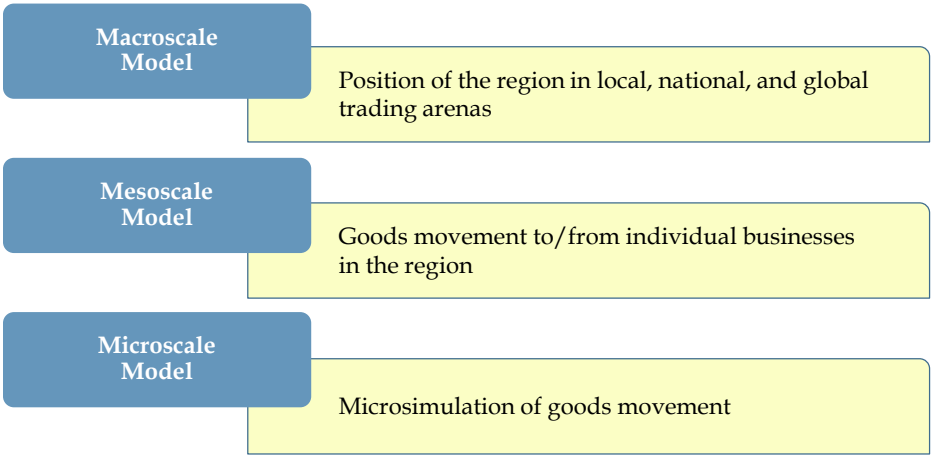
Chicago Metropolitan Agency for Planning (CMAP), adopted the three layer framework – economic, logistics, and transport – and developed a prototype freight model that consists of macroscale, mesoscale, and microscale models in a supply chain and logistics-based setting [7]. This is depicted in Figure 2-3.

The framework for CMAP was designed as part of a Federal grant with the intention of developing a freight modeling tool that addresses changing conditions in freight supply and demand in the Chicago region. A prototype mesoscale model was developed for CMAP in 2010 to 2011 that has the ability to analyze high-level goods movement data from a macroscale model at a scale that is more insightful for agency analyses. Freight supply is provided by infrastructure and carriers or other operators, while demand for freight transportation is generated by individual businesses that are trading goods with one another. The prototype models transportation decisions, such as mode choice and the use of logistical handling facilities (e.g., intermodal yards) for individual firms. In doing so, the prototype adopts an intuitive, agent-based approach for understanding the underlying economics that motivate business decisions. These individual business decisions are then effectively aggregated by the model into trip tables that can be used for a variety of highly insightful analyses. Example applications of the model included evaluating the potential demand for a new airport from the perspective of businesses across the region.

The prototype model for CMAP used a combination of observed and placeholder data and methodologies. The framework proposed under Figure 2.4 is derived from the CMAP work, which formed the basis for the development of MAG's new behavior-based freight model. It should be noted that the foundational framework for the development of the behavioral freight model was proposed in Freight Activity Microsimulation Estimator (FAME)[8]. Please see Section 6.5 for other relevant references.

[6] Kuppam, A., J. Lemp, D. Beagan, V. Livshits, L. Vallabhaneni, and S. Nippani, Development of a Tour-Based Truck Travel Demand Model Using Truck GPS Data, presented at 93rd Annual Meeting of the TRB, January 2014.
[7] Cambridge Systematics, Inc., A Working Demonstration of a Mesoscale Freight Model for the Chicago Region – Final Report and User's Guide, prepared for Chicago Metropolitan Agency for Planning, June 2011.
[8] Samimi, A., A. Mohammadian, and K. Kawamura. "A Behavioral Freight Movement Microsimulation Model: Method and Data", Journal of Transportation Letters: The International Journal of Transportation Research. 2(1):53-62, 2010.

Figure 2-3 CMAP Freight Model Components [9]



The proposed framework will enable the following things:

- The freight modeling tool also will address the changing conditions in freight supply and demand in the MAG region. Freight supply is provided by infrastructure and carriers or other operators, while demand for freight transportation is generated by individual businesses that are trading goods with one another.
- The new model will capture transportation decisions, such as mode choice and the use of logistical handling facilities (e.g., intermodal yards) for individual firms. In doing so, the model adopts an intuitive, agent-based approach for understanding the underlying economics that motivate business decisions. These individual business decisions are then effectively aggregated by the model into trip tables that can be used for a variety of highly insightful analyses.
- The model ultimately will simulate freight network performance at a detailed level in the anticipated microscale model. This component of the model will simulate individual vehicle movements, allowing MAG to analyze impacts of new infrastructure projects at a highly detailed level.

This framework would allow MAG to advance research in behavior-based freight modeling principles. Ideally, such a framework provides insights on the role of a region in today's national and global economies. The modeling framework is designed to address questions that are macroeconomic in nature with a special focus on the movement of goods into, out of, and through the region. However, the responsiveness to changes in macroeconomic conditions, such as increasing energy costs worldwide, increasing labor costs in Asia, or fluctuations in exchange rates, can only be realized if there is a national and global network that can evaluate these policies. At the moment, these models or processes do not exist. Moreover, it is beyond the scope of this study.

It is unreasonable for any region, including MAG, to develop a model that requires data far outside of its jurisdiction. Though such a model does not exist, the FHWA's FAF data does provide aggregate fixed

[9] Source: Cambridge Systematics, Inc.

interregional flows and forecasts at a national level. It may not be possible for MAG to vary inputs to see how flows between Chicago and Los Angeles might shift, but the proposed framework would allow the disaggregation of flows between Phoenix and its trading partners into supply chains, allocate flows among these supply chains, and allow testing of various polices that could cause shifts among these supply chains. Not only will this provide valuable insights to MAG, it will help make the case that multiregional (i.e., global or national) freight models should be developed down the road that could be used by MAG in evaluating macroeconomic freight policies.

2.3.1 Modeling Framework

A supply chain and logistics-based model for freight can be considered analogous to activity touring models for passenger trips. However, while passenger tours are almost always begin and end within the area served by MPO planning models, this is not the case for freight supply chains. Those freight supply chains may begin, end, or have intermediate stops, including stops at truck to truck distribution centers, outside of the model region. This is an important consideration in developing freight supply chain tours. While it is reasonable to synthesize passenger trip end for populations within the region's model, limiting freight trip ends to synthesis of firms within the region is not sufficient. It is not reasonable for any region to have sufficient information for all of the firms located outside of its regions in order to synthesize freight trips for those firms. Further, the supply chains that travel between firms outside of the region without passing through the region may be of no interest to the region, even though understanding the supply chains that are formed that impact the region are constrained by the national universe of all supply chains.

Therefore, it should be a national task to inventory and forecast all freight moments between regions. The FHWA has assumed this responsibility in preparing its FAF. It forecasts the flow of freight in complete supply chains for the entire U.S. For this study of detailed supply chains within a region, such as the Sun Corridor, it is necessary to extract its supply chains from such a national database, and to deconstruct the sum of all supply chains into specific supply chains.

The proposed steps in the supply chain model are:

1. Firm Synthesis;
2. Supplier Selection;
3. Apportionment of Commodity Flows;
4. Mode, Distribution Channel, and Path Selection; and
5. Preparation for Assignment

Mathematically, the flows that are reported in FAF are:

$$T_{ri,rj,c} = \sum_{sf \in ri, rf \in rj, p} T_{sf,rf,c,p}$$

Where:

$T_{r_i,r_j,c}$ = The freight flow (as tons or value) being shipped between ri, region *i,* and rj, region *j* of commodity *c*; and

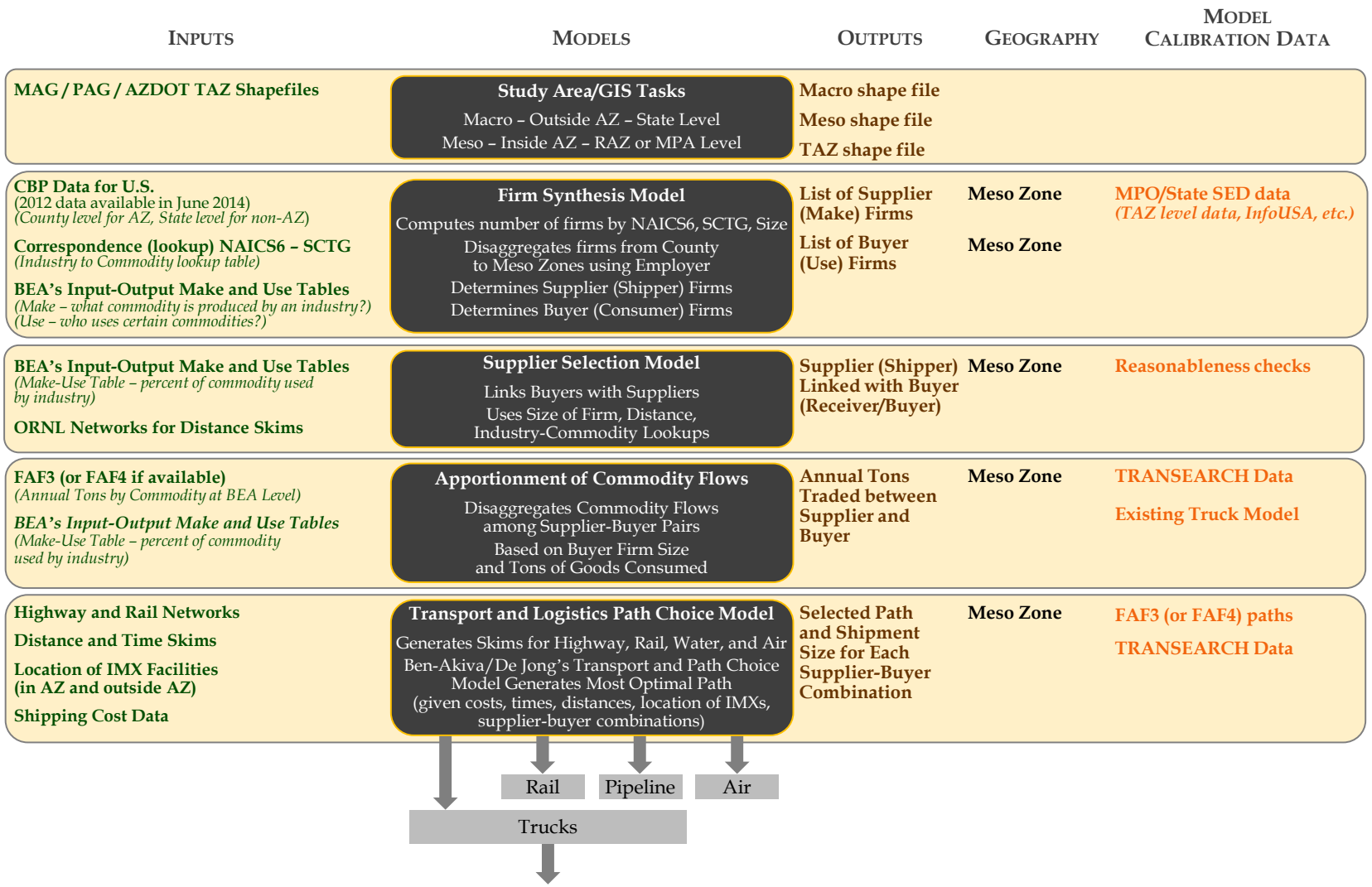
$T_{sf_i,rf_j,c,p}$ = The freight flow (as tons or value) being shipped between sf_i, a shipping firm that is located in ri, region *i,* and rf_j, a receiving firm that is located in rj, region *j,* for a commodity *c* using a path *p*.

The equality is produced by summing the freight flows of a commodity between firms over all shipping firms that are located in ri, region *i* and all receiving firms that are located in rj, region *j*, and over all paths *p* that used to connect those firms.

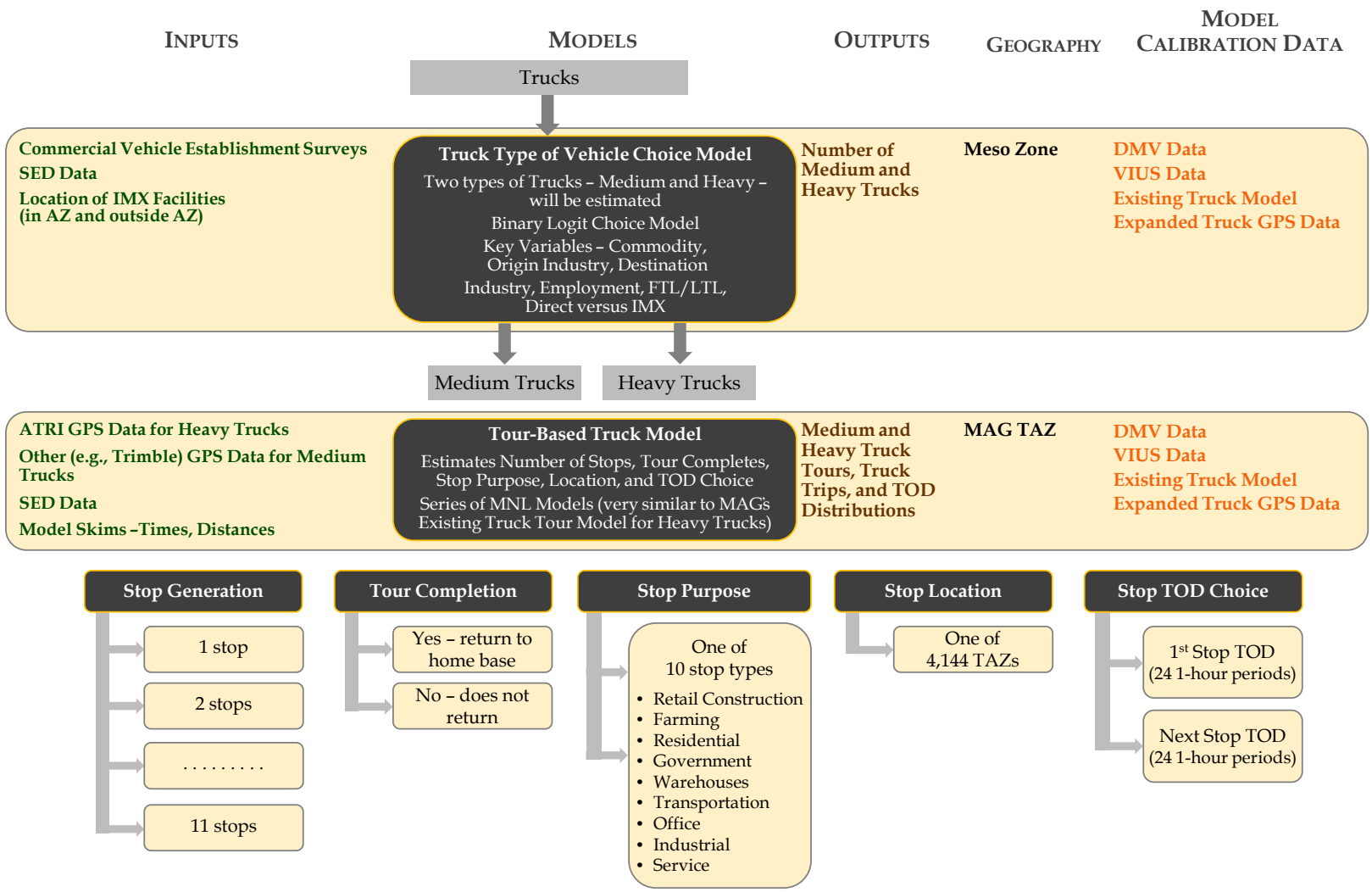
Rather than calculate the trip generation and trip distribution steps as in a traditional travel demand model, the various flows by firm and path will be calculated, and their utilities relative to each other will be used to apportion the FAF (macro) freight flows.

The overall approach and individual modeling components are depicted in Figure 2-4.

Figure 2-4 Individual Components of the Proposed Behavior-Based Freight Model [10]



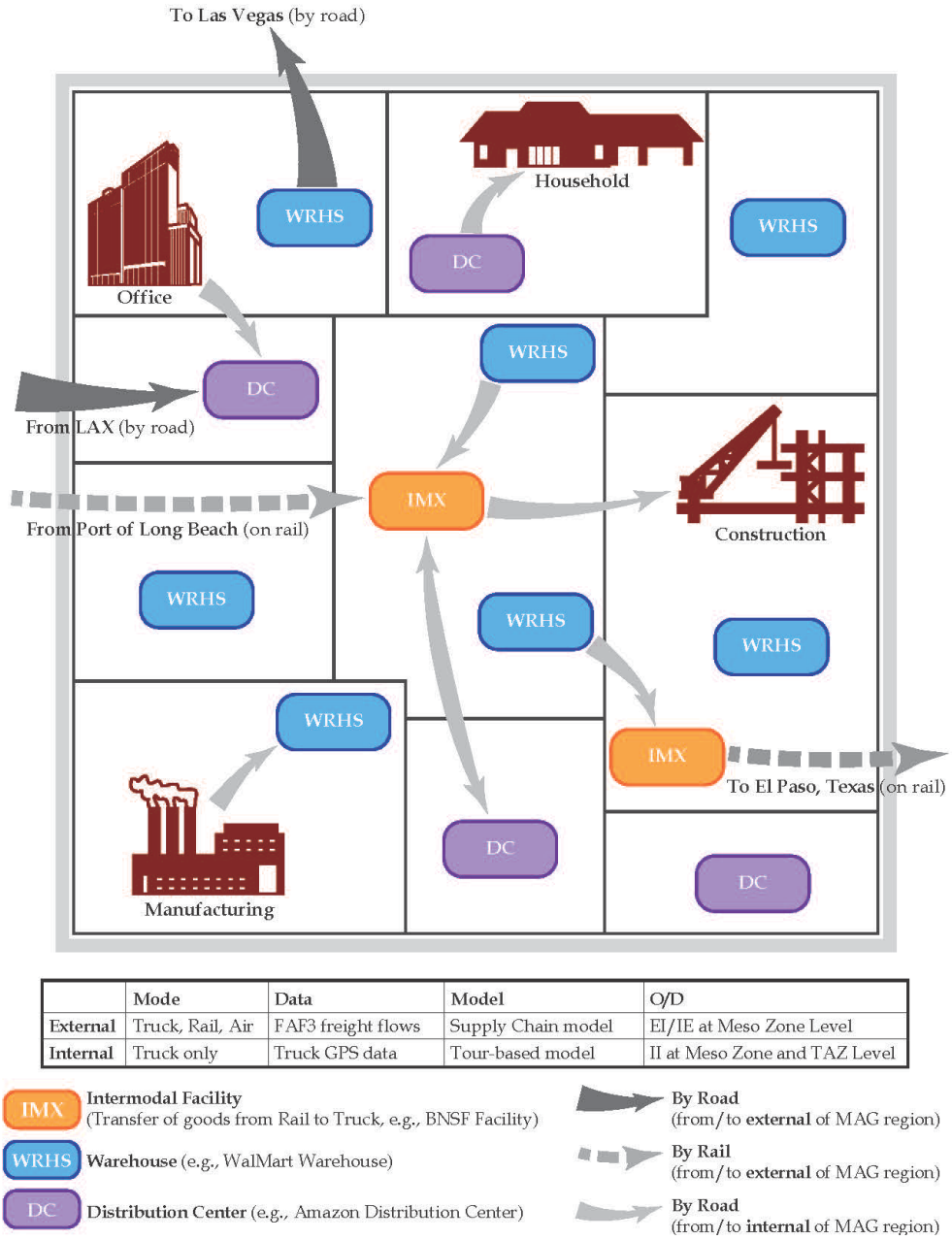
[10] Source: Cambridge Systematic, Inc.


INPUTS
MODELS
OUTPUTS
GEOGRAPHY
MODEL CALIBRATION DATA
Trucks
Commercial Vehicle Establishment Surveys
SED Data
Location of IMX Facilities (in AZ and outside AZ)
Truck Type of Vehicle Choice Model
Two types of Trucks – Medium and Heavy – will be estimated
Binary Logit Choice Model
Key Variables – Commodity, Origin Industry, Destination Industry, Employment, FTL/LTL, Direct versus IMX
Number of Medium and Heavy Trucks
Meso Zone
DMV Data
VIUS Data
Existing Truck Model
Expanded Truck GPS Data
Medium Trucks
Heavy Trucks
ATRI GPS Data for Heavy Trucks
Other (e.g., Trimble) GPS Data for Medium Trucks
SED Data
Model Skims –Times, Distances
Tour-Based Truck Model
Estimates Number of Stops, Tour Completes, Stop Purpose, Location, and TOD Choice
Series of MNL Models (very similar to MAG's Existing Truck Tour Model for Heavy Trucks)
Medium and Heavy Truck Tours, Truck Trips, and TOD Distributions
MAG TAZ
DMV Data
VIUS Data
Existing Truck Model
Expanded Truck GPS Data
Stop Generation
1 stop
2 stops
. . . . . .
11 stops
Tour Completion
Yes – return to home base
No – does not return
Stop Purpose
One of 10 stop types
• Retail Construction
• Farming
• Residential
• Government
• Warehouses
• Transportation
• Office
• Industrial
• Service
Stop Location
One of 4,144 TAZs
Stop TOD Choice
1st Stop TOD (24 1-hour periods)
Next Stop TOD (24 1-hour periods)

Study Area

The first step is the study area zone definition (within the study area and nationally), multi-modal networks that will be used in the supply chain process (Appendix B), including identification of the intermediate nodal stops (e.g., intermodal terminal, Third-Party Logistics (3PL)/truck distribution centers, etc.) on those supply chains; development of networks that connect these zones and nodes into supply chains. Figure 2-5 depicts an example of how the study area can be defined for the MAG region.

Figure 2-5 Example of Study Area [11]



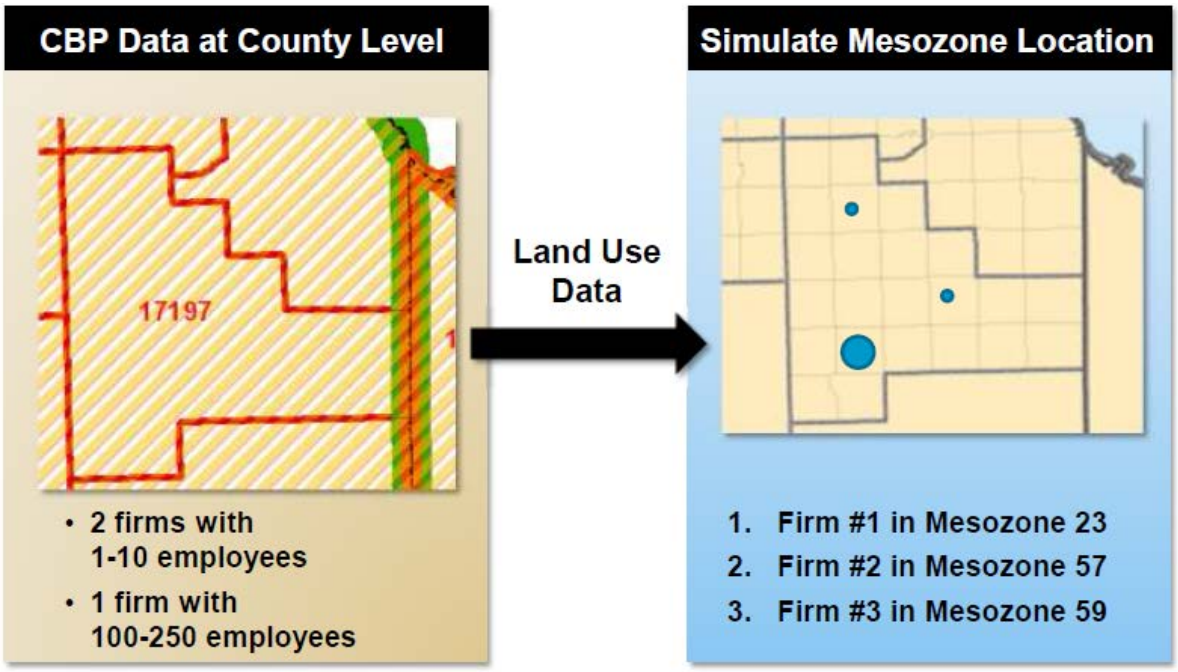
	Mode	Data	Model	O/D
External	Truck, Rail, Air	FAF3 freight flows	Supply Chain model	EI/IE at Meso Zone Level
Internal	Truck only	Truck GPS data	Tour-based model	II at Meso Zone and TAZ Level

[11] Source: Cambridge Systematics, Inc.

Firm Synthesis

The second step is the synthesis of firms. However this step, as mentioned above, was not intended to synthesize the firms in detail outside of the region. The purpose is merely to allocate the summary national supply chains (e.g., the FAF) into the separate supply chains using the Sun Corridor. For that purpose, given that it is unreasonable and unnecessary to synthesize firms in detail outside of the study region, it was deemed necessary to synthesize them in total by those parameters that are useful in allocating aggregate flows to individual supply chains (e.g., firms by industry, by size, by distance, etc.). Figure 2.6 presents an example on how the County Business Patterns (CBP) data gets simulated at the mesozone level by firm type, size, and location.

Figure 2-6 Firm Synthesis [12]



Supplier Selection

After the firms are generated or simulated in space by type, the **supplier selection model** allocates commodities from firms that produce them (suppliers) to those that buy or consume them (buyers). The approach used for this model is based on an agent-based computation economics method that captures how buyers make discrete choices about who to purchase from, and suppliers make indirect decisions about to whom to sell. A market-clearing algorithm was used in the supplier selection model (SSM) that first goes through the list of buyers and assigns its top-priority supplier. If the supplier does not have or exceeds the capacity to sell, the selecting buyer gets assigned to the next most preferred supplier, and this process is repeated until all buyers are assigned to suppliers.

Apportionment of Commodity Flows

[12] Source: Cambridge Systematics, Inc.

The next step is the application of SSM to each market and **determination of annual commodity flows** from firm to firm. The main input datasets for development and implementation of the model includes a complete list of firms (agents) in the country and the commodity Make & Use tables by industry type. These two datasets are combined to enumerate supplier and buyer firms in each market and determine their important characteristics that are influential in the decision making process of selecting suppliers and forming supply chains.

Transport and Logistics Path Choice Model

After the commodity flows are assigned to suppliers and buyers by commodity type, it needs to get assigned to a **path based on the mode of travel**. The path and mode choice are determined using a disaggregate joint model of mode choice and shipment size estimated using the Commodity Flow Survey (CFS) 2012 Microdata sample (see **Appendix A. CFS Microdata Megaregions**) as the primary source.

Tour-Based Truck Model

In order to assign the freight supply chain flows to modal networks, it is then necessary to break the supply chains into individual "links" between the stops on the supply chain. Each of the trips ("unlinked chains") will be made on a specific mode. By then, geographically aggregating the modal trips, tables of flows between origins (either the ultimate origins or an intermediate stop), and destinations (either the ultimate destination or an intermediate stop) can be constructed.

These flows might not show the "last-mile" trip to the destination as intended in the FAF from the actual local supplier to the local consumer of freight. It has been shown that the major explanatory variable for the destination of freight are the wholesale firms, who can be expected to distribute that freight locally. Therefore, in addition to the freight flows allocated from the national Origin-Destination database, to gain a complete understanding of freight movements, it is necessary to develop, as shown in the last step, **tour-based truck model**, which makes this last trip. It is probable that these tours will be made in trucks that consolidate many commodities; and that while the shipment of freight might be given between a wholesaler and his customers, the actual path of the vehicle carrying that cargo will be in a tour from that wholesaler to many suppliers in a tour until the truck is empty and returns to the wholesaler and/or takes on another load. This was aimed to be very similar to the model that one of MAG's consultants already developed for the heavy trucks.

The allocation of commodity flows to individual supply chains informs decisions about investments that might influence the choice of a supply chain. However, even if that information was not used, the information about the modal freight vehicle trips between actual intermediate stops provides invaluable information as to the volume and characteristics of freight vehicles passing through specific external stations at the boundary of a region, or at intermediate stops within the study area. This information is important to better understand vehicle movements of freight on the transportation network.

3 Review of Freight Data Sources for the Development of a Behavior-Based Freight Model

Consultant Technical Lead – Krishnan Vishwanathan (CDM Smith, Cambridge Systematics)

MAG Technical Lead – Sreevatsa Nippani, Pedro Camargo and Kyunghwi Jeon

3.1 Introduction

The main purpose of the project is to develop a new generation regional freight model based on the principles identified in the SHRP2 C20 products and consistent with MAG freight flows forecasting needs. The new freight modeling framework is envisioned to capture behavioral aspects of freight agents based on a micro-simulation approach to freight demand modeling. The new modeling paradigm would include supply chain and logistics decisions of firms including shippers, carriers and receivers. The framework is envisioned to be composed of several models, including Firm Synthesis, Supply Chain and Logistics, Input-Output flow, Truck Tour and other sub-models. The overall model should be able to provide detailed outputs by different temporal resolutions, economic sectors, industry class and other dimensions.

Robust freight data is needed to build such models and capture a range of freight movements and supply chain related choices for different commodities, types and size of firms, transportation modes, and geographies along the chain. This data can be procured from a variety of primary and secondary data sources. Sources of secondary sources such as Business Economic Area Input-Output (BEA I/O) tables, Freight Analysis Framework (FAF), Longitudinal Business Dynamics (LBD), County Business Patterns (CBP), etc. will be critical for model development.

MAG's Consultant (for the task to identify potential data sources for the project), worked with the MAG modeling team to understand the model layers and the data requirements for each layer in the initial task. A list of commercially available data as well as public databases with a detailed description of its contents was compiled. Advantages and deficiencies of each of the datasets, sources and methods utilized in compiling the datasets (if available), aggregate statistics of databases (where available), including number of records, levels of geography, exclusions from datasets, pricing (where applicable and available), contact information (where applicable) and periodicity of updates were captured.

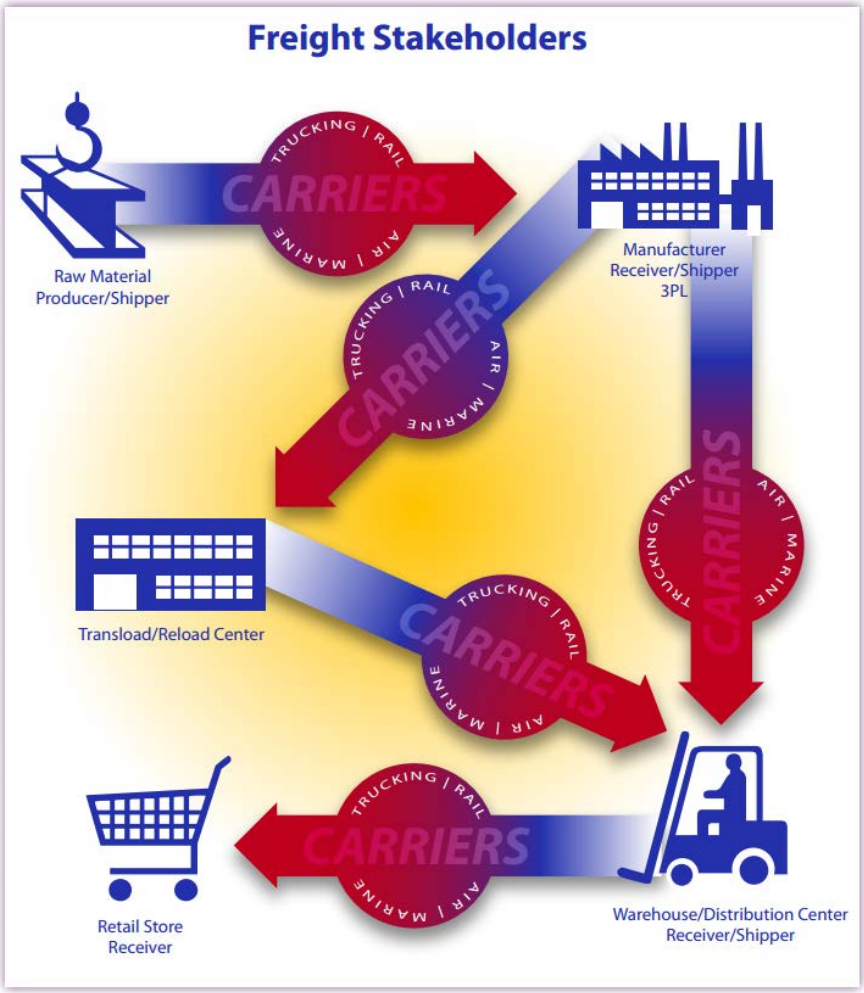
Data utilization from each source in the state of the practice and state of the art freight models was documented (whether they were candidates for estimation, calibration, and validation of models). Finally, data summary from each dataset was prepared at regional level or state level (depending upon its availability).

In order to understand and model freight flows, it is necessary to determine the various entities in the freight system (not limited to producers, shippers, carriers, forwarders and carriers, who drive freight demand). Behavioral freight models capture shippers' perception and their selection of shipment choices along each segment of a supply chain. Once the entities are identified, their mutual behavioral interactions are difficult to capture. Shipper and Carrier surveys capture behavioral data to a certain degree, but they are prohibitively expensive. Since private sector decision making is highly proprietary, understanding and modeling logistics decisions as they affect freight demand at a regional level remains a challenge for public sector agencies.

Figure 3-1 shows the supply chain and the related stakeholders. Supply Chains are highly complex, dynamic, time-sensitive and integrated systems. In order to model these integrated systems, it is necessary to capture each segment of the supply chain and the actual movement of goods by mode (truck, rail, air, etc.) along each segment. Sophistication in logistics has built upon globalization of the economy whereby materials, components and finished products are shipped worldwide. Globalization has lengthened supply lines as products may be produced offshore and can be shipped to anywhere in the world. This has resulted in more freight being carried over longer distances, with a noticeably high demand at key nodes (typically ports, airports, intermodal facilities, etc.).

Since supply chain models tend to be national and/or international in scope and are focused on modeling commodity flows, truck tour based models would model local truck movements. MAG's behavioral based freight model framework is envisioned to take a hybrid approach that combines supply chain and tour-based models. The shipments are modeled along the supply chain modeling framework and the regional pickup and delivery of these shipments are handled by truck tour models.

Figure 3-1 Supply Chain Network [13]

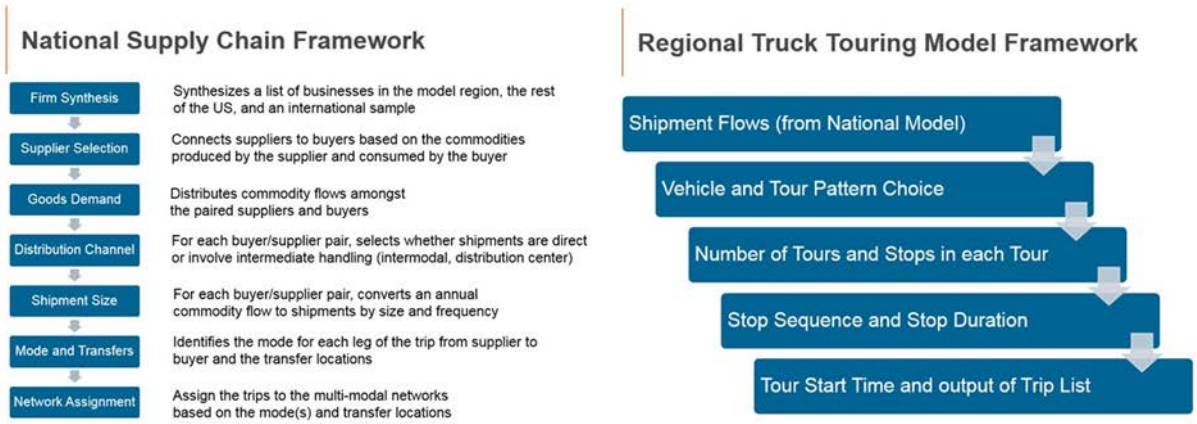


[13] Source: Arizona DOT, Multimodal Freight Analysis Study

Figure 3-2 shows the hybrid model framework. As shown in the figure, each of the components of the model requires data that may or may not be in public domain. The available and desired data (disparate datasets) can be classified into four categories:

- Free and widely available at small geographies (e.g. County Business Patterns (CBP))
- Free but restricted availability at any geography (e.g. STB Private Waybill Sample)
- Free but available only at large geographies (e.g. Freight Analysis Framework (FAF))
- Privately available for a fee (e.g. IHS TRANSEARCH)

Figure 3-2 Hybrid Model Layout [14]



In September 2014, MAG hosted a workshop to kick-off the Behavior Based Freight Model Development On-Call. From a modeling perspective, the overall consensus was to have a Financial Layer and Logistics Layer as the first two components of model hierarchy, whose results will be rolled down to Transportation Layer in the final step.

Figure 3-3 shows the interactions between the financial (production-consumption), logistics, and transportation layers along with the information layer and can help define the attributes of a behavior based freight model.

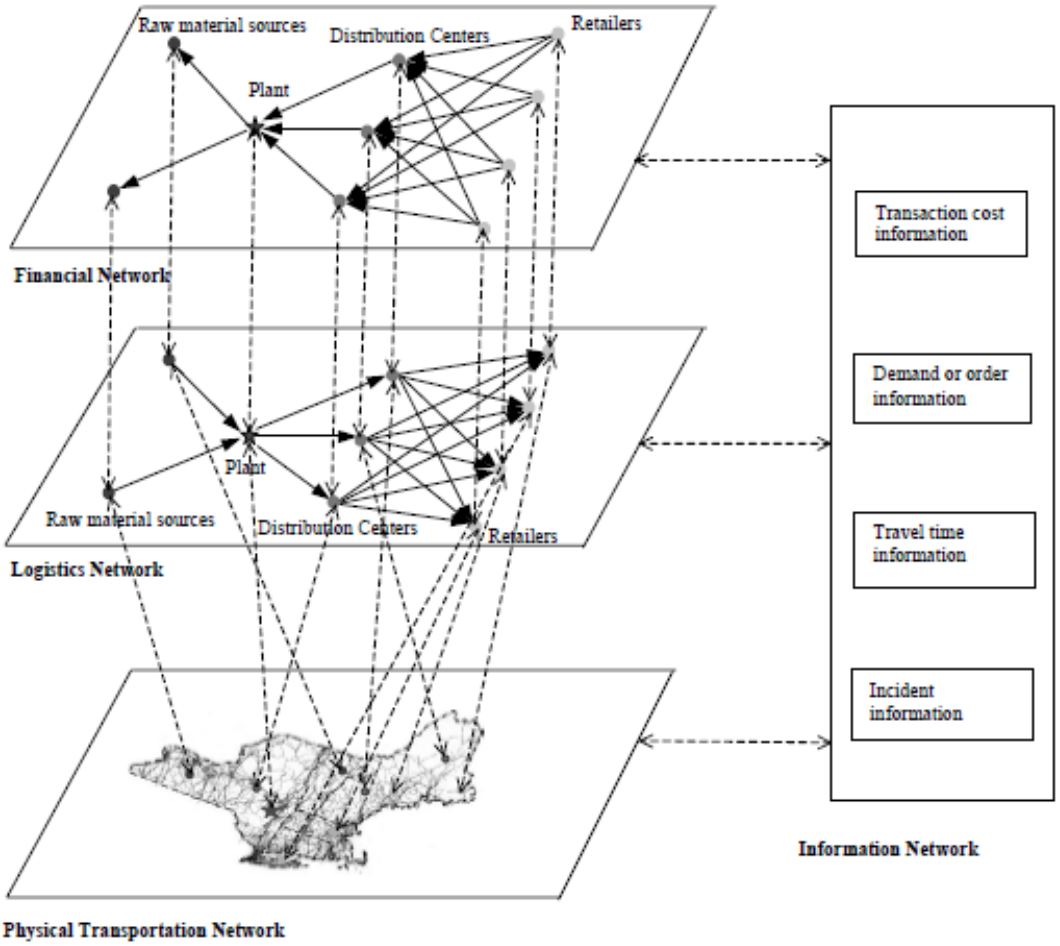
The mathematical representation of these three layers is portrayed via a series of models. These models include **firm synthesis** models which look at production, consumption, the evolution of firms, and the various factors that influence the birth, growth, location, and dissolution of firms. The second series of models focus on understanding the **logistics or supplier behavior** that help goods move efficient among the various actors of the freight story. The third series of models focuses on the **physical transportation infrastructure** on how these goods move from origin to destination, the routes taken, the modes used etc. The final series of models focuses on trucks – how they tour through a study area starting from a first pickup of the day from the depot or warehouse to the final trip back to the depot or warehouse and traverse the **logistics** and **physical transportation infrastructure** layers shown in Figure 3-3.

[14] Source: RSG

In order to develop these mathematical representations it is critical to inventory and understand the various data sources that are available and is the focus of this report.

The rest of this chapter is organized as follows. First a typology linking data types, sources, and freight models is presented. This is followed by providing metadata on each data source that helps develop firm synthesis, logistics chains, transportation chains, and truck touring models. The metadata and tables in this report identify, where relevant, data summaries in Excel. **Data summaries are identified in the field Aggregate Statistics in each of the table where it is applicable**.

Figure 3-3 Interaction between Various Networks [15]



[15] Xu, J., K. L. Hancock, and F. Southworth. "Simulation of Regional Freight Movement with Trade and Transportation Multinetworks", Transportation Research Record 1854, 2003

3.2 Data Types and Sources

Table 3-1 shows the different data types and how they can be used for freight models. Table 3-2 shows how each of the data sources presented relate to the data types identified in Table 3-1.

Table 3-1 Different Data Types for Freight Modeling

Data Types	Use in Freight Modeling
Trade Statistics	Estimation of Production-Consumption (PC) matrices for the base year
	Aggregate gravity-type models for generation and distribution at the PC level
	Value-to-weight ratios (for exported and imported goods)
National Account Data	Estimation of PC matrices for the base year
	Aggregate I/O models for generation and distribution
Transportation Statistics	Estimation of OD matrices for the base year
	Estimation of gravity-type models for generation and distribution at the OD level (less appropriate than at the PC level)
	Estimation of aggregate mode choice models
	Load factors (cargo weight to vehicle capacity)
	Models for road vehicle type choice, tour formation and empty driving/load factor if micro-data is available
Shipper surveys	Estimation of PC matrices for the base year
	Estimation of disaggregate mode choice models
	Estimation of transportation chain choice models
	Estimation of disaggregate shipment size choice models
	Estimation of disaggregate joint models (mode-shipment; mode-supplier)
	Value-to-weight ratios
Stated preference surveys	Estimation of disaggregate mode choice models
	Estimation of transportation chain choice models
	Estimation of route choice models
	Estimation of disaggregate shipment size choice models
	Estimation of disaggregate joint models (mode-shipment; mode-supplier)
	Monetary value of service attributes (e.g. value of time)
Consignment Bills and RFID data	Estimation of OD matrices for the base year (possibly PC, if tags stay on after trans-shipment or if combination of tags are registered at trans-shipment)
	Estimation of disaggregate mode choice models
	Estimation of disaggregate shipment size choice models
	Estimation of disaggregate joint models (mode-shipment; mode-supplier)
Traffic Count data	Estimation of OD matrices for the base year
	Calibration Data
Weight Data	Load factors
Network data with cost functions	Direct input for the estimation of aggregate and disaggregate mode choice models and joint models
	Indirect input for aggregate distribution models
	Direct input for the estimation of route choice models
Terminal data	Direct input for the estimation of transportation chain choice models

Table 3-2 Data Sources by Data Types

	Data Type										Spatial (Smallest Geography)	Temporal
Data Source	**Trade Statistics**	**National Account Data**	**Transportation Statistics**	**Shipper surveys**	**Stated preference surveys**	**Consignment Bills and RFID data**	**Traffic Count data**	**Weight Data**	**Network data with cost functions**	**Terminal data**		
Bureau of Economic Analysis (BEA) Input/Output Tables		✓									National	Annual
County Business Patterns (CBP)	✓										County	Annual
National Establishment Time-Series (NETS)	✓										County	Annual
Longitudinal Business Dynamics (LBD)	✓										State	Annual
Annual Survey of Manufacturers (ASM)	✓										State	Annual
Business Dynamics Statistics	✓										MSA	Annual
Business Employment Dynamics	✓										County	Quarterly
Commodity Flow Survey (CFS)			✓	✓							CSA or MSA	Every 5 years
Freight Analysis Framework (FAF)			✓								CSA or MSA	Every 5 years
Transearch			✓								County/TAZ available on demand	Annual
Surface Transportation Board (STB) Carload Waybill Sample			✓		**Project Specific**						BEA	Annual
Air Carrier Statistics			✓								Airport	Monthly
North American Transborder Freight Database			✓								State and Port of Entry/Exit	Monthly
PIERS			✓			✓				✓	Port	Annual
National Highway Planning Network (NHPN)							✓		✓		State	Unknown
National Performance Management Research Dataset (NPMRDS)							✓		✓		Traffic Message Channel	Every 5 minutes
ATRI							✓		✓		Truck Lat/Long	Second
MAG Roadway Network							✓		✓		Unknown	Unknown
Vehicle Inventory and Use Survey (VIUS)								✓			State	Every 5 years
ORNL Rail Network									✓		Unknown	Unknown
VTRIS								✓			Weight Station	Unknown
Establishment Surveys			✓	✓							Establishment	Varies by Sponsor

✓ - Applies

Table 3-3 shows the sub-models and data sources in a single table and details of each table in subsequent sections.

Table 3-3 Data Sources by Modeling Needs

Data Source	Model	Data Type	Spatial	Temporal	Mode	Commodity	Traffic Count	Data Use		
								Estimation	Calibration	Validation
Bureau of Economic Analysis (BEA) Input/Output Tables	Firm Synthesis	IO	National	Annual				✓	✗✓	✗
County Business Patterns (CBP)	Firm Synthesis	TS	County	Annual				✓	✓	✓
National Establishment Time-Series (NETS)	Firm Synthesis	TS	County	Annual				✓	✗✓	✓
Longitudinal Business Dynamics (LBD)	Firm Synthesis	TS	State	Annual				✗	✗	✓
Annual Survey of Manufacturers (ASM)	Firm Synthesis	TS	State	Annual				✗	✗	✓
Business Dynamics Statistics	Firm Synthesis	TS	MSA	Annual				✗	✓	✓
Business Employment Dynamics	Firm Synthesis	TS	County	Quarterly				✓	✓	✓
Commodity Flow Survey (CFS)	Supply Chain	TrS	CSA or MSA	Every 5 years	Truck, Rail, Air, Water, Pipeline, Other	SCTG commodities		✗	✓	✓
Freight Analysis Framework (FAF)	Supply Chain	TrS	CSA or MSA	Every 5 years	Truck, Rail, Air, Water, Pipeline, Other	SCTG commodities		✓	✓	✓
Transearch	Supply Chain	TrS	County/TAZ available on demand	Annual	Truck, Rail, Air, Water, Pipeline, Other	STCC commodities		✓	✓	✓
Surface Transportation Board (STB) Carload Waybill Sample	Supply Chain	TrS	BEA	Annual	Rail	STCC commodities		✗	✓	✓
Air Carrier Statistics	Supply Chain	TrS	Airport	Monthly	Air	None		✗	✓	✓
North American Transborder Freight Database	Supply Chain	TrS	State and Port of Entry/Exit	Monthly	Truck, Rail, Air, Water, Pipeline, Other	SITC	Yes	✗	✓	✓
PIERS	Supply Chain	TrS	Port	Annual	Water	HS		✓	✓	✓
National Highway Planning Network (NHPN)	Transportation Chain	TC, Net	State	Unknown			Yes	✓	✓	✓
National Performance Management Research Dataset (NPMRDS)	Truck Touring	TC, Net	Traffic Message Channel	Every 5 minutes	Trucks	Unknown		✓	✓	✓
ATRI	Truck Touring	TC, Net	Truck Lat/Long	Second	Trucks	Unknown		✓	✓	✓
MAG Roadway Network	Transportation Chain	TC, Net								
Vehicle Inventory and Use Survey (VIUS)	Transportation Chain	WD	State	Every 5 years	Freight Trucks and Commercial Vehicles	None		✗	✓	✓
ORNL Rail Network	Transportation Chain	TC, Net	Unknown	Unknown	Rail			✓	✓	✗✓
VTRIS	Transportation Chain	WD	Weight Station	Unknown	Freight Trucks and Commercial Vehicles	None		✗	✓	✓
Establishment Surveys	TrS, SS	Establishment	Establishment	Varies by Sponsor						

✓ - Applies ✗✓ - May apply ✗ - Does not apply

3.2.1 Data for Forecasting

A lot of the data sources presented here focus on potential application for model estimation, calibration, and validation. However, data such as TRANSEARCH, Freight Analysis Framework (FAF), and private socioeconomic data sources such as Woods & Poole provide information about future freight flows and socioeconomic conditions. Therefore, considering data coverage and limitations, careful thought must be put into how to use each of these data sources for forecasting.

3.3 Firm Synthesis

The main component of Financial Layer is Firm Synthesis. Substantial data is required to develop a firm synthesis model, which simulates regional firms by industry and by size mainly. Data is required on the industrial classification of firms, number of employees and other attributes in order to develop firm age synthesis, firm formation, location, and survival sub-models. A diverse mix of data can be used to develop firm synthesis models. These include:

- Bureau of Economic Analysis (BEA) Input/Output Tables (Table 3-4)
- County Business Patterns (CBP) data (Table 3-5)
- National Establishment Time-Series (NETS) (Table 3-6)
- Longitudinal Business Dynamics (LBD) (Table 3-7)
- Annual Survey of Manufacturers (ASM) (Table 3-8)
- Business Dynamics Statistics (BDS) (Table 3-9)
- Business Employment Dynamics (BED) (Table 3-10)
- Statistics of US Businesses (Table 3-11)
- Non Employer Statistics (Table 3-12)

The differences between some of these data sources were captured accurately by the Kauffman Foundation. From their report, "National and state-level firm or establishment birth data is captured by the Business Employer Dynamics (BED) or the Statistics of U.S. Businesses (SUSB). The BED data are compiled by the U.S. Bureau of Labor Statistics (BLS) from existing quarterly state unemployment insurance (UI) records through the Quarterly Census of Employment and Wages (QCEW) or ES-202 program. The SUSB data are collected by the U.S. Census Bureau and summarized by the U.S. Small Business Administration (Office of Advocacy). Both of these datasets include only employer firms. Employer firms represent approximately one-fourth of all firms, and many firms start with no employees. These data, therefore, are likely to lead to a substantial undercount in the rate of entrepreneurial activity, particularly for certain industries and regions. Finally, the BED and SUSB data are business-level data containing essentially no information on the owner's characteristics, while the CPS is person-level data containing very detailed information on the owner" [16].

[16] Source: Kauffman Foundation.

An important distinction to make is between firm and establishment. Here is how the Bureau of Labor Statistics (BLS) describes it: "An establishment is an economic unit that produces goods or services, usually at a single physical location, and engages in one or predominantly one activity, whereas a firm can comprise one or more establishments under common ownership by a corporate parent" [17].

Table 3-4 BEA Input/Output Table Metadata

Name	BEA Input/Output Tables
Agency/Source	Bureau of Economic Analysis
Data Type	National Accounts
Description	I/O analysis is an economic tool that measures the relationships between various industries in the economy. I/O tables show the commodity inputs that are used by each industry to produce its output, the commodities produced by each industry, and the use of commodities by final consumers.
Data Coverage	The make table shows the production of commodities by industries. The rows present the industries, and the columns display the commodities that the industries produce. The use table shows the uses of commodities by intermediate and final users. In contrast to the make table, the rows in the use table present the commodities or products, and the columns display the industries and final users that utilize them. Data available by commodity and industry. Data generally available at three levels of detail: sector (15 industry groups), summary (71 industry groups), and detail (389 industry groups). The detail tables are available only for the years the Economic Census is conducted (in the years ending with 2 and 7)
Modes of Freight	N/A
Commodities	Based on SCTG codes
Years of Data	1997 to 2013. 2013 is the most current
Format of Data	Excel Spreadsheet
Geographic Level	National
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\01_InputOutputData.xlsx
Pricing	Free
Contact Info	http://www.bea.gov/industry/io_annual.htm
Data Use	The data is used to develop production consumption equations by matching industries to commodities, thereby helping ensure the right match between commodities and industries.
Data overlaps	None
Data limitations	Not available locally. RIMS II can be used to obtain to regional I/O tables for a fee or can be obtained from the state Labor department.
Forecast	No
Model Uses	Estimation – Yes; Calibration – Maybe; Validation – No; Control Totals available – No

[17] Source: BLS.gov

Table 3-5 County Business Patterns Metadata

Name	County Business Patterns
Agency/Source	US Census Bureau
Data Type	Trade Statistics
Description	County Business Patterns (CBP) is an annual series that provides subnational economic data by industry. This series includes the number of establishments, employment during the week of March 12, first quarter payroll, and annual payroll.
Data Coverage	Statistics are available on business establishments at the U.S. level and by State, County, Metropolitan area, and ZIP code levels. Data for Puerto Rico and the Island Areas are available at the State and county equivalent levels. County Business Patterns (CBP) covers most NAICS industries excluding some (not all) crop and animal production; rail transportation; National Postal Service; pension, health, welfare, and vacation funds; trusts, estates, and agency accounts; private households; and public administration. CBP also excludes most establishments reporting government employees. Data available up to 6 digit NAICS
Modes of Freight	N/A
Commodities	N/A
Years of Data	Since 1964. 2012 is the most current year.
Format of Data	CSV
Geographic Level	County and Zip Code (from ZBP)
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\02_CBP.xlsx
Pricing	Free
Contact Info	http://www.census.gov/econ/cbp/
Data Use	The data is used to develop production consumption equations by providing employment information for generating the equations.
Data overlaps	The BLS Quarterly Census of Employment and Wages (QCEW) is a potential data source that overlaps with the CBP. Private data from InfoGroup, Woods & Poole etc. are another potential source.
Data limitations	Does not have employment information on rail transportation; National Postal Service; pension, health, welfare, and vacation funds; trusts, estates, and agency accounts; private households; and public administration. CBP also excludes most establishments reporting government employees.
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - At County and Zip Code Level but not for all industry sectors

Table 3-6 National Establishment Time-Series (NETS) Metadata

Name	National Establishment Time-Series (NETS)
Agency/Source	Walls & Associates, aggregated data from Dun & Bradstreet
Data Type	Trade Statistics
Description	Time series of business establishments. Includes following information: business name, address, contact information, headquarter linkages, related establishments in same state, years when business was active, year business started, industry classification, type of establishment, employment at location and job growth relative to peers, estimated annual sales at establishment and sales growth relative to peers, D&B credit ratings and scores, indicators such as foreign owned, import/export, minority/women owned, etc.
Data Coverage	Refined analysis of employment and business patterns; includes sole proprietors, part-time jobs and farm operations, which are not included in government data sources
Modes of Freight	N/A
Commodities	N/A
Years of Data	1990-current year. Documentation provided is for 2008; might have been slight changes to data categories, etc. since then
Format of Data	Probably would be in an Excel spreadsheet, but has lat-long data so could geocode
Geographic Level	State/MSA/County
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\03_NETS_establishments_Maricopa_Pima_Pinal_2002-2013.docx
Pricing	Unknown
Contact Info	Walls & Associates; 510-763-0641 or dwalls2@earthlink.net
Data Use	DVRPC used this data to delineate the 2010 employment centers in the region at a refined level. Data has historical information (first/last year), so can track birth and death of firms.
Data overlaps	Longitudinal Business Survey and Business Dynamic Statistics also has information about firm entry/exit, but at the aggregate level (not specific company linked to D&B number) and not as location specific (sector data for LBD and BDS is only at national level)
Data limitations	Not Free.
Forecast	No
Model Uses	Estimation – Yes; Calibration – Maybe; Validation – Yes; Control Totals Available – Yes if we assume that D&B has complete universe of establishment data

Table 3-7 Longitudinal Business Dynamics (LBD) Metadata

Name	Longitudinal Business Database
Agency/Source	U.S. Census (Center for Economic Studies)
Data Type	Trade Statistics
Description	Provides longitudinally linked data for all employer establishments contained in the Census Bureaus' business register, the Standard Statistical Establishment List (SSEL). State specific files include the following: Firm_Annual_Age_&State Firm_Annual_Entry_Exit_State_byState_1977_2012 Firm_Annual_Entry_Exit_Size&State Has information about annual age of establishment, entry and exit information by size of establishment, and entry/exit information
Data Coverage	Covers nearly all the non-farm private economy as well as some public sector activities
Modes of Freight	N/A
Commodities	N/A
Years of Data	1977-2012
Format of Data	Excel spreadsheets, have information for establishment age by initial size and size by sector and state, firm annual age and size by sector, size and state
Geographic Level	Statewide and total (nation)
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\04_LBD.xlsx
Pricing	Free
Contact Info	https://www.census.gov/ces/dataproducts/datasets/lbd.html
Data Use	Used for creation of the Business Dynamic Statistics, which provides annual measures of business dynamics for the economy and aggregated by firm and establishment characteristics.
Data overlaps	Business Dynamic Statistics, NETS also provides similar data
Data limitations	Sector level information is not available at the state level
Forecast	No
Model Uses	Estimation – No; Calibration – No; Validation – Yes; Control Totals Available – Yes to total establishments at state level but no to state level sector information.

Table 3-8 Annual Survey of Manufacturers (ASM) Metadata

Name	Annual Survey of Manufacturers
Agency/Source	Available from the U.S. Census
Data Type	Trade Statistics
Description	The survey captures manufacturing activity, products and location for public and private sectors. Provides statistics on employment, payroll, worker hours, payroll supplements, cost of materials, selected operating expenses, value added by manufacturing, capital expenditures, inventories and energy consumption. It also provides estimates of value of shipments for 1,384 classes of manufactured products.
Data Coverage	Manufacturing establishments with 1 or more paid employees or non-employers that use leased employees for manufacturing (NAICS 31-33); Arizona dataset has information broken down at 3 and 4 digit NAICS level. Statewide for other states at 2 digit level.
Modes of Freight	N/A
Commodities	N/A
Years of Data	2010 and 2011
Format of Data	Excel Spreadsheet (Arizona), have PDFs for the entire U.S. as well
Geographic Level	State (Arizona), have other states and entire U.S. available as well
Temporal Factor	Annual (except years ending in 2 and 7, then information is in Economic Census)
Aggregate Statistics	See Firm_Synthesis\05_Annual Survey of Manufacturers_2010_2013_AZ.xlsx
Pricing	N/A
Contact Info	http://www.census.gov/manufacturing/asm/
Data Use	Bureau of Economic Analysis uses ASM data to prepare annual updates of the GNP and GNP weight deflators, the annual input-output (I/O) tables, and to the GDP; Bureau of Labor Statistics uses ASM data to calculate annual productivity series, updating producer price index, etc.; State and local agencies use ASM data for economic policymaking and forecasting. Primary basis for updates to Longitudinal Research Database (LRD)
Data overlaps	None
Data limitations	None aware about
Forecast	No
Model Uses	Estimation – No; Calibration – No; Validation – Yes; Control Totals Available – Yes at State Level.

Table 3-9 Business Dynamics Statistics Metadata

Name	Business Dynamics Statistics (BDS)
Agency/Source	Available from the U.S. Census
Data Type	Trade Statistics
Description	Business Dynamics Statistics (BDS) provides for measures of net and gross job flows associated with entering, exiting, expanding, and contracting establishments including measures of job creation and destruction. Aggregate statistics are available for the nation and states, by firm characteristics and industry classification.
Data Coverage	Establishments that cover Agricultural services, forestry, and fishing, Mining, Construction, Manufacturing, Transportation and public utilities, Wholesale trade, Retail trade, Finance, insurance, and real estate, Services.
Modes of Freight	N/A
Commodities	N/A
Years of Data	1976-2012
Format of Data	Excel, CSV
Geographic Level	MSA
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\06_BusinessDynamicStatistics_MSA.xlsx
Pricing	N/A
Contact Info	http://www.census.gov/ces/dataproducts/bds/index.html
Data Use	Useful for ascertaining patterns of entrepreneurship, structural change, the gross job flows that underlie net employment change, and employment contributions by firm size and age
Data overlaps	NETS
Data limitations	Does not have self-employed, most government employees
Forecast	No
Model Uses	Estimation – No; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-10 Business Employment Dynamics Metadata

Name	Business Employment Dynamics (BED)
Agency/Source	Bureau of Labor Statistics
Data Type	Trade Statistics
Description	Business Employment Dynamics (BED) provides quarterly data on establishment openings, closings, expansions, and contractions by industry and size of firm, as well as establishment births, deaths, and survival by age, for the nation and states. BED data are generated from longitudinally linked microdata collected by the Quarterly Census of Employment and Wages (QCEW, formerly the ES-202) program.
Data Coverage	Establishments with 1 or more paid employees or non-employers that cover Agricultural services, forestry, and fishing, Mining, Construction, Manufacturing, Transportation and public utilities, Wholesale trade, Retail trade, Finance, insurance, and real estate, Services
Modes of Freight	N/A
Commodities	N/A
Years of Data	1992-Current
Format of Data	Excel, CSV
Geographic Level	County
Temporal Factor	Quarterly
Aggregate Statistics	See Firm_Synthesis\07_BED.xlsx
Pricing	N/A
Contact Info	http://www.bls.gov/bdm/
Data Use	Identify patterns of gross job creation and destruction by industry sector Track survival and identify contributions of young and old business establishments to employment growth
Data overlaps	QCEW
Data limitations	Does not have government and self-employment
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-11 Statistics of US Businesses Metadata

Name	Statistics of US Businesses
Agency/Source	US Census Bureau
Data Type	Trade Statistics
Description	The Statistics of US Businesses provides detailed annual data for U.S. business establishments by geography, industry, and enterprise size.
Data Coverage	All U.S. business establishments with paid employees. The Statistics of U.S. Businesses (SUSB) covers all NAICS industries except crop and animal production; rail transportation; National Postal Service; pension, health, welfare, and vacation funds; trusts, estates, and agency accounts; private households; and public administration. The SUSB also excludes most government employees.
Modes of Freight	N/A
Commodities	N/A
Years of Data	Since 1989. 2012 is the most current year.
Format of Data	CSV
Geographic Level	County and State
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\08_SUSB.xlsx
Pricing	Free
Contact Info	ewd.susb@census.gov
Data Use	Identify patterns of gross job creation and destruction by industry sector. Track survival and identify contributions of young and old business establishments to employment growth
Data overlaps	The data overlaps with Business Employment Dynamics and CBP (for employment).
Data limitations	Does not have establishment information on crop and animal production; rail transportation; National Postal Service; pension, health, welfare, and vacation funds; trusts, estates, and agency accounts; private households; and public administration. SUSB also excludes most establishments reporting government employees.
Forecast	No
Model Uses	Estimation – Maybe; Calibration – Yes; Validation – Yes; Control Totals Available - At County level but not for all industry sectors. At State level for all industry sectors.

Table 3-12 Non Employer Statistics Metadata

Name	Non-Employer Statistics
Agency/Source	US Census Bureau
Data Type	Trade Statistics
Description	Non-employer Statistics is an annual series that provides subnational economic data for businesses that have no paid employees and are subject to federal income tax.
Data Coverage	The data consist of the number of businesses and total receipts by industry. Most non-employers are self-employed individuals operating unincorporated businesses (known as sole proprietorships), which may or may not be the owner's principal source of income.
Modes of Freight	N/A
Commodities	N/A
Years of Data	Since 1997. 2012 is the most current year.
Format of Data	CSV
Geographic Level	County
Temporal Factor	Annual
Aggregate Statistics	See Firm_Synthesis\09_NES.xlsx
Pricing	Free
Contact Info	ewd.nonemployer.statistics@census.gov
Data Use	The data can be used to develop production consumption equations by providing employment information for generating the equations.
Data overlaps	None.
Data limitations	
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - At County level.

3.4 Logistics Chain Models

The logistics network models the supply-driven product flows from one facility to another, with freight moving from raw material sources to manufacturing plants, then on to distribution centers or warehouses or to retailers, and from there to final demand markets (i.e., customers). The objectives are to minimize total transaction costs and maximize satisfaction of the product demands involved at each stage. The following secondary data sources can help model logistics chains:

- Commodity Flow Survey (CFS) (Table 3-13)
- Freight Analysis Framework (FAF) (Table 3-14)
- TRANSEARCH (Table 3-15)
- Surface Transportation Bureau (STB) Carload Waybill Sample (Table 3-16)
- Air Carrier Statistics (Table 3-17)
- North American Trans-border Freight Database (Table 3-18)
- Port Import/Export Reporting Service (PIERS) (Table 3-19)

Table 3-13 Commodity Flow Survey (CFS) Metadata

Name	Commodity Flow Survey
Agency/Source	Available from the U.S. Census
Data Type	Transportation Statistics, Shipper Surveys
Description	The CFS data are used by policy makers and transportation planners in various federal, state, and local agencies for assessing the demand for transportation facilities and services, energy use, and safety risk and environmental concerns. Additionally, business owners, private researchers, and analysts use the CFS data for analyzing trends in the movement of goods, mapping spatial patterns of commodity and vehicle flows, forecasting demands for the movement of goods, and determining needs for associated infrastructure and equipment.
Data Coverage	The CFS covers business establishments in the following industries: Mining, Manufacturing, Wholesale trade, Select Retail and Services. The survey also covers selected auxiliary establishments (e.g., warehouses) of in-scope, multi-unit, and retail companies. Industries not covered by CFS include transportation, construction, most retail and services industries, farms, fisheries, foreign establishments, and most government-owned establishments. Shipment coverage: The CFS collects data on shipments originating from within-scope industries, including exports. Imports are not included until the point that they leave the importer's initial domestic location for shipment to another location. The survey does not cover shipments originating from business establishments located in Puerto Rico and other U.S. possessions and territories.
Modes of Freight	Truck, Rail, Air, Water, Pipeline, Other
Commodities	SCTG Commodities
Years of Data	Since 1993 and conducted in years ending in 2 and 7
Format of Data	Downloadable from American Fact Finder
Geographic Level	Consolidated Statistical Area or Metropolitan Statistical Area as defined by OMB
Temporal Factor	Once every Five years
Aggregate Statistics	See Logistics_Chain\01_2012_CFS_data.xlsx
Pricing	Free
Contact Info	http://www.census.gov/econ/cfs/
Data Use	Used to analyze trends in goods movement over time, develop models and analytical tools for policy analysis and investment decisions, forecast future demand for goods movement, analyze and map spatial patterns of commodity and vehicle flows.
Data overlaps	
Data limitations	Temporal and spatial resolution makes it difficult to account for small geography swings in the economy
Forecast	No
Model Uses	Estimation – No; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-14 Freight Analysis Framework (FAF) Metadata

Name	Freight Analysis Framework (FAF)
Agency/Source	FHWA
Data Type	Transportation Statistics
Description	FAF integrates data from a variety of sources to create a comprehensive picture of freight movement among states and major metropolitan areas by all modes of transportation.
Data Coverage	FAF provides estimates of freight measures available in FAF3 include value, tons, and domestic ton-miles by mode of transportation, for type of commodity, between and within states or the 123 domestic FAF regions, and to and from 8 foreign regions for exports and imports.
Modes of Freight	Truck, Rail, Air, Water, Pipeline, Other
Commodities	SCTG Commodities
Years of Data	Since 1997 and based on the CFS as the primary data source.
Format of Data	Access Tables
Geographic Level	Consolidated Statistical Area or Metropolitan Statistical Area as defined by OMB and rest of state
Temporal Factor	Once every Five years
Aggregate Statistics	See Logistics_Chain\02_FAF3_summary_032015.xlsx
Pricing	Free
Contact Info	http://faf.ornl.gov/fafweb/Default.aspx
Data Use	The FAF helps users understand the movement of freight between regions within the US and between the US and the rest of the world.
Data overlaps	TRANSEARCH provides similar data but uses STCC instead of SCTG codes and data is available at county level. Data can also be purchased from TRANSEARCH at the Zip code and TAZ levels.
Data limitations	Temporal and spatial resolution makes it difficult to account for small geography as well as swings in the economy
Forecast	Yes
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-15 TRANSEARCH Metadata

Name	TRANSEARCH
Agency/Source	IHS Global Insight
Data Type	Transportation Statistics
Description	TRANSEARCH is an annual database of U.S. county-level freight movement data used for freight modeling and forecasting.
Data Coverage	TRANSEARCH includes market-to-market flow data for more than 450 individual commodities and seven modes of transportation at the county level
Modes of Freight	Truck, Rail, Air, Water, Pipeline, Other
Commodities	STCC Commodities
Years of Data	Since 1995 and available yearly
Format of Data	Access database
Geographic Level	US County and BEA regions (outside of geography of interest)
Temporal Factor	Annual
Aggregate Statistics	See Logistics_Chain\03_TRANSEARCH.xlsx
Pricing	Ranges from $28,000 to $500,000
Contact Info	844-301-7334
Data Use	Used to analyze trends in goods movement over time, develop models and analytical tools for policy analysis and investment decisions, forecast future demand for goods movement, analyze and map spatial patterns of commodity and vehicle flows.
Data overlaps	FAF provides similar data but uses SCTG instead of STCC codes and data is not available at TAZ level in the standard tables but available as special tabulations
Data limitations	Methodology not very transparent
Forecast	Yes
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-16 Surface Transportation Board (STB) Carload Waybill Sample Metadata

Name	Carload Waybill Sample
Agency/Source	Surface Transportation Board
Data Type	Transportation Statistics
Description	The Carload Waybill Sample is a stratified sample of carload waybills for all U.S. rail traffic submitted by those rail carriers terminating 4,500 or more revenue carloads annually.
Data Coverage	Under statue (49 CFR Part 1244) each railroad is required to file waybill sample information for all line-haul revenue waybills terminated on its lines if it terminates at least 4,500 revenue carloads in any of the three preceding years, or if it terminates at least 5% of the revenue carloads terminating in any state in any of the three preceding years.
Modes of Freight	Rail
Commodities	STCC Commodities
Years of Data	Since 1972 and available yearly (2013 most current)
Format of Data	ASCII Files
Geographic Level	BEA Regions
Temporal Factor	Annual
Aggregate Statistics	See Logistics_Chain\04_STBWaybill2013_040615.xlsx
Pricing	Free
Contact Info	http://www.stb.dot.gov/stb/industry/econ_waybill.html
Data Use	Used to understand rail flow and for development of state transportation plans. Also used to develop TRANSEARCH rail tables.
Data overlaps	None
Data limitations	The Public Use Waybill Sample is restricted to BEA Regions. The confidential Waybill Sample data is available at the county level but is subject to approval by STB.
Forecast	Yes: MAG has 2009, 2020, 2035, 2040 and 2050 Waybill Sample data at the County level. Since the dataset is confidential, data summary cannot be provided in this report.
Model Uses	Estimation – No; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-17 Air Carrier Statistics Metadata

Name	Air Carrier Statistics
Agency/Source	US DOT Office of the Secretary – Research (OST-R)
Data Type	Transportation Statistics
Description	The Air Carrier Statistics database contains domestic and international air carrier traffic information on US and foreign air carriers having at least one point of service in the United States.
Data Coverage	Large carriers with annual operating revenues of $20 million or more
Modes of Freight	Air
Commodities	None
Years of Data	Since 1990 and available yearly (2014 most current)
Format of Data	CSV
Geographic Level	Airport
Temporal Factor	Monthly
Aggregate Statistics	See Logistics_Chain\05_AirCarrierStats_1995-2014_PHXTUS.xlsx
Pricing	Free
Contact Info	
Data Use	Used to understand air flow and for development of state transportation plans.
Data overlaps	None
Data limitations	Commodity information and value of goods is not available.
Forecast	No
Model Uses	Estimation –No; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-18 North American Trans-border Freight Database Metadata

Name	Trans-border Freight Database
Agency/Source	US DOT Office of the Secretary – Research (OST-R)
Data Type	Transportation Statistics
Description	The Trans-border Freight Database contains freight flow information by commodity type and mode for US Exports and Imports from Canada and Mexico.
Data Coverage	Beginning with the 1997 data, the North American Trans-border Freight Data represents official U.S. trade with Canada and Mexico for shipments that entered or exited the United States by surface modes of transport (other than air or maritime vessel). The data from April 1993 to December 1996 included official U.S. trade with Canada and Mexico by surface modes and trans-shipments that moved from a third country through Canada or Mexico to the United States or from the United States to a third country through Canada or Mexico. For this time period, it was not possible to separate transshipment activity from the official trade activity at the detailed level. Due to customer requests, BTS discontinued the inclusion of transshipment activity in the North American Trans-border Freight Data beginning with the January 1997 data month. This allows customers to perform comparable trade analyses by mode of transportation.
Modes of Freight	Truck, Rail, Air, Water, Pipeline, Other
Commodities	http://apps.bts.gov/programs/international/transborder/commodity.html
Years of Data	Since 1990 and available yearly (2014 most current)
Format of Data	CSV
Geographic Level	State and Customs Port
Temporal Factor	Monthly
Aggregate Statistics	Logistics_Chain\06_Transborder_Freight.xlsx
Pricing	Free
Contact Info	http://apps.bts.gov/programs/international/transborder/TBDR_QA.html
Data Use	Used to understand cross border freight movements and can help develop freight model.
Data overlaps	None
Data limitations	Commodity information not available at port level (only aggregate ports). Also only value available and weight is not available
Forecast	No
Model Uses	Estimation – No; Calibration – Yes; Validation – Yes; Control Totals Available - Yes

Table 3-19 Port Import/Export Reporting Service (PIERS) Metadata

Name	PIERS
Agency/Source	IHS PIERS
Data Type	Transportation Statistics, RFID, Terminal Data
Description	The PIERS database provides information on U.S. foreign waterborne imports and exports. The database also reports trade shipment statistics for cargo movements between ports in Mexico and South America to major trade partners around the world.
Data Coverage	Electronically filed or hard-copy freight vessel manifests submitted to the U.S. Customs Service are the source of information in the database. Vessel manifest data are collected from all foreign trade carriers at major port locations in the U.S., Mexico, Latin America, and Asia. Manual vessel manifests filed at smaller port locations are not reported in the database. Shipments are collected for the following: Containerized, Break bulk, Dry bulk, Tankers
Modes of Freight	Water
Commodities	6-digit Harmonized System Commodity classification (HS)
Years of Data	2014 Most recent
Format of Data	Excel
Geographic Level	State and Customs Port
Temporal Factor	Monthly
Aggregate Statistics	None
Pricing	$799/month for pro version
Contact Info	Online contact form (https://www.piers.com/ContactUs)
Data Use	Used to understand freight manifest, shipper and consignee party names (useful for supply chain modeling)
Data overlaps	US Census Bureau Foreign Trade Statistics has some overlaps
Data limitations	Cost
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - Unknown

3.5 Transportation Chain Data

The physical transportation infrastructure as goods and services move through the network from manufacturer/distributor to the final customer is manifested via the transportation chain as goods move from different modes and intermodal centers. Information such as distance, travel time, and capacity is needed for each link along the transportation chain, allowing intermodal as well as single-mode routing of freight traffic. The following sources are available for representing the transportation network:

- National Highway Performance Network (NHPN) (Table 3-20)
- ORNL Rail Network (Table 3-21)
- Vehicle Inventory and Use Survey (VIUS) (Table 3-22)
- Vehicle Travel Information System (VTRIS) (Table 3-23)
- MAG Roadway Network

Table 3-20 National Highway Planning Network (NHPN) Metadata

Name	National Highway Planning Network (NHPN)
Agency/Source	FHWA
Data Type	Traffic Count Data & Network Data with Cost Functions
Description	The NHPN is a geospatial network database that contains line features representing more than 450,000 miles of highways in the United States.
Data Coverage	This database covers all principal arterials and rural minor arterials.
Modes of Freight	Truck
Commodities	N/A
Years of Data	2005 to Current
Format of Data	Shape file
Geographic Level	State
Temporal Factor	Unknown
Aggregate Statistics	See Transportation_Chain\01_NHPN.xlsx
Pricing	Free
Contact Info	http://www.fhwa.dot.gov/planning/processes/tools/nhpn/
Data Use	Used as base network for freight modeling applications. Also can be used for validation of models.
Data overlaps	MAG Roadway network
Data limitations	Does not cover every facility type. Also, the AADT information is not updated more frequently.
Forecast	Yes
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available - No

Table 3-21 ORNL Railroad Network Metadata

Name	CTA Railroad Network
Agency/Source	CTA, ORNL
Data Type	Network Data with Cost Functions
Description	The CTA Railroad Network is a representation of the North American railroad system that contains every railroad route in the US, Canada, and Mexico that has been active since 1993. The network is an extension of the Federal Railroad Administration's strategic network.
Data Coverage	This database covers every railroad route in the US, Canada, and Mexico.
Modes of Freight	Rail
Commodities	N/A
Years of Data	1993 to Current
Format of Data	Shape file
Geographic Level	Unknown
Temporal Factor	Unknown
Aggregate Statistics	See Transportation_Chain\02_ORNL_Rail.xlsx
Pricing	Free
Contact Info	http://www-cta.ornl.gov/transnet/RailRoads.html (Bruce Peterson, 865-946-1352, PetersonBE@ornl.gov)
Data Use	The data can be used for network analysis and developing railroad skims.
Data overlaps	None
Data limitations	
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Maybe; Control Totals Available - No

Table 3-22 Vehicle Inventory and Use Survey (VIUS) Metadata

Name	VIUS
Agency/Source	US Census Bureau
Data Type	Weight Data
Description	VIUS provides physical and operational characteristic data of nationwide private and commercial truck fleet. The physical characteristic data include weight, number of axles, overall length, body type, etc. for medium and heavy trucks. The operational characteristic data include commodities handled, distance traveled, mileage, etc.
Data Coverage	The VIUS excludes vehicles owned by Federal, state, or local governments; ambulances; buses; motor homes; farm tractors; and non-powered trailer units. Additionally, trucks that were included in the sample but reported to have been sold, junked, or wrecked prior to the survey year (date varies) were deemed out-of-scope.
Modes of Freight	Freight Trucks and Commercial Vehicles
Commodities	N/A
Years of Data	1997, 2002
Format of Data	Access
Geographic Level	State
Temporal Factor	The VIUS has not been updated since 2002 and current plans to develop a new version are unknown.
Aggregate Statistics	None
Pricing	Free
Contact Info	http://www.census.gov/svsd/www/vius/2002.html
Data Use	Used for development of truck payload factors.
Data overlaps	VTRIS
Data limitations	Lack of more current data. FHWA has plans to develop a new VIUS database.
Forecast	No
Model Uses	Estimation – No; Calibration – Yes; Validation – Yes; Control Totals Available - No

Table 3-23 Vehicle Travel Information System (VTRIS) Metadata

Name	VTRIS
Agency/Source	FHWA
Data Type	Weight Data
Description	VTRIS provides information on vehicle travel characteristics.
Data Coverage	Unknown.
Modes of Freight	Freight Trucks and Commercial Vehicles
Commodities	N/A
Years of Data	Unknown
Format of Data	Stand Alone Executable
Geographic Level	Weight Station
Temporal Factor	Unknown
Aggregate Statistics	None
Pricing	Free
Contact Info	http://www.fhwa.dot.gov/ohim/ohimvtis.cfm
Data Use	Used for development of truck payload factors.
Data overlaps	VTRIS
Data limitations	More information is not available.
Forecast	No
Model Uses	Estimation – No; Calibration – Yes; Validation – Yes; Control Totals Available - No

3.6 Truck Touring Models

Given the preponderance of trucks as a major contributor of freight, it is necessary to develop truck tour models to determine the impacts trucks have on congestion, VMT, air quality etc. Therefore it is necessary to develop truck touring models as part of behavior based freight models. Two data sources that are available include:

- ATRI Truck GPS Data (Table 3-24)
- National Performance Management Research Dataset (NPMRDS) (Table 3-25)

Table 3-24 ATRI Truck GPS Data Metadata

Name	ATRI Truck GPS Data
Agency/Source	ATRI
Data Type	Traffic Count and Network Data
Description	The American Transportation Research Institute (ATRI) collects truck position data throughout the U.S. and North America from a large sample of trucks that use onboard, wireless communications systems.
Data Coverage	Each record contains truck id, geographic information at lat/long level, and temporal information regarding location of a truck.
Modes of Freight	Freight Trucks
Commodities	N/A
Years of Data	2005 to Current
Format of Data	CSV Files
Geographic Level	Individual Truck Location
Temporal Factor	Every second
Aggregate Statistics	None
Pricing	Contact ATRI
Contact Info	Dan Murray 651-641-6162
Data Use	Used for development of truck touring models and for developing speed and travel time profiles.
Data overlaps	NPMRDS
Data limitations	Sample of trucks so need to develop truck expansion factors.
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available – No but expansion methods are being developed using ODME (see Pinjari FDOT Report16F [18])

[18] http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PL/FDOT-BDK84-977-20-rpt.pdf

Table 3-25 National Performance Management Research Dataset (NPMRDS) Metadata

Name	NPMRDS
Agency/Source	FHWA
Data Type	Traffic Count and Network Data
Description	Vehicle Probe Data available for passengers and freight. ATRI data makes up the freight data.
Data Coverage	Each record contains truck id, geographic information at lat/long level, and temporal information regarding location of a truck.
Modes of Freight	Freight Trucks
Commodities	N/A
Years of Data	2013 to current
Format of Data	CSV Files
Geographic Level	Traffic Message Channel
Temporal Factor	Every 5 minutes
Aggregate Statistics	None
Pricing	Free
Contact Info	Jeff Purdy 202-366-6993 Jeffrey.Purdy@dot.gov Vince Mantero 202-366-2997 Vince.Mantero@dot.gov
Data Use	Used for development of truck touring models and for developing speed and travel time profiles.
Data overlaps	ATRI Truck GPS
Data limitations	Sample of trucks so need to develop truck expansion factors.
Forecast	No
Model Uses	Estimation – Yes; Calibration – Yes; Validation – Yes; Control Totals Available – No

3.7 Other Data Sources

The other data sources that are relevant to MAG for the development of the Behavioral Freight Model include:

- Inland Ports Foreign Trade Zones: Increasingly, inbound cargo is transferred directly from an ocean carrier to railcars and then transported to an inland location, away from the more congested port itself, for further processing and distribution. These inland locations or intermodal centers serve as inland ports. Details about Arizona Inland Port Foreign Trade Zones can be found in (Other_Data\Inland Ports_Free Trade Zones\FTZ_InlandPorts.xlsx).
- Land Use Information: The MAG demographics office has compiled information about land use in the state and a summary of land use information of Maricopa, Pima, and Pinal Counties is provided in (Other_Data\Land_Use\CAAG_MAG_PAG_LU.xlsx).
- Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW): The Quarterly Census of Employment and Wages Program is a cooperative program involving the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor and the State Employment Security Agencies (SESAs). The QCEW program produces a comprehensive tabulation of employment and wage information for workers covered by State unemployment insurance (UI) laws and Federal workers covered by the Unemployment Compensation for Federal Employees (UCFE) program.
- Publicly available files include data on the number of establishments, monthly employment, and quarterly wages, by NAICS industry, by county, by ownership sector, for the entire United States. These data are aggregated to annual levels, to higher industry levels (NAICS industry groups, sectors, and super-sectors), and to higher geographic levels (national, State, and Metropolitan Statistical Area (MSA)). The QCEW program serves as a near census of monthly employment and quarterly wage information by 6-digit NAICS industry at the national, State, and county levels. At the national level, the QCEW program publishes employment and wage data for nearly every NAICS industry.
- At the State and area level, the QCEW program publishes employment and wage data down to the 6-digit NAICS industry level, if disclosure restrictions are met. In accordance with BLS policy, data provided to the Bureau in confidence are not published and are used only for specified statistical purposes. BLS withholds publication of UI-covered employment and wage data for any industry level when necessary to protect the identity of cooperating employers. Totals at the industry level for the States and the Nation include the non-disclosable data suppressed within the detailed tables. However, these totals cannot be used to reveal the suppressed data. Summary of the QCEW data for the state of Arizona as well as Maricopa, Pima, and Pinal Counties is provided in (Other_Data\QCEW\QCEW.xlsx).
- MAG staff has collected information on Warehouses, Motor Carriers, Air Freight Carriers, Contract Logistics Companies, International Freight Forward Custom Brokers, Railroads and Airports for the region from the MAG Employer database. This table is provided in (Other_Data\Freight Related Employers).

3.8 Summary

This chapter discussed the various data sources that are available and/or needed by MAG to develop a behavior based freight model. From the overview, many public and private data sources are available to MAG to develop a behavior based freight model. However, based on the preceding sections there are gaps in the data in terms of geography, model needs, and mode. The following sections provide a summary of the data gaps.

3.8.1 Geography Gaps

Commodity flows across the nation are very aggregate and values and volumes of commodities by corridor also are difficult to create from available data, except at a very aggregate level. Some of the current data bases are synthesized or inputted from other not directly correlated data sets. The importance of intermodal locations and rail/highway transfer points cannot be readily determined with current data. Published data, whether rail, air or highway, do not include the modal travel times. Such information is critical in fully estimating competitive demand for transportation by mode, corridor or commodity/industry and is an existing gap.

At the state and local level, knowing local distribution, retail and warehousing movements are a major gap when examining data needs. Little is known about such movements, yet they account for a large portion of the VMT. Trip patterns, weights, vehicle configuration, and routing are not systematically collected and travel times, routing, number of stops, weights, etc. are needed.

3.8.2 Model Needs Gaps

Developing a behavior based freight model requires intimate knowledge of the commodities/products being moved, the markets and origins for those commodities, the demand nodes for the commodities and the cost and market share by modes of the overall transportation system. The datasets identified above fill some of the gaps, but cost information is a gap (and the toughest to obtain) that needs to be filled so that decision making by freight stakeholders are reflected properly in the model.

3.8.3 Mode Gaps

The data gaps dealing with the origin, destination, shipper, receiver, intermodal nodes, weight, value, commodity, etc. are all applicable to each of the modes. The data sources for the differing modes do vary somewhat in data detail and completeness. Air suffers the most as far as value of goods transported is concerned, while rail analysis needs more information on origin and destination to be made available in the public domain. Local wholesale, and retail distribution along with information on commercial vehicles is a missing data gap w.r.t. trucks.

4 Analysis of Datasets used in the Model Development

Consultant Technical Lead – Arun Kuppam (Cambridge Systematics)

MAG Technical Lead – Shuyao Hong, Pedro Camargo, and Sreevatsa Nippani

Contributors from ATRI – Dan Murray and Jeff Short

Based upon the data needs for each of the sub-models, there were specific datasets that were identified to develop these sub-models. MAG acquired or utilized the following datasets to develop various sub-models:

- National Establishment Time Series (NETS) Data
- American Transportation Research Institute (ATRI) Truck GPS data
- StreetLight Truck GPS data
- IHS Global Insight's TRANSEARCH data
- Freight Analysis Framework (FAF 4.1) data

In this Chapter, each of these datasets is described briefly, besides providing a snapshot summary.

4.1 National Establishment Time Series (NETS) Data

The first step in the supply chain modeling process is to develop a Firm Synthesizer (refer to Chapter 5) that simulates firms in the region by industry/sector type, employee size and location. Various Firm lifecycle events like Firm Birth, Firm Relocation (within the modeling region, from the modeling region to outside of modeling region, from outside of modeling region to within the modeling region) and Firm Dissolution were sought to be modeled. Firms outside of the MAG region were not included as part of Firm Synthesis.

To be able to develop Firm Synthesizer model, obtaining panel-data for Firms was critical. National Establishment Time-Series (NETS) data was determined to be the best source to develop the model, as it covers more businesses than traditional sources like ES-202 database. Moreover, the database has been demonstrated to be the best to cover establishments with employee size 1-4. Developed by Walls and Associates in association with Dun and Bradstreet (D&B), the database has twenty-three annual snapshots (taken every January) of the full Duns Marketing Information (DMI) file consisting over 52.4 million record between January 1990 and January 2012. The time series data consists of over 350 fields in the following categories [19, 20, 21]

[19] NETS Database Description 2013 Revised
[20] Walls, D. (2007). National Establishment Time Series Database: Data Overview. Presented at 2007 Kauffman Symposium on Entrepreneurship and Innovation Data.
[21] Neumark, D., Wall, B. and Zhang, J. (2005). Business establishment dynamics and employment growth. Entrepreneurship Seminar Series. Hudson Institute Center for Employment Policy.

- Business name, address and contact information (including officer, title, phone number, CBSA codes and longitude and latitude).
- Headquarters linkages (including the D-U-N-S Number of the topmost domestic firm in a "Family Tree" of companies, as well as the parent company and headquarters; number of establishments reporting to a headquarters; and whether the ownership has changed 1990-present).
- Number of establishments related to each establishment in its final year.
- Years when business was active ("1989" is earliest year in Database and, currently, "2011" is the latest year) and year business started.
- Industry classification (primary SIC, two secondary SICs, and whether the primary market changed 1990-present).
- Type of establishment (Single location, headquarters, or branch; public or private; legal status: proprietorship, partnership, corporation or non-profit and "cottage" businesses).
- Employment at location and job growth relative to peers (3-digit SIC).
- Estimated annual sales at the establishment and its sales growth relative to peers.
- D&B credit ratings and PayDex Scores (January, minimum and maximum for previous year).
- Special indicators: foreign-owned, import/export, government contracts, minority-owned, women-owned, and gender of officer.
- Relocation information (origin and destination of significant moves, employment and sales in move year, origin and destination latitude-longitudes, distance of move and whether the establishment moved more than once 1990-2011).

NETS database was purchased by MAG, paving way to better understand the spatial dynamics of megaregional economic activity. As stated previously, the NETS database has records on individual business establishments for the entire United States. It consists of firm-related characteristics like year of establishment, employment size, and business type along with location and relocation information.

The key analyses that was envisioned to be carried out in this task included spatio-temporal analysis of firm formation and dissolution by considering determinants like firm internal attributes (size, age, growth), market area characteristics, agglomeration economies (diversity and availability of skilled labor), regional accessibility and transportation access, and network structure and topology as locational factors in the form of network centrality indices. Key business establishment events were analyzed over time that include establishment formation (or birth) and closure (or death), including business establishment relocation, forming the critical components of life-cycle events of business establishments. Results from this analysis were plotted to examine trends, outliers and evolving distribution of business establishments.

The various analyses that were performed using the data were as follows –

1. Tracking Firm Life Cycle changes from 1991 to 2012.
2. Tracking the Distribution of Firm Births, Dissolutions and Relocations by Industry Types from 1991 to 2012.
3. Development of Firm Growth Model, Firm Relocation Mode and Firm Dissolution Model, using 80 percent of NETS data for mega-region.
4. Model validation using 20 percent of NETS data for mega-region.

Each of these analyses will be presented in detail, in Chapter 5.

4.2 American Transportation Research Institute (ATRI) Truck GPS data

4.2.1 Overview

Traditionally data used for truck trip distribution and touring models is sourced by collecting information directly from truck drivers – often through truck trip diaries or surveys. Over the past 5 to 10 years, however, truck GPS data has emerged as an alternative to such methods. Truck GPS data offers a far more robust look at truck movement within a region while at the same time providing a substantially lower cost per trip for the data. In 2012 MAG took the step of acquiring and analyzing truck GPS data from the American Transportation Research Institute (ATRI) as part of its Truck Travel Model Update. In 2015, a larger dataset was acquired by MAG from ATRI to update the MAG truck tour model.

4.2.2 Data Collection and Coverage

The trucking industry employs telematics to manage and track its operations. The most useful aspect of these data to the industry is to manage operations in real-time. The data, however, have a secondary use to industry-trusted organizations such as ATRI, through its unique access to trucking industry data, aggregates and anonymizes information from thousands of trucking companies. ATRI is often called upon to analyze these data in support of state, local and regional transportation planning efforts.

Truck GPS data within an extended mega-regional modeling area (as shown in Figure 4-1) was collected from ATRI's global database for a two week period in April 2014 and October 2014, as well as a one week period in January 2014 and July 2014 (see Figure 4-2 for the temporal coverage). It should be noted that the modeling area has boundaries beyond Maricopa County – and includes Pinal County and sections of Gila, Pima and Yavapai counties.

Figure 4-1 Mega-regional Modeling Area (green) and ATRI Data Study Area (green + blue)

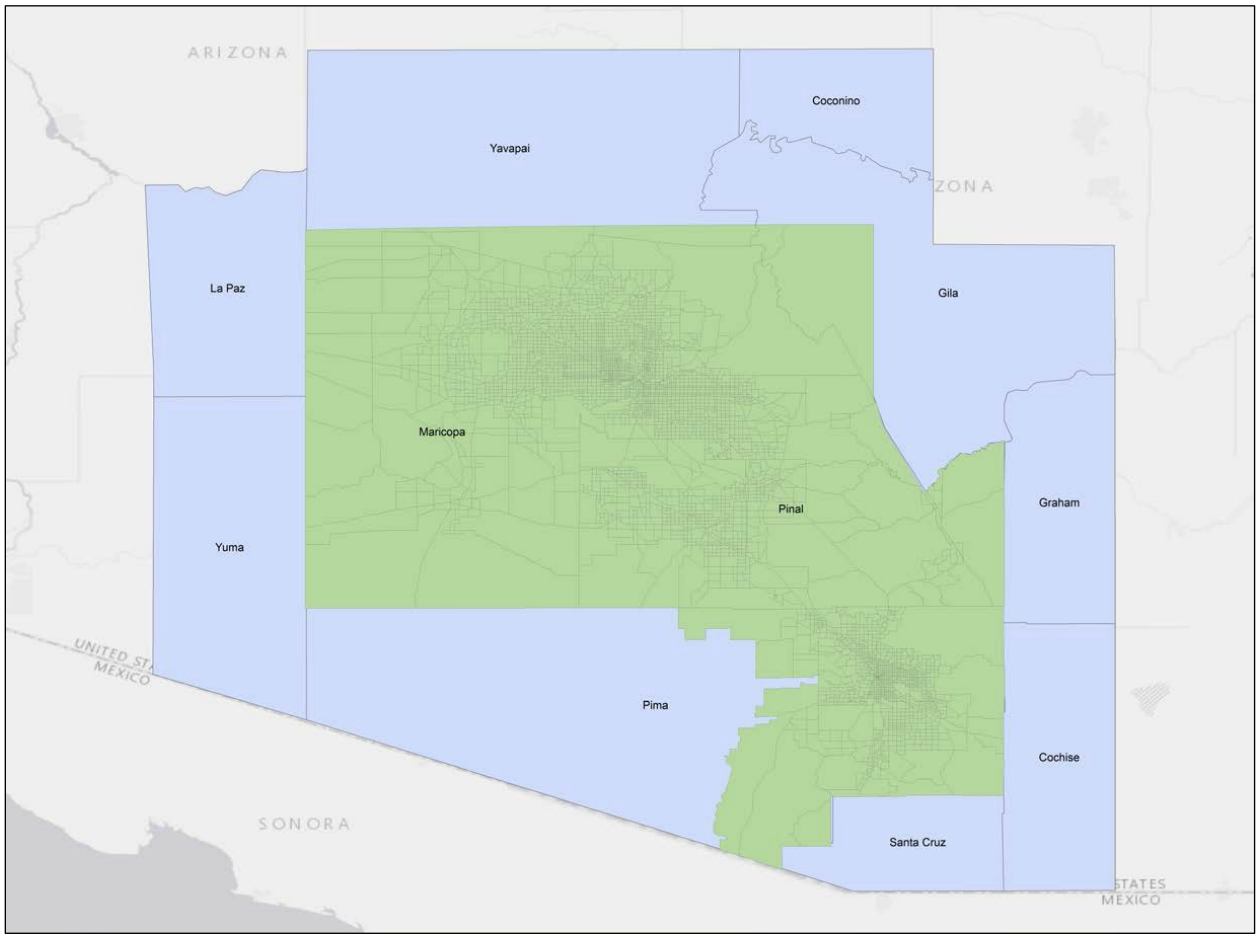
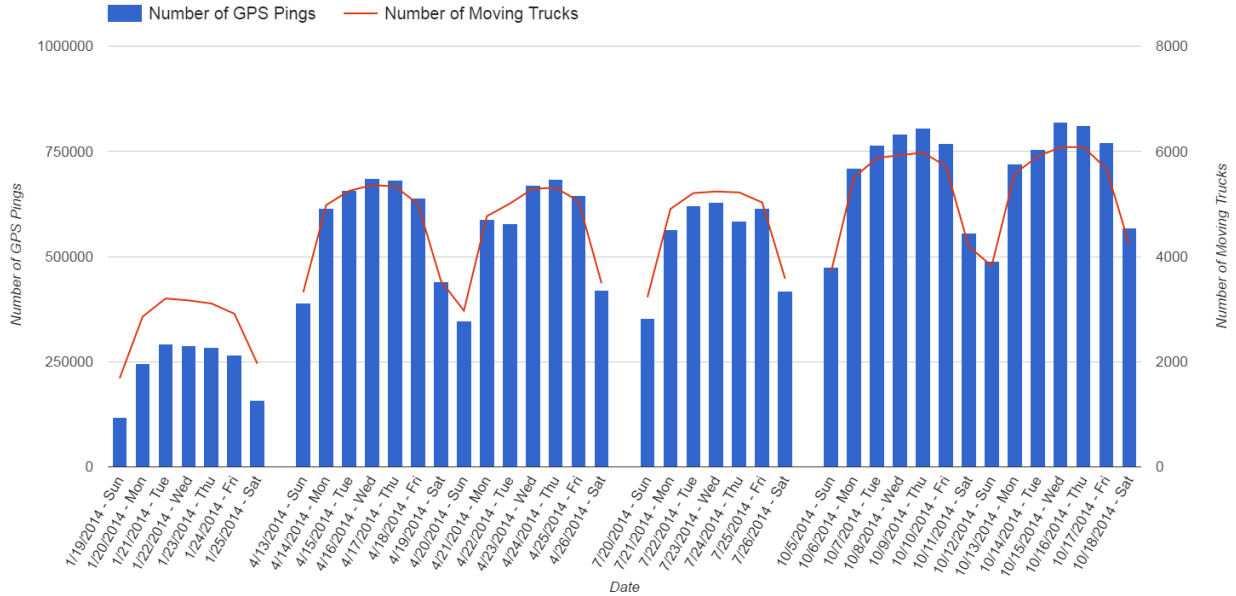


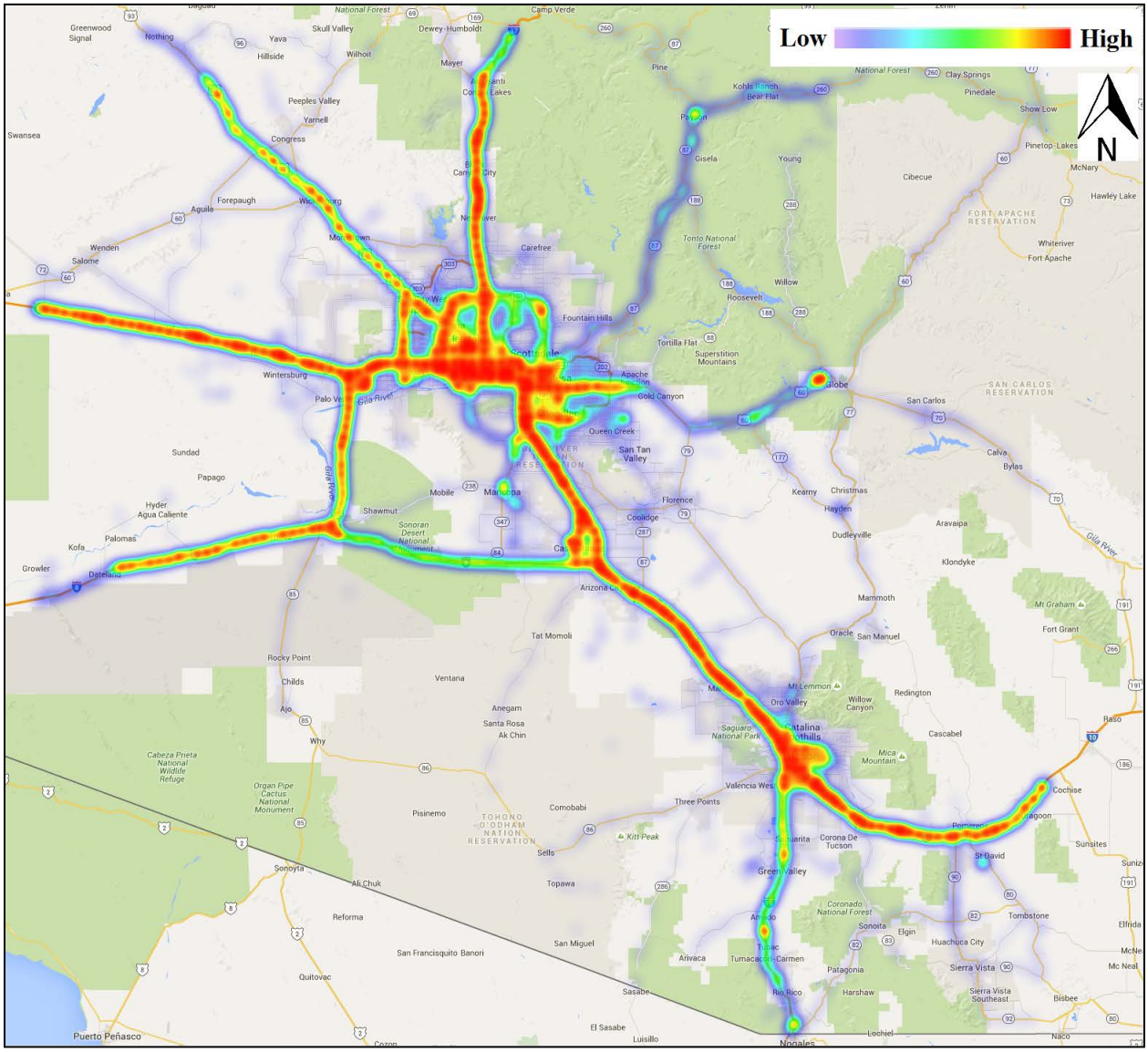
Figure 4-2 Temporal Coverage of ATRI Data



4.2.3 Data Transformations and Processing

Timestamps of all data captured were converted to Arizona's Mountain Standard Time. A dataset of truck position points within the appropriate time periods was first identified using a simple bounding box around the Figure 4-1 area. Next, the subset of data was refined further to include only position points within the Figure 4-1 area based on the x/y coordinates using a map matching procedure. Figure 4-3 depicts the spatial concentration of GPS pings in ATRI data within the study area.

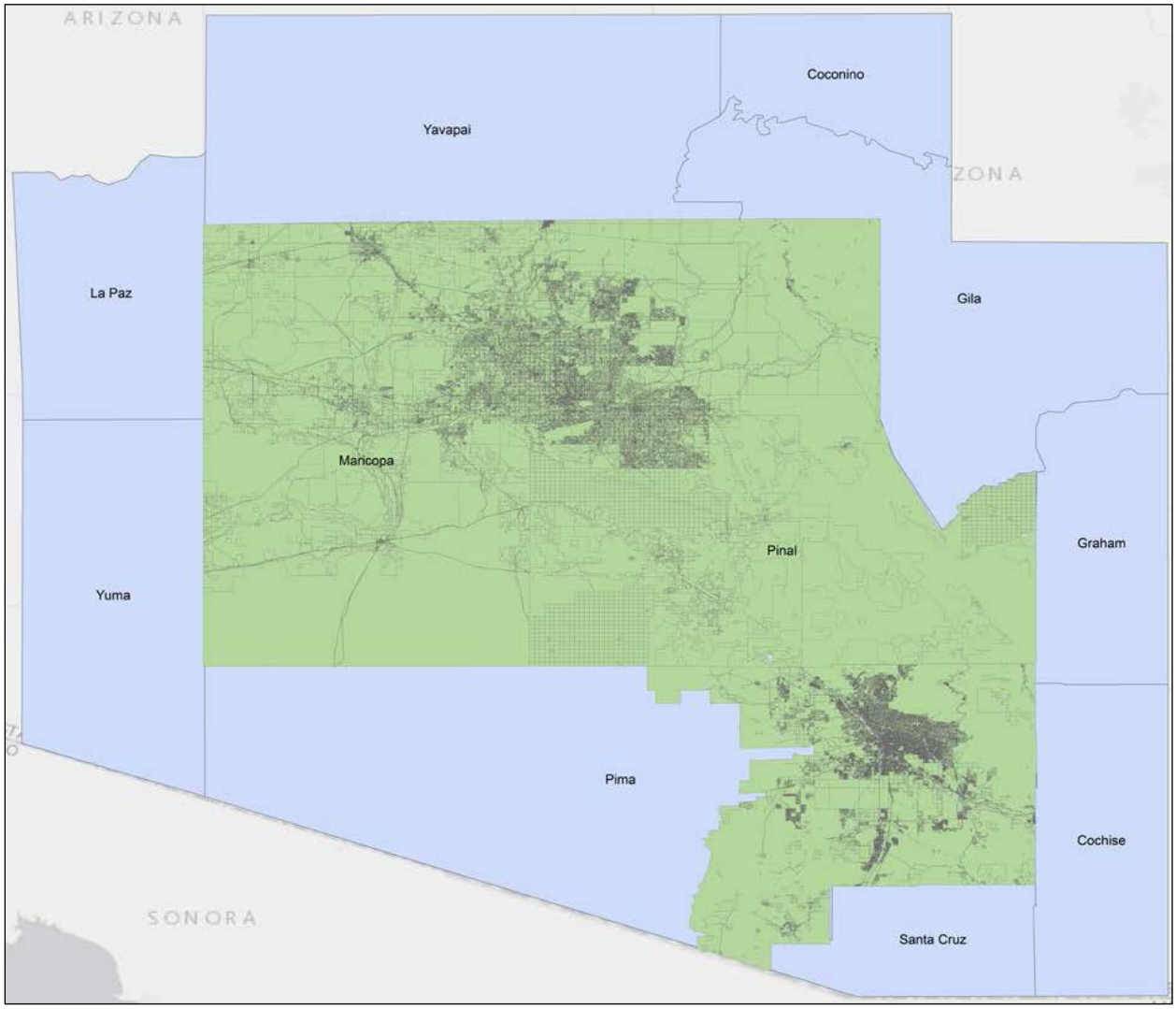
Figure 4-3 Heat Map of GPS Pings in ATRI Data



The data were then joined with traffic analysis zones (TAZ) and Land Use (LU) shapefiles (see Figure 4-4) that were provided by MAG and refined/improved by ATRI. Through the use of these shapefiles each data point within the study area was assigned TAZ and LU values. This was accomplished using the special join tool in ESRI's ArcGIS software package. TAZ data was first joined to the truck GPS dataset. Shapefile preparation was required for this step as there was overlap among the county TAZ files. Land use across the five county segments included in the study area was not standardized across the original files, therefore a crosswalk table between land use and stop purposes were developed. The standardized LU categories (i.e. stop purposes) included the following:

- Construction
- Farming
- Government
- Households
- Industrial
- Warehousing
- Office
- Retail
- Service/Other
- Transportation

Figure 4-4 Land Use Shapefile in the Modeling Area



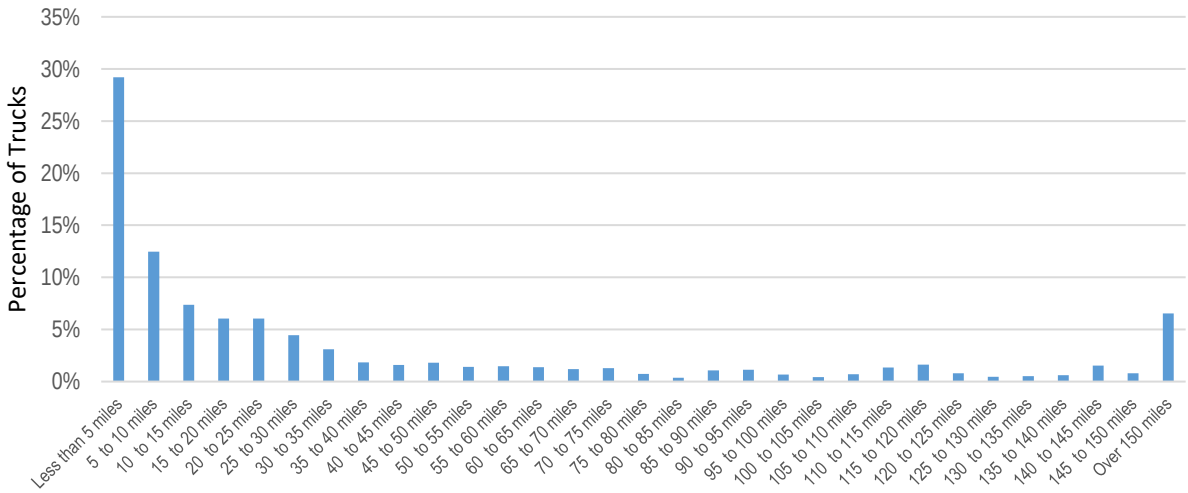
The truck GPS data was next processed into truck trips to create an LU and TAZ origin/destination matrix. Using the date/time stamp the amount of time between pings was calculated for each observation. Using the amount of time between pings, and the distance between pings (based on the longitude/latitude data) an average speed between pings was calculated. In instances when the speed drops below five miles per hour (MPH) the truck is considered to be stopped, though this does not automatically mean the truck has reached a destination.

Once a truck is considered to be stopped, the amount of time stopped begins to aggregate based on the previously calculated amount of time between pings. The stop is ignored if the truck begins moving before the amount of time stopped reaches 5 minutes or if the truck moves more than 0.5 miles away from the geographic center of stop pings. If the amount of time stopped exceeds 240 minutes (four hours), the stop is considered as the end of that particular trip.

A stop table was produced with a unique truck identifier, stop date/time stamp, coordinates of stop location, stop duration, and associated TAZ and Land Use values. Other outputs of this process include

TAZ-level truck trip matrix and Stop Purpose-to-Stop Purpose truck trip matrices. Several trip profiles were assembled based on the trip matrices. The first of these is trip length distribution as displayed in Figure 4-5. Not surprising due to the size of the region, the majority of trips were short movements.

Figure 4-5 Distribution of Truck Trip Lengths (ATRI Data)



Next examination was the distribution of stop duration and stops per truck. Figure 4-6 depicts the distribution of truck tours by stop duration, where the average stop duration is 42 minutes. Figure 4-7 shows the distribution of truck tours by stops per truck. The average is 3.39 stops per truck for all tours combined. However, there is a large variation in this metric across different tour purposes as shown in Figure 4-8. Retail, warehouse and industrial (or manufacturing) tour types are by far generating most number of stops per truck. This is seen by the curves going up the left side of the cube below while the rest of the tour types have pretty flat curves indicating a fewer number of stops per truck.

Figure 4-6 Distribution of Truck Stop Duration (ATRI Data)

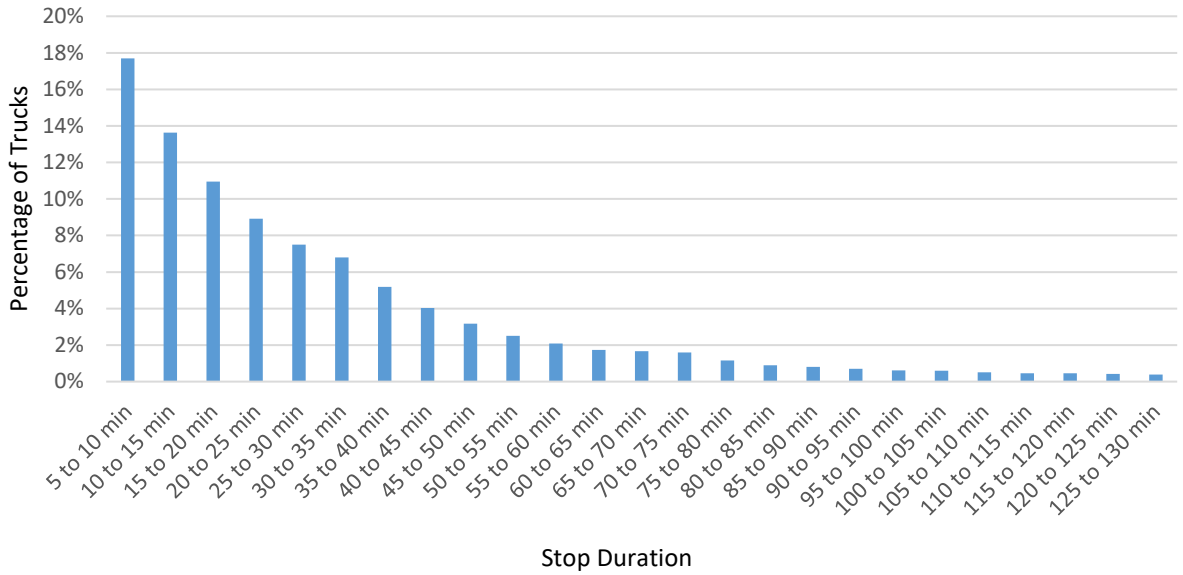


Figure 4-7 Distribution of Stops per Truck based on ATRI Data

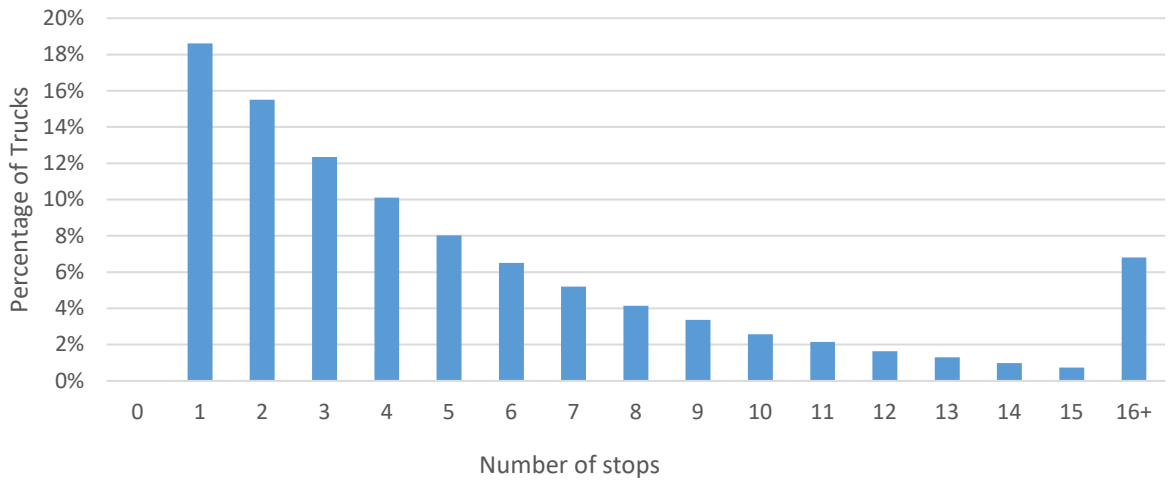
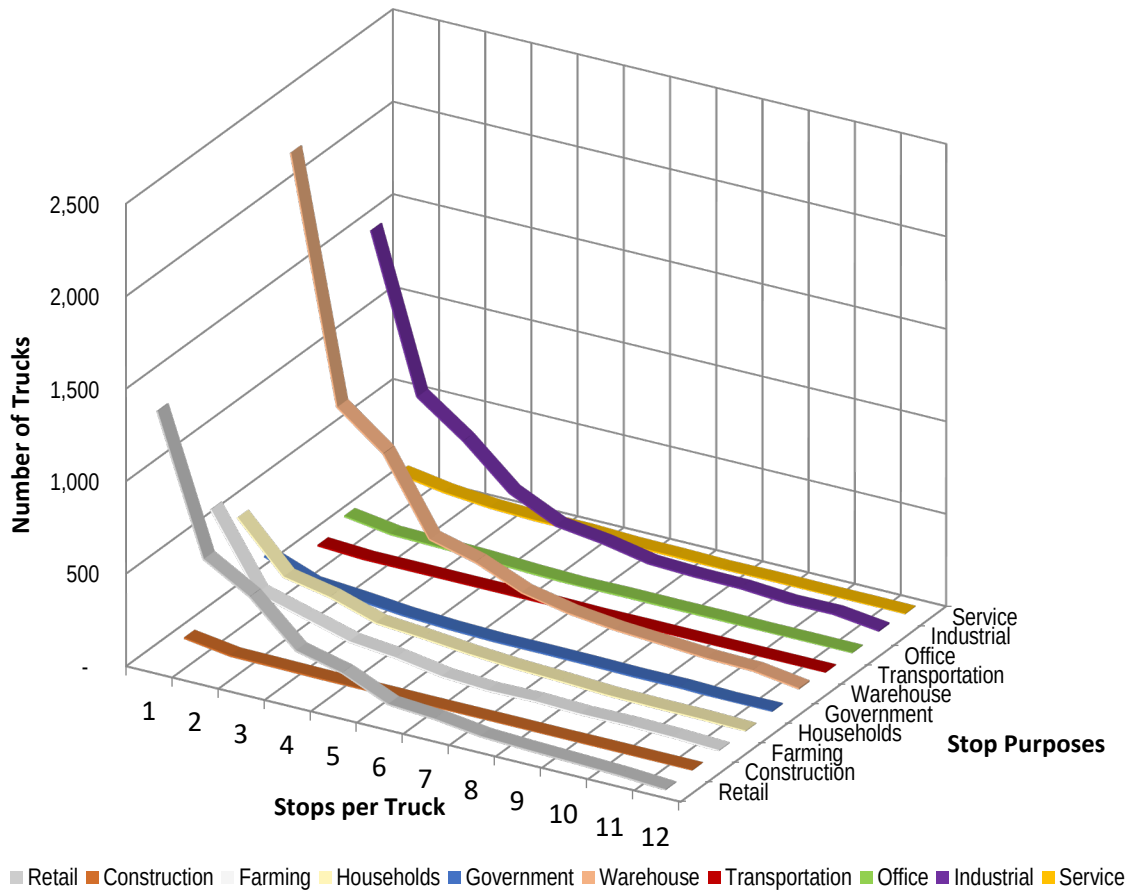


Figure 4-8 Stops per Truck by Tour Purpose based on ATRI Data



4.3 StreetLight Truck GPS data

4.3.1 Overview

Under a separate contract, MAG has contributed its own funds to purchase StreetLight Commercial Light and Medium Duty Trips GPS data ("StreetLight data") from INRIX for the development of light and medium truck tour models. The data contains the trip points for all Light and Medium Duty Commercial Vehicle trips that started or ended in any of the MAG zones in the April of 2015. MAG staff provided in-kind support to process the StreetLight data for the truck tour model estimation for light and medium trucks, following a similar procedure as outlined in Section 4.2 for ATRI data. Like other GPS data, the StreetLight data represents each ping as a single row. Table 4-1 shows the data dictionary for columns.

Table 4-1 StreetLight Data Dictionary

Columns	Format	Description
Vehicle ID	text	The ID of the device as provided by INRIX. Vehicle ID is not persistent across all trips for a given vehicle. It can and does change.
Vehicle Weight	text	The weight classification of the vehicle, as determined by provider: • 'Medium' indicates Medium Duty (14,000 lbs – 26,000 lbs) • 'Light' indicates Light Duty (< 14,000 lbs).
Trip ID	big integer	A unique identifier for the specific trip. Trip ID is unique within each month of data.
Point Index	integer	The sequence of this point (1 to N) within the trip.
Point Flag	text	The type of point with values: - 'SP' indicates the start point; - 'EP' indicates the end point*; - 'WP' indicates a waypoint. *: StreetLight system stop definition is when there is no movement within 5 meters for 5 minutes.
Point Latitude	real	The latitude, in degrees, for this point.
Point Longitude	real	The longitude, in degrees, for this point.
Point Time (epoch)	integer	The timestamp of the point in UTC-epoch format.

4.3.2 Data Statistics and Coverage

The StreetLight data contains a total of 22.3 million records in which 22 million are marked as collected from medium trucks. Details statistics are shown in Table 4-2. As a first step in the GPS data processing, the raw data were filtered in order to improve data consistency and relevancy:

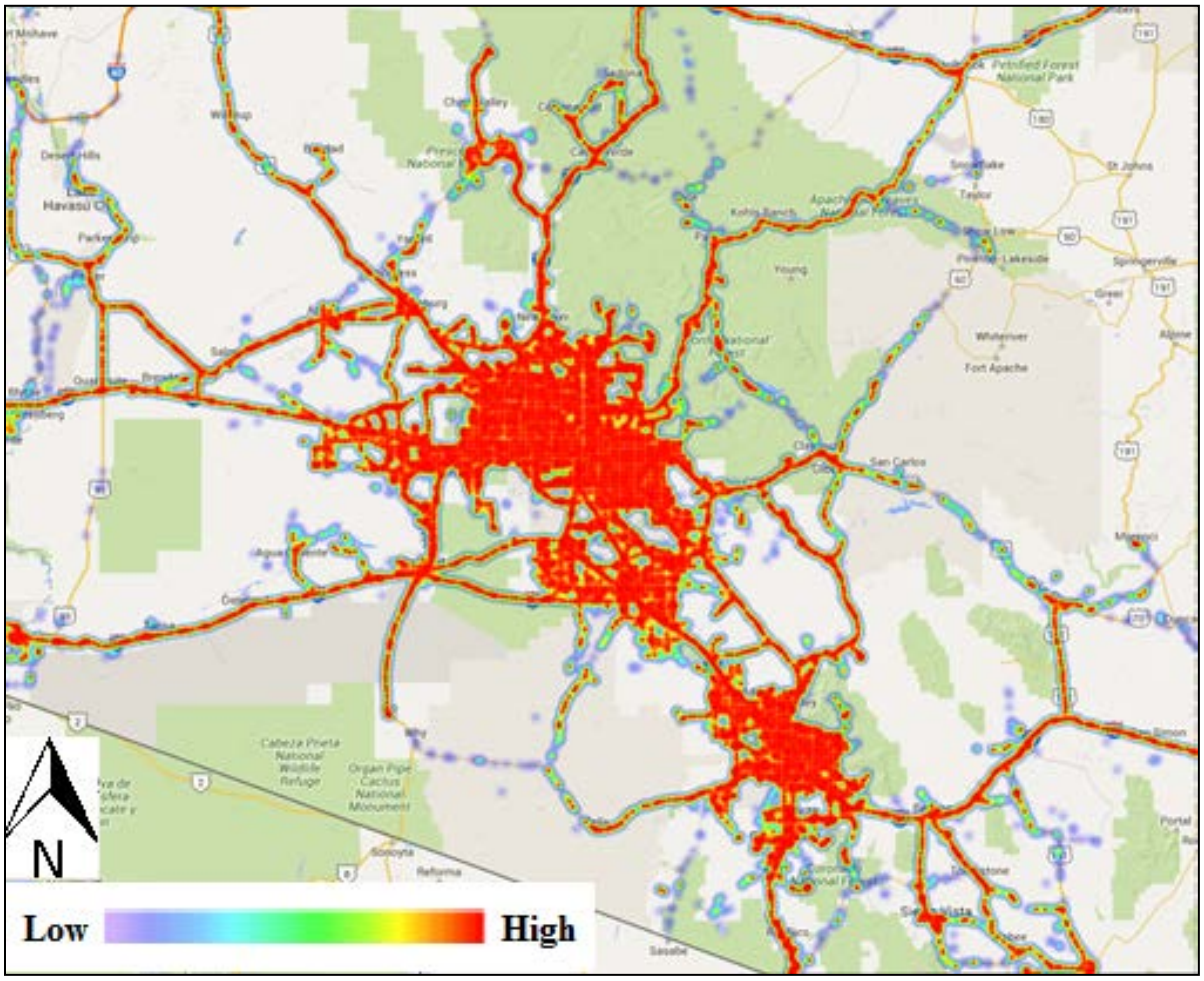
- Duplicate records were removed from the sample (258 trucks with 20,922 records),
- Removal of truck records with less than 10 pings (519 trucks with 3097 records).
- Removal of truck records with geographic coverage (defined as the diagonal of the bounding box that encloses all its pings) less than 1 mile (197 trucks with 6,021 records).
- Removal of truck records with operational time span (i.e. the duration between its first and last pings) less than 30 minutes (504 trucks with 4,447 records).

Only a very small fraction of records was removed after the initial data cleaning indicating the quality of data is decent. In terms of spatial coverage, as shown in Figure 4-9, the data covered the mega-region study area well: all major highways and urbanized areas have high concentration of GPS pings. It is worth noting that the data are not limited to the mega-region modeling boundary because they were collected from trucks that either started or ended in any of the mega-regional modeling area during the study period.

Table 4-2 StreetLight Data Statistics

	Medium Truck	Light Truck
Number of records	21,990,194 (98.4%)	358,205 (1.6%)
Number of unique Vehicle IDs	16,692 (94.4%)	1,000 (5.6%)
Number of unique Trip IDs	1,194,388 (98.2%)	22,366 (1.8%)

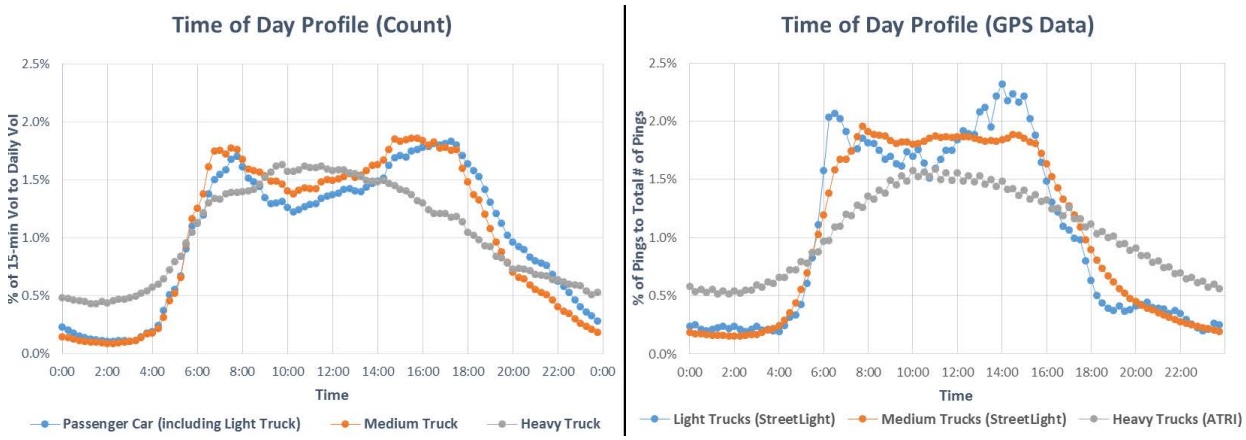
Figure 4-9 Heat Map of GPS Pings in StreetLight Data



After converting the timestamp to Arizona's Mountain Standard Time, the time of day profile was calculated and compared with time of day profile based on vehicle classification counts as a sanity check. Considering vehicle class definitions may differ between traffic counts and GPS data as well as the fact that GPS data (either StreetLight or ATRI) are essentially a sample of the whole truck population, the time of day profile comparison in Figure 4-10 demonstrate an acceptable degree of match. As expected,

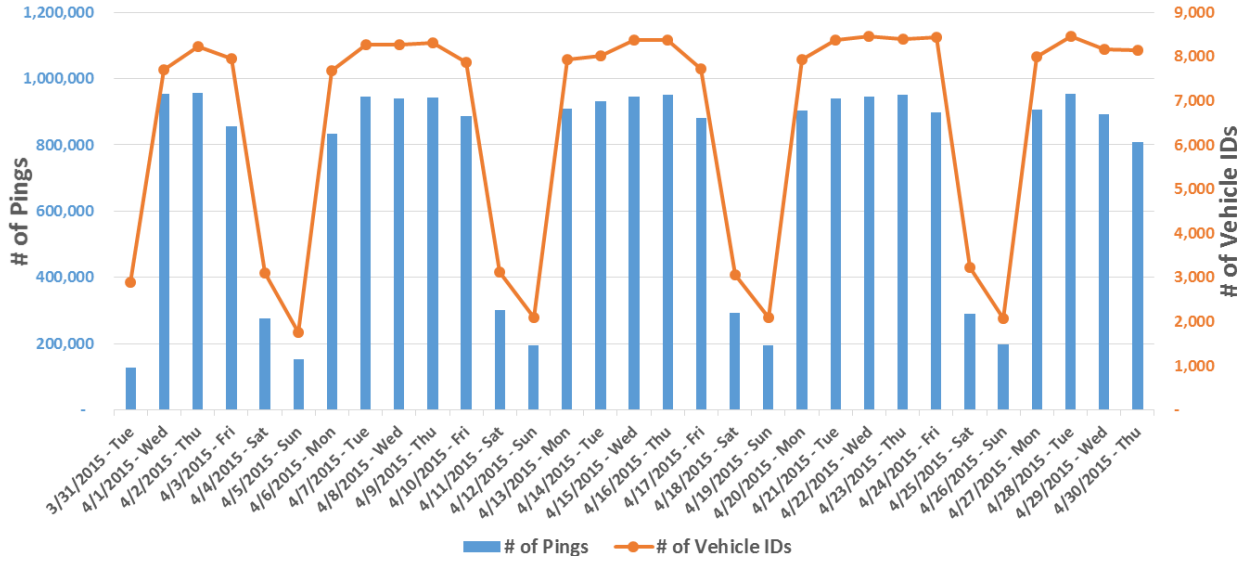
the peak operation time period for light and medium trucks was clearly concentrated around normal business hours from 6 am to 5 pm, compared to a more "spread-out" pattern observed in heavy trucks.

Figure 4-10 Time of Day Profile: Truck GPS Data vs Traffic Counts



The data were also examined in terms of its daily coverage within the month of April 2015. As shown in Figure 4-11, the amount of data (either measured by number of records or number of unique vehicle IDs) shows a consistent weekly pattern of variation across the one-month study. Although, because time zone adjustment, there are seven hours of data in the end of March 31st, 2015 and the last seven hours of April 30th, 2015 is partially missing.

Figure 4-11 Temporal Coverage of StreetLight Data



4.3.3 Data Processing

MAG used the similar procedure, as discussed in the Section 4.2.3 for ATRI data, to derive stops, associated land use, and trip statistics from StreetLight data for the light and medium truck tour model estimation purpose. The first step is identifying the truck stops and the procedure is below:

1) The first and the last record of a truck will be considered as stops.

2) If travel time between consecutive GPS records is less than 5 mph, than the truck is considered stopped
    a. The time stamp for the stop is the same as the first GPS ping that defines the stop
    b. The coordinates for the stop are the geographic center of the coordinates of both GPS pings
3) If a truck traveled less than 0.5 miles between stops, than it is assumed that they are a single stop
    a. The coordinates for the resulting stop are the geographic center of the coordinates all GPS pings that define such stop
4) Stops shorter than 5 minutes are considered to be associated with normal traffic and are discarded.
5) If a truck stops for more than 4 hours, this stop is considered the end of current tour, and if the truck moves from the stop later, this stop is the beginning of the subsequent tour.

Approximately 956,000 stops were identified by the algorithm. After filtering out stops that happened outside the modeling area, in the middle of freeways, freeway ramps, or arterial streets, or during the weekend, about 750,000 stops were retained. Figure 4-12 is the distribution of stop duration within 24 hours. Nearly 80% of stops last within 4 hours. Figure 4-13 and Figure 4-14 shows the distribution of number of stops for medium and light trucks, respectively.

Figure 4-12 Distribution of Truck Stop Duration (StreetLight Data)

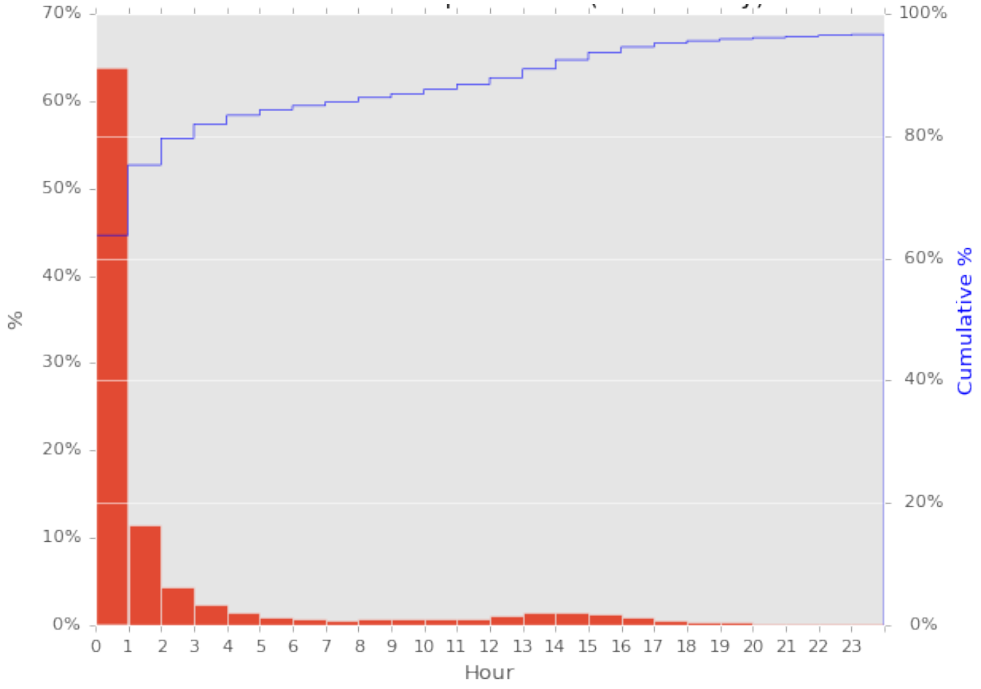


Figure 4-13 Distribution of Number of Stops for Medium Truck (per tour) (StreetLight Data)

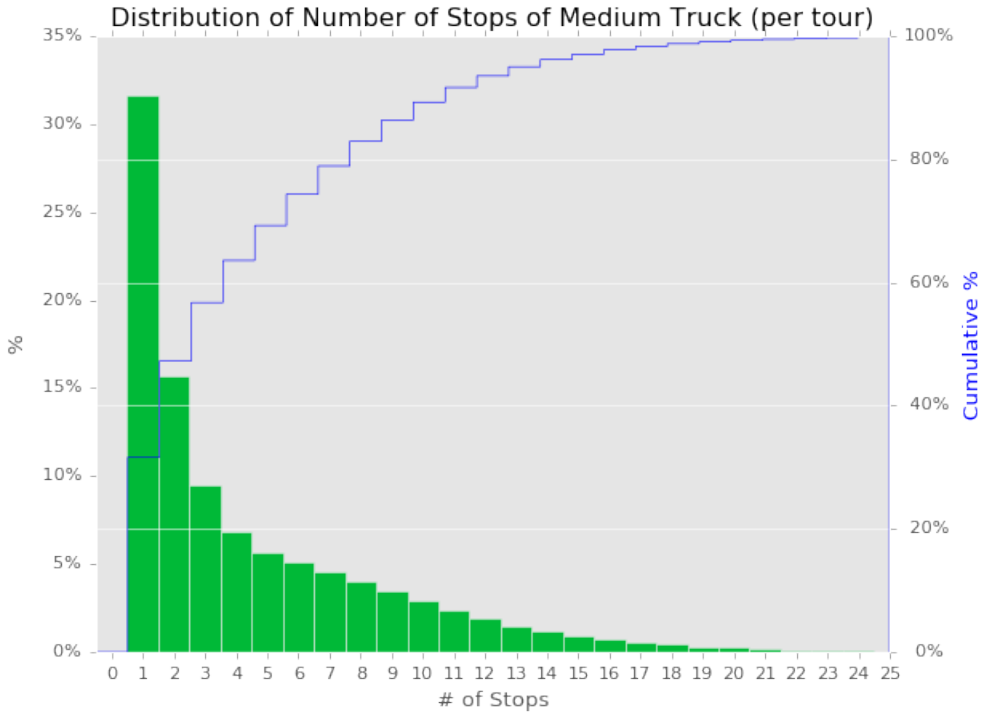
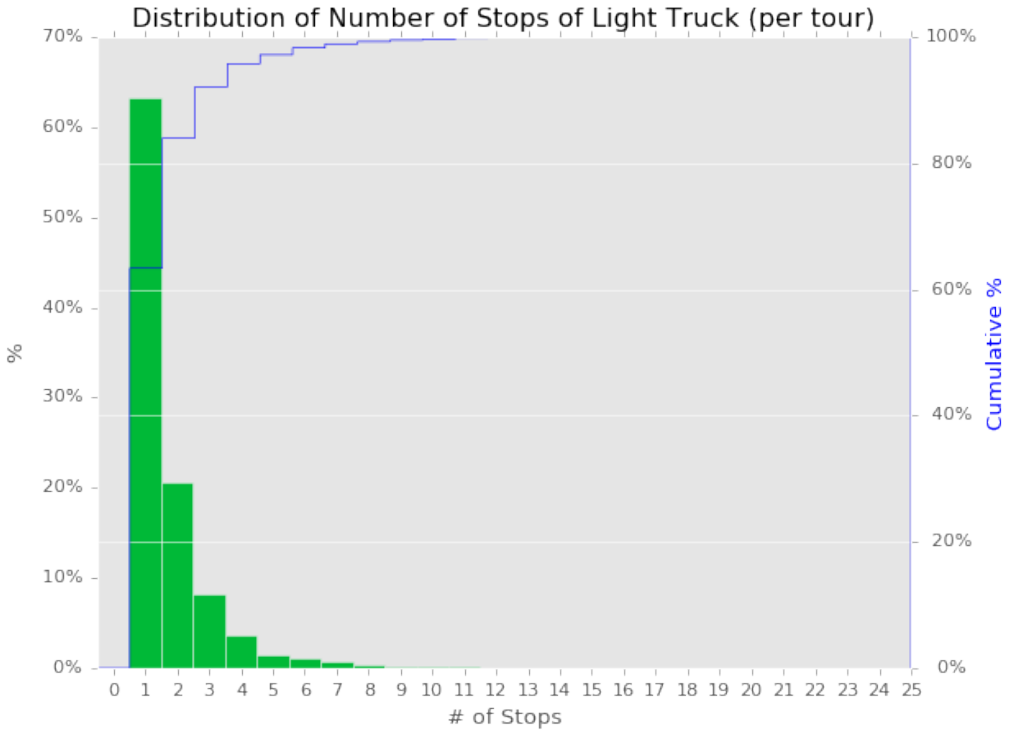


Figure 4-14 Distribution of Number of Stops for Light Trucks (per tour) (StreetLight Data)



Each stop was then assigned a stop purpose based on nearest land use or establishments as well as crosswalk tables between land use and stop purpose and between establishment NAICS code and stop purposes. Table 4-3 and Table 4-4 shows the summary of stop and tour statistics as well as tour purposes and stops per tour from the StreetLight data. Similarly to ATRI data processing, a stop table, a TAZ-to-TAZ trip matrix and a Stop Purpose-to-Stop Purpose trip matrix were produced for the model estimation.

Table 4-3 StreetLight Truck Tour Statistics

Truck Types	Tours	Share
Light	3023	2.10%
Medium	142,495	97.90%
Total	**145,518**	

Table 4-4 Summary of Tour Purposes and Stops per Tour based on StreetLight Data

Stop Purposes	Tours	Share	Average Number of Stops
Retail	24,199	16.70%	4.35
Construction	7,270	5.00%	4.14
Farming	1,888	1.30%	3.97
Households	58,990	40.70%	5.2
Government	11,328	7.80%	3.66
Warehousing	169	0.10%	4.85
Transportation	2,526	1.70%	6.24
Office	5,768	4.00%	4.43
Industrial	13,981	9.60%	4.67
Service/Other	18,881	13.00%	4.73

4.4 IHS Global Insight's TRANSEARCH Data

4.4.1 Overview

TRANSEARCH is an annual database of U.S. county-level freight movement data used for freight modeling and forecasting, is produced by the Trade and Transportation consulting practice within IHS Consulting.

TRANSEARCH 2013 release enhancements included improved industry freight data and air data, along with an improved methodology in agriculture flow processing and freight flow modeling. The US Department of Agriculture (USDA)'s estimates for production in 2013 have been incorporated in the latest TRANSEARCH release, which was acquired by MAG. TRANSEARCH 2013 was updated based upon the new 2012 NAICS (North American Industry Classification System) codes.

TRANSEARCH captures market-to-market flow data for more than 500 individual commodities [22] and seven modes of transportation:

- For-hire truckload
- For-hire less-than-truckload
- Private truck
- Rail/highway intermodal
- Air
- Water

Volume is the primary measurement of commodity flows, and is presented in terms of annual short tons. The short tons can be converted to other measures, such as

- Units (such as truck counts)
- Dollar Value
- Vehicle-miles traveled (VMT)
- Ton-miles

TRANSEARCH combines primary shipment data obtained from many of the nation's largest rail and truck freight carriers with information from public, commercial and proprietary sources to generate a base year estimate of freight flows at the county level. Once the base year is completed, a separate model is used to produce a 30-year forecast of freight flows. These projections are driven primarily by IHS Economics' long-term U.S. Macroeconomic and Business Markets Insights forecasts. Data from the TRANSEARCH model have been used by Federal government agencies, more than thirty U.S. states, various local governments, freight carriers of all modes, besides being utilized by several Consultants.

4.4.2 Commodity Groupings

Standard Transportation Commodity Codes (STCC) are used in the TRANSEARCH development process to organize and present commodity information for a variety of reasons, including the following:

- Compatibility with Surface Transportation Board (STB) Waybill data.
- Convertibility from international codes.
- The suitability of STCC to transportation.

IHS continuously updates translation tables allowing the data to be presented by other commodity coding systems, such as Standard Classification of Transported Goods (SCTG) system.

[22] TRANSEARCH 2013 Methodology Documentation

4.4.3 STB Confidential Carload Waybill Sample

For public agencies which have authorization from the Surface Transportation Board (STB), TRANSEARCH rail traffic is developed from the confidential version of those agencies annual Carload Waybill Sample. The Waybill Sample is a statistically based stratified sample of shipments terminated by U.S. rail carriers. The Waybill Sample file contains detailed information on the origin, destination, commodity and volume of each sampled movement. All carriers terminating 4,500 or more carloads per year are required to report, resulting in data capture from all Class I and Class II railroads. Carriers moving fewer than 4,500 annual loads are sampled only if they act as haulage agents for larger railroads.

4.4.4 Data Capture Exclusions

The following types of movements are not captured in TRANSEARCH [23]:

- Drayage for inland waterways, pipelines, international air, and rail carload transfers.
- Non-manufactured goods – primary (raw) products from fisheries and logging camps, and waste.
- Small package and mail shipments moved entirely over-the-road.
- Military and other government trucks.
- Household goods and local service trucks (such as utility repair vehicles).
- Domestic pipeline flows (although some cross-border pipeline flows are included).

4.4.5 TRANSEARCH Data Acquisition by MAG

TRANSEARCH database was purchased by MAG with a base year of 2013. It contains freight flows to, from, within, and through the study area are the five Arizona Counties (complete coverage of Maricopa County, Pinal County and Pima County; partial coverage of Gila County and Yavapai County) which are wholly or partially contained in the MAG-PAG mega-region. The database has Commodities at STCC4 level (Standard Transportation Commodity Code), Equipment Types, Highway Network, Highway Routes, Modes, Regions, Rail Flows, TRANSEARCH flows at Region Level, TRANSEARCH flows at TAZ level, etc. The flow tables from this database would be converted to standard trip tables for use in the model.

In the latest acquisition, IHS provided 2013 base year TRANSEARCH commodity flow data at the Travel Analysis Zone (TAZ) level for MAG Modeling Area for the Activity Based Model. The "Modes" field included Truck (Less-Than-Truckload, Private, Truckload), Rail (Carload and Intermodal), Water, Air, and Pipeline. Commodity classification was based upon four-digit Standard Transportation Commodity Code (STCC), as noted earlier. The "Units" hauling commodity flow consisted of Short tons (2000 pounds), Truck equivalents, Railcars (for non-intermodal rail) and trailers/containers (for intermodal). The value was expressed in Commodity dollars per ton table.

The "Equipment Type" includes coding for Dry Van, Reefer (Refrigerated Truck), Flat Bed, Tank, Dump and Other. Inland Trade Identification consists of an indication for Import, Export, Domestic, and North American Free Trade Agreement (NAFTA) flag for each record. Highway routing indicators

[23] TRANSEARCH 2013 Methodology Documentation

like First/Last Segments, From/To Federal Information Processing Standard (FIPS) codes, and Entry/Exit road on records, links, segment tables, along with shape files, were acquired.

Since the Regional Behavioral Based Freight Model was envisioned to encompass various modes like Rail, MAG needed to work with AZ Carload Waybill Sample. IHS integrated AZ Carload Waybill Sample (provided by ADOT - Arizona Department of Transportation) into TRANSEARCH database and provided Origin and Destination information for each commodity flow record along with routing. Apart from 2013 base year data, MAG acquired forecast year database for 2020, 2030 and 2040 in the same format as that of base year. Commodity values for forecast years were chained to base year dollar. MAG provided TAZ shape files for the MAG modeling area to be used by IHS for spatial aggregation of TRANSEARCH data.

As part of the data purchase, MAG obtained full access rights to the entire TRANSEARCH data that was acquired independently by ADOT. MAG accessed the following data from Arizona Department of Transportation:

- Arizona Freight Finder Database for 2013 (Establishment-level freight generating data, including details of value of sales, number of employees, tons by commodity and mode, address, size of facility, NAICS classification, and other metrics).
- Highway and Rail shape-files for 2013 (US network shape file and associated routing tables to allow for custom routing assignments). MAG was aware of the fact that access to rail data was by non-disclosure agreement with Surface Transportation Board (as ADOT is the custodian of STB sample data for Arizona).

4.4.6 Data Analysis Summary from TRANSEARCH

There were several types of data analysis that were performed using TRANSEARCH data. Table 4-5, Table 4-6, and Table 4-7 illustrate county-wise annual Commodity Flow summary from 2013 to 2045, for Maricopa County, Pinal County and Pima County respectively. Table 4-8 shows county-to-county annual commodity flow in tons for 2013. Table 4-9 shows county-to-county annual commodity flow in dollars for 2013. Table 4-10 shows county-to-county annual commodity flow in tons for 2045. Table 4-11 shows county-to-county annual commodity flow in dollars for 2045, indexed to 2013 dollar. These tables showcase the projected robust growth for MAG-PAG mega-region.

For the first time at MAG, the evaluation of commodity flow growth by rail for Maricopa County was possible for years 2013 to 2040. Table 4-12 shows Internal-Internal Commodity Flow Growth by Rail for Maricopa County. Table 4-13 shows Internal-External Commodity Flow Growth by Rail for Maricopa County. Table 4-14 shows External-Internal Commodity Flow Growth by Rail for Maricopa County. Table 4-15 shows the External-External Truck flows to the Mega-region. Table 4-16 shows Arizona-NAFTA Annual Commodity Flow Summary by truck mode. Table 4-17 shows MAG/PAG Mega-Region and NAFTA Annual Commodity Flow Summary by truck mode. Table 4-18 shows the 2013 Mega-Region Inter-county Flows by Truck by SCTG2 (in tons). Table 4-19 shows the 2013 Mega-Region Inter-county Flows by Truck by SCTG2 (in dollars). Table 4-20 shows the 2013 Mega-Region Inter-county Flows by Rail by SCTG2 (in tons). Table 4-21 shows the 2013 Mega-Region Inter-county Flows by Rail by SCTG2 (in dollars).

MAG staff developed STCC2 to SCTG2 crosswalk, after working with the Consultant for the project, to replicate each of the above calculations using SCTG2. The Supply Chain Model (including Mode Choice

Model) has 43 commodities at SCTG2 level; the model itself was developed using Freight Analysis Framework data (version 4.1) database for 2015. Most importantly, the outputs from that model were verified using TRANSEARCH data for 2015 (using compound annual growth rate from 2013) and after using SCTG2-STCC2 crosswalk. The mega-region External-External Truck flow was developed for each of SCTG2 codes (1-43), showcasing the versatility of TRANSEARCH database, in the absence of such data from Freight Analysis Framework data (version 4.1).

Table 4-5 Maricopa County Commodity Flow

Year	I-I Weight (in Tons)	I-I Flow Value (dollars)	I-E Weight (in Tons)	I-E Flow Value (dollars)	E-I Weight (in Tons)	E-I Flow Value (dollars)	E-E* Weight (in Tons)	E-E* Flow Value (dollars)
2013	48,026,672	44,206,555,866	21,437,708	31,331,078,244	28,856,407	47,939,136,203	32,650,700	111,653,726,556
2020	53,124,613	52,593,356,834	26,241,396	42,667,617,359	39,303,029	62,623,497,388	41,841,975	145,933,097,022
2030	65,727,888	82,142,086,869	29,168,271	57,806,920,444	53,656,952	85,513,287,582	54,652,379	198,895,542,366
2035	68,936,041	94,378,392,325	31,356,073	65,160,285,243	63,005,781	101,849,304,342	63,084,484	235,089,214,210
2040	70,958,579	107,495,117,100	34,391,913	74,991,834,496	73,252,241	121,407,971,740	75,609,357	290,022,555,282
2045	73,172,778	119,467,571,603	39,096,251	88,475,061,218	94,180,101	151,035,210,060	95,813,192	370,213,830,241
	379,946,571	500,283,080,598	181,691,613	360,432,797,004	352,254,511	570,368,407,315	363,652,087	1,351,807,965,678
***: Flows are external-external to the megaregion**								

Table 4-6 Pinal County Commodity Flow Summary

Year	I-I Weight (in Tons)	I-I Flow Value (dollars)	I-E Weight (in Tons)	I-E Flow Value (dollars)	E-I Weight (in Tons)	E-I Flow Value (dollars)	E-E* Weight (in Tons)	E-E* Flow Value (dollars)
2013	1,975,224	347,448,813	8,406,869	2,855,367,842	6,553,197	4,047,197,720	32,650,700	111,653,726,556
2020	2,521,642	363,874,158	12,063,601	3,538,418,130	8,460,387	5,115,854,545	41,841,975	145,933,097,022
2030	3,641,945	417,552,245	16,350,589	4,404,562,833	8,925,009	6,540,962,899	54,652,379	198,895,542,366
2035	4,147,157	444,680,012	18,894,542	4,842,502,549	9,189,928	7,352,780,614	63,084,484	235,089,214,210
2040	4,125,472	473,566,937	20,103,534	5,391,832,315	10,155,441	8,288,421,217	75,609,357	290,022,555,282
2045	4,415,413	507,576,825	22,391,031	6,117,194,911	11,295,235	9,443,137,772	95,813,192	370,213,830,241
	20,826,853	2,554,698,990	98,210,166	27,149,878,580	54,579,197	40,788,354,768	363,652,087	1,351,807,965,678

Table 4-7 Pima County Commodity Flow Summary

Year	I-I Weight (in Tons)	I-I Flow Value (dollars)	I-E Weight (in Tons)	I-E Flow Value (dollars)	E-I Weight (in Tons)	E-I Flow Value (dollars)	E-E* Weight (in Tons)	E-E* Flow Value (dollars)
2013	7,355,113	10,155,640,861	6,669,678	6,909,958,877	7,739,566	12,979,085,552	32,650,700	111,653,726,556
2020	7,225,675	12,423,761,651	7,553,766	9,246,074,254	10,745,110	16,468,247,934	41,841,975	145,933,097,022
2030	6,481,456	16,201,100,764	8,657,376	12,690,168,561	14,355,265	22,831,379,218	54,652,379	198,895,542,366
2035	6,574,309	18,788,169,760	9,516,564	14,645,737,653	16,021,586	26,963,063,418	63,084,484	235,089,214,210
2040	7,554,020	21,513,485,733	10,643,022	16,668,172,231	17,109,984	32,229,437,752	75,609,357	290,022,555,282
2045	8,269,495	23,719,352,153	11,890,127	19,425,173,249	19,222,186	39,161,654,672	95,813,192	370,213,830,241
	43,460,068	102,801,510,922	54,930,532	79,585,284,826	85,193,696	150,632,868,547	363,652,087	1,351,807,965,678

Table 4-8 2013 County-to-County Annual Commodity Flow in Tons

	Maricopa	Pinal	Pima	Total
Maricopa	48,026,667	3,234,034	1,352,905	52,613,606
Pinal	5,455,512	1,975,224	1,045,434	8,476,170
Pima	1,291,326	937,577	7,355,113	9,584,016

Table 4-9 2013 County-to-County Annual Commodity Flow in Dollars

	Maricopa	Pinal	Pima	Total
Maricopa	44,206,556,267	1,254,331,068	2,369,456,478	47,830,343,814
Pinal	951,682,222	347,448,813	117,003,299	1,416,134,334
Pima	1,641,369,184	223,847,967	10,155,640,861	12,020,858,012

Table 4-10 2045 County-to-County Annual Commodity Flow in Tons

	Maricopa	Pinal	Pima	Total
Maricopa	73,403,399	2,726,919	3,640,258	79,770,576
Pinal	17,697,851	4,415,413	1,622,219	23,735,484
Pima	3,001,250	1,649,213	8,269,495	12,919,958

Table 4-11 2045 County-to-County Annual Commodity Flow in Dollars indexed to 2013

	Maricopa	Pinal	Pima	Total
Maricopa	121,061,479,690	2,493,141,583	6,339,447,506	129,894,068,779
Pinal	2,155,253,292	507,576,825	188,608,584	2,851,438,701
Pima	3,969,638,413	674,465,774	23,719,352,153	28,363,456,340

Table 4-12 Internal-Internal Commodity Flow Growth by Rail for Maricopa County

County	Year	Movement	Commodity (Tons)	Commodity (Value)	Mode	CAGR (Tons)	CAGR (Value)
Maricopa	2013	I-I	44,080	40,193,889	Rail		
Maricopa	2020	I-I	56,710	54,998,428	Rail	3.66	4.58
Maricopa	2030	I-I	73,225	86,358,076	Rail	3.03	4.60
Maricopa	2035	I-I	84,179	111,156,618	Rail	2.98	4.73
Maricopa	2040	I-I	101,209	155,450,494	Rail	3.13	5.14

Table 4-13 Internal-External Commodity Flow Growth by Rail for Maricopa County

County	Year	Movement	Commodity (Tons)	Commodity (Value)	Mode	CAGR (Tons)	CAGR (Value)
Maricopa	2013	I-E	869,568	1,729,423,104	Rail		
Maricopa	2020	I-E	1,078,636	2,207,125,078	Rail	3.13	3.55
Maricopa	2030	I-E	1,370,037	2,953,016,942	Rail	2.71	3.20
Maricopa	2035	I-E	1,513,912	3,305,821,404	Rail	2.55	2.99
Maricopa	2040	I-E	1,726,239	3,780,708,341	Rail	2.57	2.94

Table 4-14 External-Internal Commodity Flow Growth by Rail for Maricopa County

County	Year	Movement	Commodity (Tons)	Commodity (Value)	Mode	CAGR (Tons)	CAGR (Value)
Maricopa	2013	E-I	6,371,328	12,930,458,482	Rail		
Maricopa	2020	E-I	7,931,069	16,620,042,634	Rail	3.18	3.65
Maricopa	2030	E-I	9,855,229	20,568,546,380	Rail	2.60	2.77
Maricopa	2035	E-I	10,961,397	22,710,424,180	Rail	2.50	2.59
Maricopa	2040	E-I	12,560,496	25,667,415,765	Rail	2.55	2.57

Table 4-15 2013 External-External Truck Flow for Mega-Region

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
3	0	0	0	0	0	0	0	0	262	20	3	0	13	21	0	0	0	**320**
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
5	0	0	0	0	0	0	0	0	0	90	103	0	38	90	0	0	0	**320**
6	0	0	0	0	0	0	0	0	0	7	566	0	18	7	0	0	0	**597**
7	0	0	0	0	0	0	0	35	0	109	0	0	0	109	0	0	0	**254**
8	0	0	0	0	0	0	56	0	0	62	0	0	2	62	0	0	0	**183**
9	0	0	0	0	0	0	0	0	0	20	3	0	15	20	0	0	0	**60**
10	0	0	35	0	74	2	87	68	37	0	580	0	71	5,675	0	0	0	**6,629**
11	0	0	1	0	39	517	0	0	0	530	0	0	12	379	0	0	0	**1,477**
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
13	0	0	4	0	41	10	0	1	4	133	53	0	0	133	0	0	0	**378**
14	0	0	35	0	74	2	87	68	37	5,853	712	0	71	0	0	0	0	**6,938**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Total	**0**	**0**	**74**	**0**	**227**	**532**	**230**	**171**	**341**	**6,824**	**2,021**	**0**	**240**	**6,496**	**0**	**0**	**0**	**17,156**

Note - Truck types captured are "Truck Truckload" Or "Truck L-T-L" Or "Truck PVT" Or "Truck NEC"; They are Heavy Trucks + Medium Trucks

Table 4-16 2013 Arizona – NAFTA Annual Commodity Flow Summary

Freight Flow Year	AZ-NAFTA Weight (in Tons)	AZ-NAFTA Flow Value (dollars)	NAFTA-AZ Weight (in Tons)	NAFTA-AZ Flow Value (dollars)	E-E (NAFTA Pass through AZ) Weight (in Tons)	E-E (NAFTA Pass through AZ) Flow Value (dollars)
2013	1,147,006	5,178,402,540	3,347,617	7,638,381,929	3,065,768	15,299,118,031
2020	1,706,208	8,008,530,441	5,113,863	11,670,043,787	4,687,248	23,824,344,185
2030	2,624,240	13,188,818,428	8,073,348	18,487,111,509	7,350,114	38,706,799,888
2035	3,229,801	16,830,741,014	10,104,167	23,295,042,763	9,185,030	49,301,154,423
2040	4,133,996	22,419,888,593	12,898,263	29,904,475,436	11,842,376	64,791,455,137
2045	5,589,487	31,887,452,338	16,964,575	39,635,099,328	16,026,458	89,686,352,300
Total	**18,430,737**	**97,513,833,354**	**56,501,832**	**130,630,154,753**	**52,156,994**	**281,609,223,964**

Note: These commodity volumes are hauled by trucks only.

Table 4-17 2013 MAG/PAG Mega-Region and NAFTA Annual Commodity Flow Summary

Freight Flow Year	Megaregion-NAFTA Weight (in Tons)	Megaregion-NAFTA Flow Value (dollars)	NAFTA-Megaregion Weight (in Tons)	NAFTA-Megaregion Flow Value (dollars)	E-E* (NAFTA Pass through Megaregion) Weight (in Tons)	E-E* (NAFTA Pass through Megaregion) Flow Value (dollars)
2013	457,324	2,271,242,675	2,734,005	6,485,403,680	3,113,226	15,394,863,203
2020	676,650	3,468,074,972	4,179,675	9,930,726,770	4,758,770	23,966,737,065
2030	1,029,298	5,627,389,414	6,603,656	15,797,836,055	7,463,519	38,926,821,615
2035	1,261,095	7,118,139,203	10,561,352	25,664,146,675	9,327,280	49,575,659,387
2040	1,606,819	9,366,501,699	13,902,686	34,097,307,282	12,025,154	65,143,084,898
2045	2,163,784	13,074,988,745	13,902,686	34,097,307,282	16,268,859	90,154,899,541
Total	**7,194,970**	**40,926,336,708**	**51,884,060**	**126,072,727,744**	**52,956,808**	**283,162,065,709**

Note: These commodity volumes are hauled by trucks only; The Megaregion is assumed to be Maricopa County, Pinal County and Pima County only; The E-E flows have both NAFTA flows routed via state highways along with NAFTA related flows produced/consumed outside of the megaregion but within AZ.

Table 4-18 2013 Mega Region Inter-County Flows by Truck by SCTG2 (in tons)

Tons (by Truck)	4013-4013	4013-4021	4013-4019	4021-4013	4021-4021	4021-4019	4019-4013	4019-4021	4019-4019
SCTG	MC - MC	MC - Pinal	MC - Pima	Pinal - MC	Pinal-Pinal	Pinal - Pima	Pima - MC	Pima - Pinal	Pima-Pima
1	12,418	1,108	553	98,366	51,836	3,895	2,461	1,878	375
2	0	0	29,049	0	0	10,726	0	0	2,200
3	311,440	134,697	29,276	173,334	144,328	29,323	6,926	4,922	9,154
4	120,451	17,537	11,115	43,650	26,154	23,926	3,624	914	4,824
5	54,941	1,171	2,036	590	124	72	46	0	99
6	357,502	5,415	22,805	1,319	966	815	2,937	394	13,747
7	2,939,304	174,765	129,128	334,618	125,481	22,598	138,752	68,803	241,700
8	10,630	186	315	0	0	0	0	0	0
9	0	0	0	0	0	0	191	0	3,969
10	0	0	0	0	0	0	1,026	283	2,184
11	40,837,052	4,101,908	409,048	972,037	1,131,500	817,145	110,595	932,314	7,866,607
12	6,109,354	18,255	13,575	5,979,296	1,705,354	833,540	63,119	11,834	616,886
13	70,273	3,224	1,316	43,362	20,097	3,647	2,762	3,835	22,055
14	0	0	0	0	0	0	138,299	38,871	26,276
15	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	6,111,090	406,009	43,664	93,668	27,782	26,381	92,581	53,986	1,328,404
20	54,169	1,753	6,197	0	0	0	1,091	26	1,479
21	31,558	473	3,190	0	0	0	1,091	26	1,479
22	94,048	344	2,709	42,811	1,027	8,313	0	0	0
23	146,496	2,166	10,493	4,220	337	2,272	1,271	75	6,567
24	471,752	6,114	21,615	9,204	1,048	1,535	15,113	705	13,926
25	76,043	6,341	4,224	8,453	794	774	0	0	0
26	483,242	17,799	31,002	19,439	1,146	3,347	23,908	3,318	35,755
27	25,813	352	398	0	0	0	0	0	0
28	236,431	5,834	11,419	32	0	0	14,119	1,242	12,891
29	296,591	1,975	8,468	491	46	48	8,850	404	24,037
30	112,978	869	5,460	297	0	28	2,884	24	9,955
31	5,773,935	75,964	128,814	1,040,699	133,924	96,041	258,560	375,368	1,246,179
32	361,910	2,412	11,359	28,995	2,952	2,069	12,963	379	8,122
33	745,392	9,846	31,160	7,154	899	1,812	21,333	1,425	66,581
34	287,148	12,097	9,262	656	634	86	4,070	4,180	17,358
35	591,298	2,626	18,711	493	27	124	8,712	98	40,208
36	273,876	6,046	33,834	376	323	260	204	0	1,050
37	19,857	0	618	0	0	0	328	0	49,557
38	29,470	527	1,528	0	0	0	2,600	52	5,783
39	92,393	768	6,325	728	88	126	836	0	2,347
40	40,350	993	2,855	34	22	0	403	13	2,791
41	1,874,286	151,612	6,085	40,358	34,071	1,081	131,654	64,696	25,893
42	0	0	0	0	0	0	0	0	0
43	4,062,118	160,768	593,959	34,083	0	110	498,740	23,214	222,002

Table 4-19 2013 Mega Region Inter-County Flows by Truck by SCTG2 (in dollars)

Tons (by Truck)	4013-4013	4013-4021	4013-4019	4021-4013	4021-4021	4021-4019	4019-4013	4019-4021	4019-4019
SCTG	MC - MC	MC - Pinal	MC - Pima	Pinal - MC	Pinal-Pinal	Pinal - Pima	Pima - MC	Pima - Pinal	Pima-Pima
1	29,612,878	2,599,958	1,335,635	229,299,851	120,792,450	9,090,441	5,772,229	4,375,239	891,389
2	0	0	7,088,014	0	0	2,617,190	0	0	536,872
3	204,400,005	95,373,068	21,196,836	115,408,655	85,486,447	19,699,265	4,400,937	3,414,063	3,253,014
4	48,147,381	4,682,521	3,400,974	10,620,751	6,363,601	5,821,588	1,810,575	370,285	2,473,602
5	217,269,342	4,600,495	8,063,989	2,416,480	507,651	295,178	187,942	0	404,343
6	770,354,874	11,741,218	53,532,625	3,502,566	2,565,165	2,164,982	7,800,676	1,046,306	36,517,020
7	2,375,291,032	91,193,885	121,008,223	148,535,385	68,223,759	28,118,185	73,779,951	27,273,372	165,628,339
8	22,249,846	389,475	659,816	0	0	0	0	0	0
9	0	0	0	0	0	0	2,925,783	0	68,295,059
10	0	0	0	0	0	0	202,131	55,724	430,510
11	292,434,880	29,373,842	2,929,200	6,960,775	8,102,695	5,851,593	791,970	6,676,317	56,332,922
12	50,351,196	150,455	111,881	49,279,316	14,054,944	6,869,754	520,203	97,531	5,084,164
13	16,248,006	360,821	787,309	3,156,889	1,451,126	1,061,810	301,327	261,756	2,866,392
14	0	0	0	0	0	0	273,122,588	78,695,032	52,058,003
15	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	5,187,086,935	423,749,499	13,589,802	9,961,578	2,954,582	2,805,588	17,869,224	43,311,849	1,197,847,334
20	692,102,129	11,671,508	70,955,483	0	0	0	22,841,201	544,512	30,979,593
21	660,906,228	9,905,391	66,806,732	0	0	0	22,841,201	544,512	30,979,593
22	38,617,232	141,318	1,112,436	17,578,456	421,813	3,413,260	0	0	0
23	469,196,935	6,961,763	35,445,695	13,390,629	1,085,226	7,577,521	5,453,522	253,057	25,894,120
24	1,730,182,194	22,402,869	79,045,838	34,549,825	3,927,591	5,745,430	54,498,124	2,582,385	50,634,152
25	11,243,021	937,532	624,509	1,249,737	117,336	114,427	0	0	0
26	300,816,636	8,900,702	16,855,741	10,276,827	825,471	2,594,485	18,449,036	2,441,537	28,100,842
27	50,200,690	684,564	773,788	0	0	0	0	0	0
28	465,777,744	11,789,952	24,694,271	69,844	0	0	27,416,741	2,410,843	25,049,661
29	871,241,652	6,590,590	27,610,933	1,778,132	91,263	95,730	25,412,446	1,192,837	73,577,596
30	658,008,612	5,110,363	35,605,145	1,757,813	0	109,580	19,270,798	92,735	64,222,073
31	1,034,499,707	16,486,835	32,761,371	101,082,445	19,980,880	16,331,220	50,449,272	25,313,689	200,168,110
32	1,099,568,475	7,871,228	51,691,010	55,401,183	8,123,221	4,234,134	34,847,785	989,120	31,258,984
33	2,428,918,053	29,605,874	102,942,538	21,024,381	2,622,572	5,338,734	59,354,147	3,955,259	213,016,724
34	5,141,734,504	96,408,885	75,163,700	5,513,398	7,342,015	698,085	35,883,520	38,561,721	167,472,294
35	6,032,742,937	21,413,161	181,363,391	2,851,405	149,361	705,804	76,414,797	629,988	342,085,678
36	2,386,049,402	56,385,611	315,153,843	1,299,553	1,099,372	831,677	1,254,364	0	6,268,397
37	2,464,973,241	0	160,822,058	0	0	0	74,596,976	0	13,387,124,626
38	596,556,627	8,772,584	24,492,883	0	0	0	178,145,461	1,018,546	297,888,487
39	408,180,800	3,351,975	26,707,604	3,592,634	426,633	563,793	5,127,815	0	12,959,389
40	308,193,959	6,456,529	18,675,670	193,613	152,975	0	3,105,288	557,317	38,629,085
41	472,606,547	38,800,237	1,270,300	10,082,668	8,717,551	195,314	33,109,364	16,555,947	4,706,737
42	0	0	0	0	0	0	0	0	0
43	11,465,675,877	318,939,896	1,075,506,088	158,597,266	0	511,087	850,950,991	28,620,887	652,895,294

Table 4-20 2013 Mega Region Inter-County Flows by Rail by SCTG2 (in tons)

Tons (by Rail)	4013-4013	4013-4021	4013-4019	4021-4013	4021-4021	4021-4019	4019-4013	4019-4021	4019-4019
SCTG	MC - MC	MC - Pinal	MC - Pima	Pinal - MC	Pinal-Pinal	Pinal - Pima	Pima - MC	Pima - Pinal	Pima-Pima
1	0	0	0	0	0	0	0	0	0
2	30,000	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	7,040	0	0	0	0	0	0	0	0
20	0	0	0	3,360	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0
23	7,040	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0

Table 4-21 2013 Mega Region Inter-County Flows by Rail by SCTG2 (in dollars)

Tons (by Rail)	4013-4013	4013-4021	4013-4019	4021-4013	4021-4021	4021-4019	4019-4013	4019-4021	4019-4019
SCTG	MC - MC	MC - Pinal	MC - Pima	Pinal - MC	Pinal-Pinal	Pinal - Pima	Pima - MC	Pima - Pinal	Pima-Pima
1	0	0	0	0	0	0	0	0	0
2	7,320,000	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	8,414,929	0	0	0	0	0	0	0	0
20	0	0	0	733,399	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0
23	24,458,960	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0

4.5 Freight Analysis Framework (FAF 4.1) data

The Freight Analysis Framework (FAF) integrates data from a variety of sources to create a comprehensive picture of freight movement among states and major metropolitan areas by all modes of transportation. With data from the 2012 Commodity Flow Survey and additional sources, FAF Version 4.1 (FAF4.1) provides estimates for tonnage and value, by commodity type, mode, origin, and destination for 2015, the most recent year, and forecasts through 2045. The FAF4.1 Regional Database for 2015 and forecasts through 2045 is available in Microsoft Access format. The database consists of the data tables and descriptive information. The records in the flow table contain the following information:

- Foreign Origin - The foreign origin, if any, for the record. The foreign zones are those specified in the FAF include the countries of Canada and Mexico.
- Domestic Origin - The domestic origin for the record. The domestic zones are the 123 FAF3 regions covering the United States.
- Domestic Destination - The domestic destination for the record. The domestic zones are the same as described and defined under "Domestic Origin" above.
- Foreign Destination - The foreign destination, if any, for that record. The foreign zones are the same as those described above under "Foreign Origin."
- SCTG - The commodity being reported in that record according the Standard Classification of Transported Goods (SCTG). This is reported at a two-digit level.
- Domestic Mode - The modes used for domestic transportation according to the coding convention in FAF4.1 are shown in Table 4-22.
- Foreign Mode – It comprises of the mode of arrival to zones of entry for imports or mode of departure from zones of exit for exports; they do not include modes used in foreign countries to and from foreign ports and airports.
- Tons - The flow in annual kilotons for that record in the year 20XX. These years include the base year flow in 2015 and forecast flows for 2020, 2025, 2030, 2035, 2040 and 2045.
- Value – The value of commodity flow in millions of dollars in the year 20XX. These years include the base year value in 2015 and forecast flows for 2020, 2025, 2030, 2035, 2040 and 2045, indexed to 2015 dollar.
- FAF Region – The 132 domestic regions as defined in the FAF4.1 user guide.
- State – The 50 States and Washington, D.C.
- Foreign Region – The 8 international regions defined in Table 4 were based on information obtained from the United Nations Statistics Division. Specific countries included in each region can be found at the United Nations site.

- Zone of Entry – This is the FAF region or state where an import enters the United States. This term refers to the entire region or state and does not limit the entry to any single location. This replaces the term "international gateway" used previously.
- Zone of Exit – This is the FAF region or state where an export leaves the United States. This term refers to the entire region or state and does not limit the exit to any single location.

FAF4.1 uses Multiple Modes and Mail, rather than intermodal to represent commodities that move by more than one mode. Intermodal typically refers to containerized cargo that moves between ship and surface modes or between truck and rail, and repeated efforts to identify containerized cargo in the CFS have proved unsuccessful. Shipments reported as Multiple Modes can include anything from containerized cargo to coal moving from mine to railhead by truck and rail to harbor. The "Mail" component recognizes that shippers who use parcel delivery services typically do not know what modes were involved after the shipment was picked up. It also reports the foreign inbound and outbound mode used, if any.

Using the Data Extraction Tool (DET), the 2015 base year Internal-External, External-to-Internal and Internal-to-Internal Flows were extracted for the mega-region for each of the SCTG2 categories. Table 4-24 shows the 2015 Mega-Region Commodity Flow Summary by SCTG2 (in tons) using FAF4 Data. Similarly, Table 4-24 shows the 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in millions of dollars) using FAF4 Data.

Figure 4-15 shows the Mega-Region Annual Commodity Flow Summary by SCTG2 (in kilo tons). From the figure, SCTG19 (Coal and Petroleum Products) has the highest commodity flow volume for the Mega Region, in terms of tons. Figure 4-16 shows the Mega-Region Annual Commodity Flow Summary by SCTG2 (in million dollars). From the figure, SCTG35 (Electronic and Other Electrical Equipment and Components, and Office Equipment) has the highest commodity flow volume for the Mega Region, in terms of value.

Table 4-22 FAF Modes [24]

Code	Mode	Description
1	Truck	Includes private and for-hire truck. Does not include truck that is part of Multiple Modes and Mail or truck moves in conjunction with domestic air cargo.
2	Rail	Includes any common carrier or private railroad. Does not include rail that is part of Multiple Modes and Mail.
3	Water	Includes shallow draft, deep draft, Great Lakes and intra-port shipments. Does not include water that is part of Multiple Modes and Mail.
4	Air (includes truck-air)	Includes shipments move by air or a combination of truck and air in commercial or private aircraft. Includes air freight and air express. In the case of imports and exports by air, domestic moves by ground to and from the port of entry or exit are categorized with Truck.
5	Multiple Modes and Mail	Includes shipments by multiple modes and by parcel delivery services, U.S. Postal Service, or couriers (capped at 150 pounds). This category is not limited to containerized or trailer-on-flatcar shipments.
6	Pipeline	Includes crude petroleum, natural gas, and product pipelines. Note: Does include flows from offshore wells to land which are counted as Water moves by the U.S. Army Corps of Engineers. Does not include pipeline that is part of Multiple Modes and Mail.
7	Other and Unknown	Includes movements not elsewhere classified such as flyaway aircraft, and shipments for which the mode cannot be determined.
8	No Domestic Mode	Includes shipments that have an international mode, but no domestic mode and is limited to import shipments of crude petroleum transferred directly from inbound ships to a U.S. refinery at the zone of entry. This is done to ensure a proper accounting of import flows, while avoiding assigning flows to the domestic transportation network that do not use it.

[24] FAF4 User Guide: https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/FAF4_0%20User%20Guide.pdf

Table 4-23 2015 Mega-Region Commodity Flow Summary by SCTG2 (in tons) using FAF4 Data

SCTG2	SCTG2 Name	I-I (Ktons)	I-E (Ktons)	E-I (Ktons)	Total Ktons - 2015
1	Live Animals and Live Fish	88	317	17	422
2	Cereal Grains	1,350	81	1,221	2,652
3	Other Agricultural Products	4,699	1,982	848	7,529
4	Animal Feed and Products of Animal Origin, nec.	2,654	1,042	1,349	5,046
5	Meat, Fish, Seafood, and their Preparations	177	269	624	1,070
6	Milled Grain Products and Preparations, and Bakery Products	208	263	968	1,439
7	Other Prepared Foodstuffs and Fats and Oils	6,624	2,517	5,227	14,367
8	Alcoholic Beverages	1,302	433	791	2,526
9	Tobacco Products	5	2	6	12
10	Monumental or Building Stone	129	69	1	199
11	Natural Sands	5,591	3	62	5,657
12	Gravel and Crushed Stone	47,494	313	462	48,269
13	Nonmetallic Minerals, nec.	501	1,271	306	2,078
14	Metallic Ores and Concentrates	11,235	197	367	11,799
15	Coal	5,524	51	3,675	9,249
16	Crude Petroleum	0	11	0	11
17	Gasoline and Aviation Turbine Fuel	1,110	8	1,267	2,385
18	Fuel Oils	3,669	30	816	4,515
19	Coal and Petroleum Products, nec.	3,471	22,191	29,147	54,809
20	Basic Chemicals	2,201	1,092	1,061	4,353
21	Pharmaceutical Products	116	263	89	468
22	Fertilizers	425	211	1,934	2,570
23	Chemical Products and Preparations, nec.	190	227	835	1,252
24	Plastics and Rubber	604	660	1,397	2,661
25	Logs and Other Wood in the Rough	337	45	24	406
26	Wood Products	2,816	455	1,615	4,886
27	Pulp, Newsprint, Paper, and Paperboard	177	208	739	1,123
28	Paper or Paperboard Articles	472	228	483	1,183
29	Printed Products	92	49	263	404
30	Textiles, Leather, and Articles of Textiles or Leather	129	102	716	947
31	Nonmetallic Mineral Products	15,979	1,308	2,376	19,663
32	Base Metal in Primary or Semi-Finished Forms and in Finished Basic Shapes	1,595	503	1,879	3,977
33	Articles of Base Metal	713	609	985	2,307
34	Machinery	612	234	580	1,427
35	Electronic and Other Electrical Equipment and Components, and Office Equipment	985	662	880	2,527
36	Motorized and Other Vehicles (including Parts)	217	764	1,602	2,583
37	Transportation Equipment, nec.	33	31	84	148
38	Precision Instruments and Apparatus	46	76	75	197
39	Furniture, Mattresses and Mattress Supports, Lamps, Lighting Fittings, and Illuminated Signs	707	330	551	1,588
40	Miscellaneous Manufactured Products	526	394	894	1,813
41	Waste and Scrap	10,596	434	199	11,229
42	Miscellaneous Transported Products				
43	Mixed Freight	4,087	1,312	1,206	6,605
	Total:	**139,487**	**41,245**	**67,619**	**248,350**

Table 4-24 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in millions of dollars) using FAF4 Data

SCTG2	SCTG2 Name	I-I (M$)	I-E (M$)	E-I (M$)	Total Value (Million$) - 2015
1	Live Animals and Live Fish	217	859	30	1,105
2	Cereal Grains	303	30	443	775
3	Other Agricultural Products	4,624	2,089	1,636	8,349
4	Animal Feed and Products of Animal Origin, nec.	926	1,509	1,174	3,610
5	Meat, Fish, Seafood, and their Preparations	783	1,381	2,615	4,779
6	Milled Grain Products and Preparations, and Bakery	412	469	1,399	2,280
7	Other Prepared Foodstuffs and Fats and Oils	4,579	4,118	7,475	16,172
8	Alcoholic Beverages	2,605	267	1,357	4,230
9	Tobacco Products	224	19	139	382
10	Monumental or Building Stone	24	16	1	41
11	Natural Sands	83	2	25	110
12	Gravel and Crushed Stone	299	40	25	364
13	Nonmetallic Minerals, nec.	37	169	53	259
14	Metallic Ores and Concentrates	5,398	796	107	6,301
15	Coal	231	2	58	291
16	Crude Petroleum	0	7	0	7
17	Gasoline and Aviation Turbine Fuel	1,005	7	837	1,849
18	Fuel Oils	3,564	31	679	4,273
19	Coal and Petroleum Products, nec.	1,675	5,725	8,721	16,121
20	Basic Chemicals	866	1,346	1,528	3,741
21	Pharmaceutical Products	12,314	11,434	3,789	27,538
22	Fertilizers	211	78	920	1,209
23	Chemical Products and Preparations, nec.	906	984	3,572	5,462
24	Plastics and Rubber	2,160	2,087	4,901	9,148
25	Logs and Other Wood in the Rough	23	93	9	125
26	Wood Products	1,719	266	1,230	3,215
27	Pulp, Newsprint, Paper, and Paperboard	198	272	691	1,161
28	Paper or Paperboard Articles	797	425	946	2,169
29	Printed Products	525	298	1,207	2,030
30	Textiles, Leather, and Articles of Textiles or Leather	1,158	1,150	6,432	8,740
31	Nonmetallic Mineral Products	2,192	495	2,252	4,939
32	Base Metal in Primary or Semi-Finished Forms and	1,898	2,371	3,395	7,664
33	Articles of Base Metal	1,946	2,321	3,997	8,264
34	Machinery	5,655	4,880	11,374	21,909
35	Electronic and Other Electrical Equipment and	10,969	15,691	26,790	53,450
36	Motorized and Other Vehicles (including Parts)	2,322	8,268	12,241	22,831
37	Transportation Equipment, nec.	3,114	7,258	3,940	14,312
38	Precision Instruments and Apparatus	1,151	3,610	6,732	11,494
39	Furniture, Mattresses and Mattress Supports,	4,076	1,405	2,519	8,000
40	Miscellaneous Manufactured Products	2,260	13,225	11,012	26,497
41	Waste and Scrap	1,023	477	258	1,757
42	Miscellaneous Transported Products				
43	Mixed Freight	14,188	6,890	8,283	29,362
	Total:	**98,660**	**102,859**	**144,795**	**346,313**

Figure 4-15 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in kilo tons)

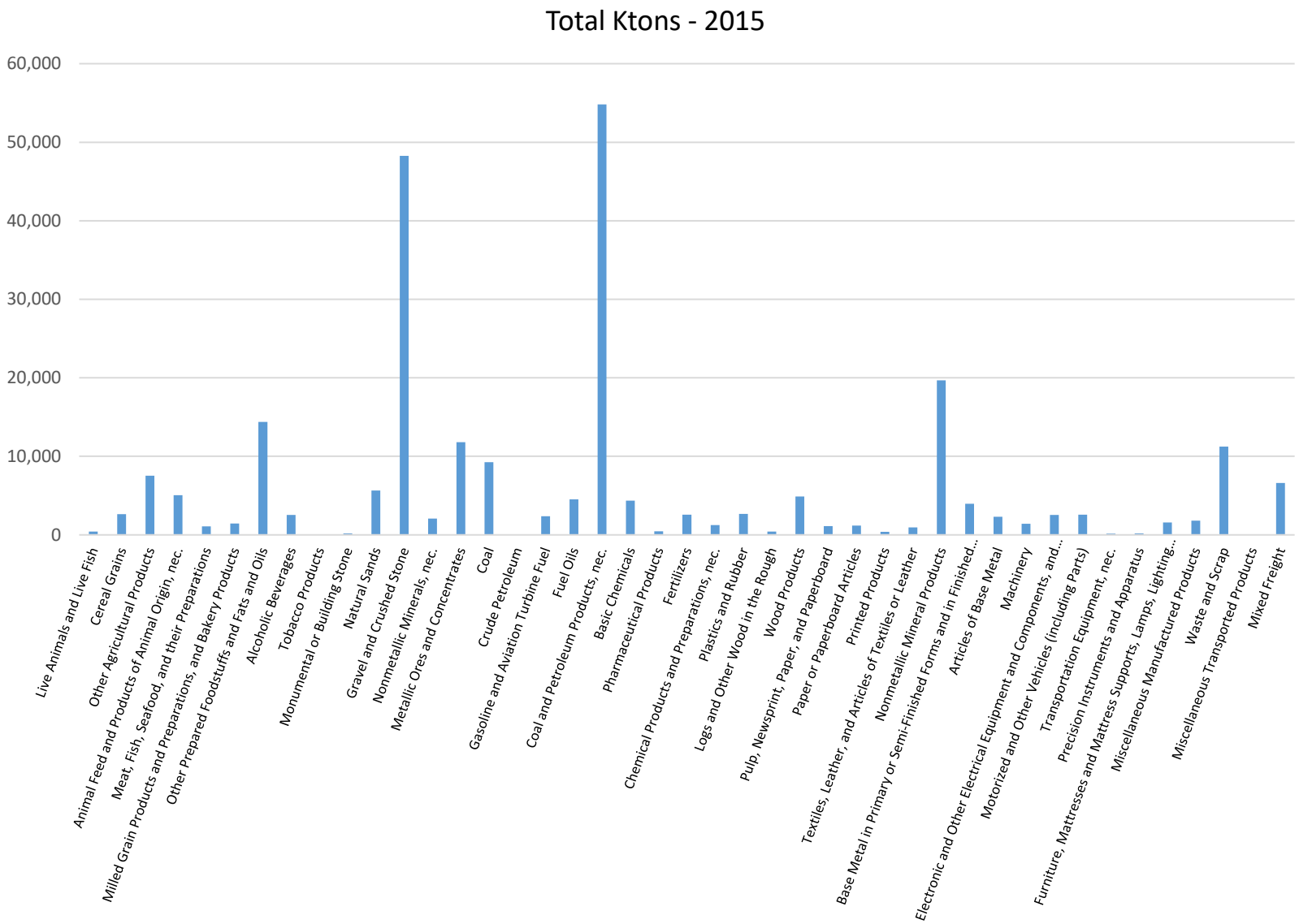
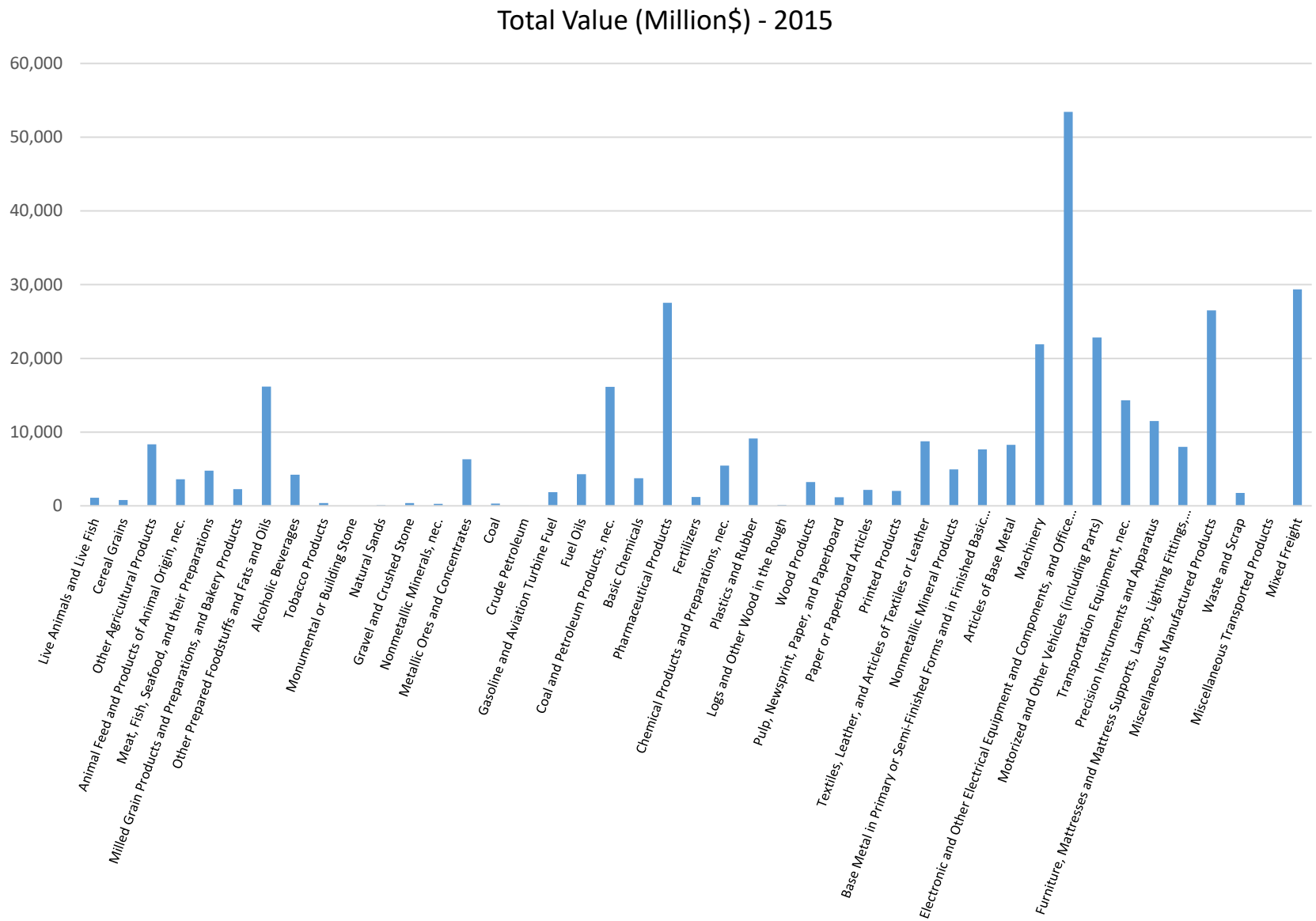


Figure 4-16 2015 Mega-Region Annual Commodity Flow Summary by SCTG2 (in Millions of dollars)



5 Firm Synthesis Model

Consultant Technical Lead –Srinath Ravulaparthy and Arun Kuppam (Cambridge Systematics)

MAG Technical Lead – Kyunghwi Jeon, Daehyun You, and Sreevatsa Nippani

Integrated land-use and transportation models address regional issues related to congestion, mobility, and economic competitiveness. Essential to this is freight transport, in the form of economic vitality along with distribution and complex interactions between various firms. Within this context, it is critical to capture and accurately describe the behavioral dynamics of these firms in the region including firm demographic events of birth (formation), death (dissolution), growth (expansion), decline (contraction), and migration (relocation) patterns. The study of firm demography is termed as firmographics that models the evolution of firms in time and space in a given region, capturing firm births, deaths, migration, and growth/decline. This chapter describes the methodology for modeling firm demographics using a microsimulation approach for the Phoenix-Tucson megaregion. The model is based on principles of firmographics that predicts the location, magnitude, and size of firms in the study region. This model uses the National Establishment Time Series (NETS) database as a seed table. A series of econometric models are estimated to simulate the firm events that consider determinants such as firm internal attributes (size, age, and growth) and external attributes (market area characteristics, transportation costs, agglomeration economies).

The simulated results are validated with observed firm demographic trends along with zonal-level employment estimated using various goodness-of-fit measures. For instance, using a 20% validation sample, the firm synthesizer accurately projects firm events of birth, death, and relocation from 2007 to 2012. In addition, total zonal employment predicted in the region also closely resembles the observed trend, thereby also capturing the observed spatial distribution of employment in the megaregion.

5.1 NETS Data Description

The main database used for the development of the firm synthesis model is the 2012 National Establishment Time Series (NETS) database that includes more than half a million business establishments in the study region of MAG and PAG, with longitudinal information about their industries, location, headquarters, and revenue over the period 1990 to 2012. The NETS database is constructed by taking a series of snapshots based on the Dun and Bradstreet (D&B) archival national establishment data that has been thoroughly quality controlled and updated [25]. In the database, the unit of observation is a business establishment that produces goods or services at a single physical location, e.g., a single store or a business establishment. NETS database tracks every business establishment from its birth through physical moves (or relocation) it makes, capturing any changes in ownership and recording establishment's death if it occurs. In the process the database also records information on location of the business, employment, sales, and industry type for each year from 1990 to 2012.

[25] Walls, D. (2007). National Establishment Time Series Database: Data Overview. Presented at 2007 Kauffman Symposium on Entrepreneurship and Innovation Data.

One of the highly desirable features of the NETS database is that it covers all the establishments in the study area, which reflects the fact that NETS is designed to capture the universe rather than a sample of establishments. Moreover, the NETS data also captures the business relocation that in contrast to other data sources that researchers have used to study business establishment and employment dynamics [26]. With all its advantages, NETS data also reports higher employment at the national and regional levels primarily because it covers more businesses than traditional sources such as ES-202 database. This is especially true for small businesses.

5.1.1 Firm Events Analysis

In this analysis, the focus was on studying the three lifecycle events of the firm beginning with birth, relocation, and death of the firm (or business establishment) in the study area. First, birth of a firm is defined as the beginning (or first) year in which the business establishment was recorded in the NETS database for the MAG and PAG study area. Second, relocation is defined as change in address (or physical location) from one location to another. Finally, firm death is defined as the last year in which the firm and its internal attributes were recorded in the panel from 1990 to 2012. We note that in this analysis we do not include businesses that were either formed or closed outside of the study area. Furthermore, to make the analysis tractable we classify the business establishments based on the two-digit North American Industrial Classification System (NAICS). The 16 types of industries we use to represent are: a) agriculture, forest, fishing and hunting [11]; b) mining [21]; c) utilities [22]; d) construction [23]; e) manufacturing [31-33]; f) wholesale trade [42]; g) retail trade [44-45]; h) transportation and warehousing [48-49]; i) information [51]; j) professional, scientific, management, administrative and waste management services [54-56]; k) health care services [62]; l) arts, entertainment, recreational, accommodation, and food services [71-72]; m) other services (expect public administration) [81]; n) finance, insurance, real estate, rental, and leasing (FIRE) [52 53]; o) public administration and armed forces [92]; and p) educational services [61].

As part of this analysis, Table 5-1 illustrates the evolving distribution of business establishments in MAG and PAG area from 1990 to 2012. For this purpose, the distribution is a function of births, deaths and migration that includes businesses moving into and outside of the study area. We note that the beginning year in the panel being 1990, the total number of business establishments (or survivors in 1990) includes all the firms as a result the events related to birth, death and migration are set to zero. Furthermore, the total number of business establishments F_t in year t (after 1990) is given as $F_t = S_{t-1} + B_t + I_t - D_t - O_t$. Additionally, we also represent in Figure 5-1 the evolving distribution of births, deaths, and overall business establishment population in the study area. Note that births and deaths presented in Figure 5-1 are represented as number of events (birth or death) per 1000 firms.

[26] Neumark, D., Wall, B. and Zhang, J. (2005). Business establishment dynamics and employment growth. Entrepreneurship Seminar Series. Hudson Institute Center for Employment Policy.

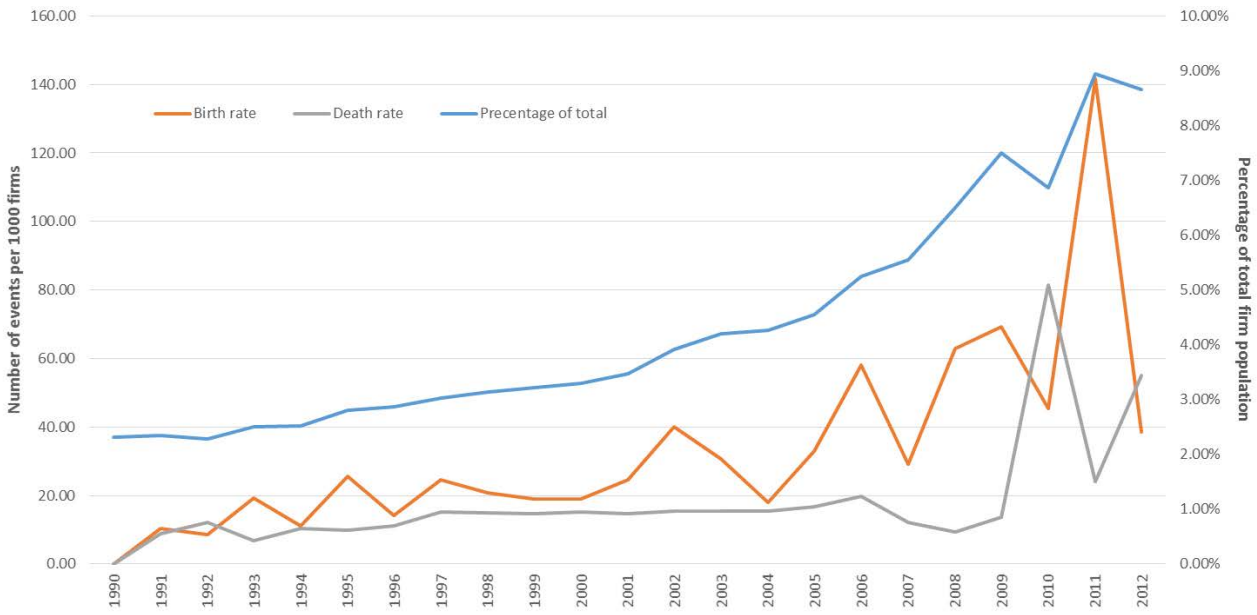
Table 5-1 Evolutionary Transition of Firm Population in MAG and PAG from 1990 to 2012

Year	Births (B)	Deaths (D)	Move-in (I)	Move-out (O)	Survivors (S)	Total (F)	Percentage of total
1990	0	0	0	0	130,884	130,884	2.31%
1991	10,488	8,800	285	164	130,884	132,693	2.34%
1992	8,531	12,076	207	116	132,693	129,239	2.28%
1993	19,346	6,796	294	156	129,239	141,927	2.51%
1994	11,196	10,338	333	139	141,927	142,979	2.53%
1995	25,669	9,800	293	119	142,979	159,022	2.81%
1996	14,134	11,234	289	149	159,022	162,062	2.86%
1997	24,705	15,231	229	143	162,062	171,622	3.03%
1998	20,836	14,846	204	176	171,622	177,640	3.14%
1999	19,072	14,718	283	144	177,640	182,133	3.22%
2000	19,103	15,319	454	210	182,133	186,161	3.29%
2001	24,712	14,598	755	349	186,161	196,681	3.48%
2002	39,978	15,327	763	360	196,681	221,735	3.92%
2003	30,618	15,531	1,108	466	221,735	237,464	4.20%
2004	18,047	15,337	1,405	616	237,464	240,963	4.26%
2005	32,943	16,762	1,102	441	240,963	257,805	4.56%
2006	58,158	19,806	879	485	257,805	296,551	5.24%
2007	29,239	12,156	779	465	296,551	313,948	5.55%
2008	62,895	9,262	1,232	676	313,948	368,137	6.51%
2009	69,350	13,576	1,309	950	368,137	424,270	7.50%
2010	45,332	81,365	1,399	1,026	424,270	388,610	6.87%
2011	141,577	24,142	1,330	967	388,610	506,408	8.95%
2012	38,620	55,151	0	0	506,408	489,877	8.66%

Note that birth and death rates are represented as number of events (birth or death) per 1000 firms.

Overall there is an increase in firm population in MAG and PAG region with an average growth rate of 4.35 percent (growth trend in blue line in Figure 5-1). The overall growth in firm population is further enhanced by persistent birth rate of business establishments with an annual average of 33 new born per 1000 firms (birth rate in orange line in Figure 5-1) for study period from 1990 to 2012. In contrast, rate of business establishment closures in MAG and PAG are fairly stable over the years with an annual average rate of 18 deaths per 1000 firms (death rate in gray line in Figure 5-1) in the population from 1990 to 2012. However, the most noticeable effect of "recession" is significant in year 2010 with high-closure rate of 81.37 per 1000 firms and low birth rate of 45 new born per 1000 firms.

Figure 5-1 Distribution of Birth and Death Rates of Business Establishments from 1990 to 2012



Furthermore, we also illustrate the firm behavioral dynamics of individual lifecycle events related to birth, death and relocation by the two-digit NAICS code as discussed earlier. These are presented below from Table 5-2 to Table 5-4 and Figure 5-2 to Figure 5-4.

Table 5-2 Distribution of Firm Births by Industry Types from 1991 to 2012

Year	Agriculture	Mining	Utilities	Construction	Manufacturing	Wholesale Trade	Retail trade	Transport	Information	FIRE	Professional Services	Educational Services	Healthcare Services	Arts and Accommodation	Other	Public Administration	Total
1991	114	17	2	1,266	603	668	1,329	211	168	995	2,013	124	636	1,062	1,246	34	10,488
1992	65	11	5	984	581	604	1,141	194	200	915	1,843	130	457	662	708	31	8,531
1993	102	40	23	1,443	1,030	1,365	2,955	377	431	2,113	4,058	338	1,777	1,313	1,845	136	19,346
1994	83	14	9	1,086	704	777	1,638	269	227	1,196	2,210	131	1,082	724	983	63	11,196
1995	114	30	20	2,065	1,144	1,856	4,260	470	476	2,654	5,666	353	2,137	1,465	2,909	50	25,669
1996	87	17	18	1,331	794	987	2,228	351	377	1,460	3,122	200	902	839	1,326	95	14,134
1997	157	36	25	2,450	1,206	1,738	4,268	490	586	2,592	5,118	270	1,528	1,554	2,598	89	24,705
1998	148	30	11	2,094	1,079	1,275	3,305	407	603	1,948	4,976	278	1,257	1,345	1,949	131	20,836
1999	151	12	23	1,931	985	1,047	2,847	421	631	1,789	4,473	287	1,457	1,209	1,701	108	19,072
2000	136	33	18	1,794	1,162	966	2,870	464	716	1,897	4,098	351	1,003	1,516	1,930	149	19,103
2001	176	23	14	2,679	1,168	1,137	3,693	592	782	2,646	5,465	363	1,684	1,565	2,501	224	24,712
2002	313	39	38	3,964	1,355	1,436	4,769	826	905	4,398	9,789	503	4,534	1,932	4,813	364	39,978
2003	450	25	27	2,626	1,076	1,066	2,858	598	1,005	4,285	10,890	288	1,418	1,430	2,438	138	30,618
2004	161	13	19	1,946	696	803	2,369	417	542	2,366	4,787	197	836	1,213	1,538	144	18,047
2005	274	28	21	2,931	1,003	1,037	3,314	690	718	4,414	12,915	308	1,486	1,347	2,276	181	32,943
2006	469	33	41	5,180	1,456	1,638	4,453	1,074	1,199	8,221	24,874	536	3,003	1,985	3,723	273	58,158
2007	232	23	32	2,689	840	813	2,296	655	677	3,875	10,766	275	1,743	2,187	1,942	194	29,239
2008	402	25	30	5,144	1,362	2,162	5,255	1,638	1,250	7,995	27,287	443	4,348	1,588	3,800	166	62,895
2009	513	26	30	4,394	1,105	2,165	3,718	1,460	1,099	6,306	40,214	311	2,718	1,379	3,741	171	69,350
2010	423	32	30	3,468	854	1,681	3,725	902	851	4,724	20,544	288	2,539	1,340	3,813	118	45,332
2011	1,499	38	71	11,943	2,632	5,984	11,262	3,485	3,542	18,919	57,637	657	6,990	3,108	13,512	298	141,577
2012	434	23	47	2,615	875	1,713	3,020	1,074	978	6,400	11,871	602	2,724	1,141	4,454	649	38,620

Table 5-3 Distribution of Firm Deaths by Industry Types from 1991 to 2012

Year	Agriculture	Mining	Utilities	Construction	Manufacturing	Wholesale Trade	Retail trade	Transport	Information	FIRE	Professional Services	Educational Services	Healthcare Services	Arts and Accommodation	Other	Public Administration	Total
1991	119	25	8	1,193	515	655	1,606	150	132	944	1,300	39	869	448	794	3	8,800
1992	171	25	11	1,564	697	797	2,099	255	180	1,540	1,769	107	823	655	1,370	13	12,076
1993	98	20	10	886	436	502	1,209	144	93	865	919	74	485	462	589	4	6,796
1994	206	32	8	1,280	771	700	1,671	175	157	1,143	1,763	126	786	499	965	56	10,338
1995	110	19	5	1,421	502	620	1,461	201	151	969	1,612	64	576	1,073	1,010	6	9,800
1996	107	25	10	1,245	710	949	1,904	201	177	1,282	2,071	103	665	863	899	23	11,234
1997	133	32	23	1,432	835	1,259	2,370	295	340	1,850	3,090	161	1,242	719	1,427	23	15,231
1998	106	11	14	1,273	733	1,152	2,344	287	288	1,698	3,099	152	1,300	701	1,666	22	14,846
1999	96	28	18	1,371	853	1,184	2,407	277	311	1,554	3,050	160	1,018	925	1,432	34	14,718
2000	118	19	19	1,420	799	1,043	2,633	305	397	1,565	3,412	144	1,022	807	1,572	44	15,319
2001	115	29	19	1,429	820	990	2,401	312	372	1,548	3,128	175	900	788	1,521	51	14,598
2002	103	23	9	1,710	883	1,009	2,537	335	528	1,588	3,279	182	794	832	1,489	26	15,327
2003	107	30	22	1,711	845	950	2,384	361	414	1,817	2,936	180	1,178	747	1,669	180	15,531
2004	126	17	20	1,440	878	838	2,188	327	489	1,627	3,657	192	1,049	800	1,528	161	15,337
2005	139	21	14	1,608	773	795	2,590	388	554	1,754	3,398	221	1,140	1,120	2,164	83	16,762
2006	321	27	20	1,727	830	866	2,107	455	686	2,661	6,126	160	1,306	922	1,409	183	19,806
2007	110	18	7	1,128	571	563	1,450	237	377	1,387	3,146	156	898	776	1,242	90	12,156
2008	63	12	8	698	460	393	1,031	213	256	1,119	1,823	161	795	1,167	857	206	9,262
2009	91	10	16	1,158	528	653	1,777	287	405	1,946	2,981	155	935	1,498	1,092	44	13,576
2010	655	64	72	7,201	2,804	3,296	9,044	1,923	2,001	10,435	28,478	609	4,608	3,372	6,574	229	81,365
2011	187	30	36	2,669	1,089	1,351	2,984	610	531	3,051	6,355	236	1,569	1,206	2,180	58	24,142
2012	433	41	35	6,226	1,547	2,732	4,544	1,103	1,402	6,609	22,060	295	2,557	1,683	3,752	132	55,151

Table 5-4 Distribution of Firm Relocations by Industry Types from 1991 to 2011

Year	Agriculture	Mining	Utilities	Construction	Manufacturing	Wholesale Trade	Retail trade	Transport	Information	FIRE	Professional Services	Educational Services	Healthcare Services	Arts and Accommodation	Other	Public Administration	Total
1991	31	16	9	474	280	331	415	69	65	354	588	56	164	158	209	8	3,227
1992	16	10	2	422	266	294	363	63	61	344	534	18	163	117	195	5	2,873
1993	21	12	6	502	300	378	403	66	85	375	690	32	190	107	176	22	3,365
1994	19	16	8	443	321	313	381	59	78	377	614	30	163	119	177	14	3,132
1995	16	13	11	445	339	358	401	79	78	414	711	43	216	100	224	10	3,458
1996	17	9	5	516	312	351	402	65	95	405	788	41	229	125	211	14	3,585
1997	19	14	7	588	352	413	486	83	95	453	951	57	286	188	299	15	4,306
1998	18	11	6	575	340	439	428	75	92	456	973	35	222	154	295	17	4,136
1999	25	6	3	491	279	306	403	71	97	399	887	65	256	116	270	10	3,684
2000	32	11	2	661	358	386	510	110	132	477	1,051	56	230	148	325	14	4,503
2001	39	13	2	749	432	434	737	134	149	544	1,377	82	299	224	443	18	5,676
2002	58	9	3	710	385	384	640	119	152	620	1,290	89	379	257	444	30	5,569
2003	82	8	4	989	457	440	807	161	234	842	2,084	94	520	310	659	33	7,724
2004	54	13	10	993	504	461	929	165	225	799	2,058	142	532	370	695	28	7,978
2005	40	12	8	812	458	429	667	142	168	565	1,622	86	380	238	467	27	6,121
2006	40	9	4	770	384	324	540	147	162	710	1,484	81	413	243	427	30	5,768
2007	38	8	7	732	340	321	534	118	142	631	1,485	80	408	230	440	22	5,536
2008	93	8	4	745	375	364	661	156	162	706	1,649	107	533	233	464	15	6,275
2009	70	11	12	916	410	398	693	216	194	797	3,372	94	523	242	528	17	8,493
2010	75	10	15	844	414	379	594	184	156	773	3,658	115	563	247	474	14	8,515
2011	75	6	15	802	374	357	633	196	191	881	3,805	88	545	239	500	11	8,718

Figure 5-2 Composition of Births by Industry Types from 1991 to 2012

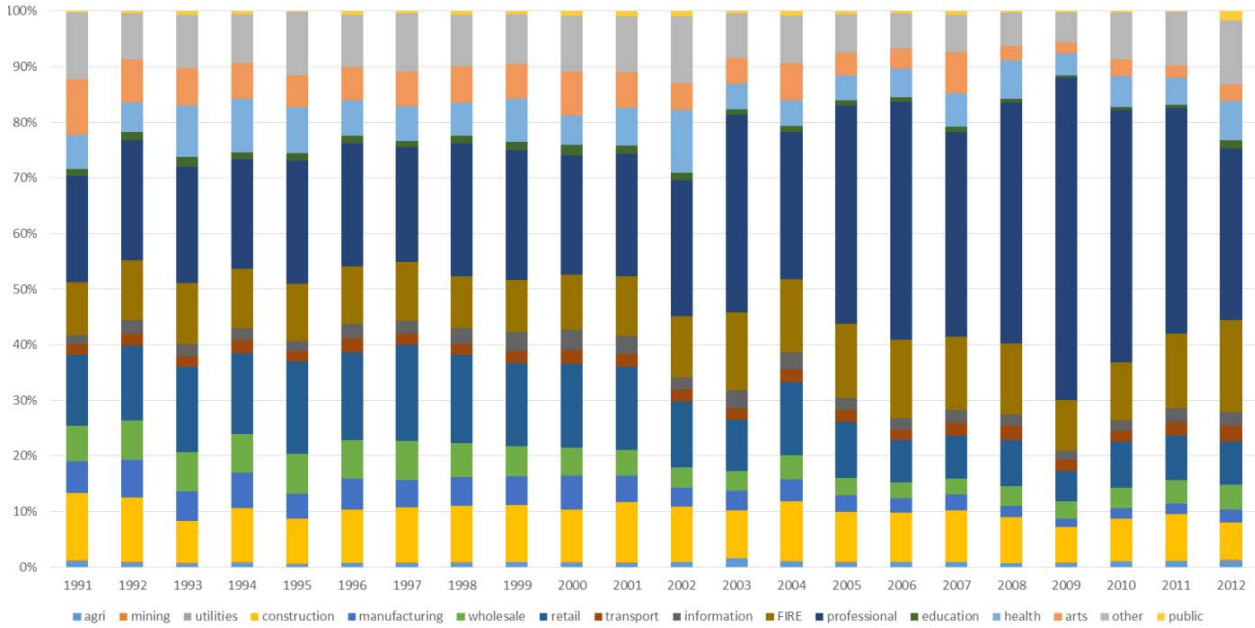


Figure 5-3 Composition of Deaths by Industry Types from 1991 to 2012

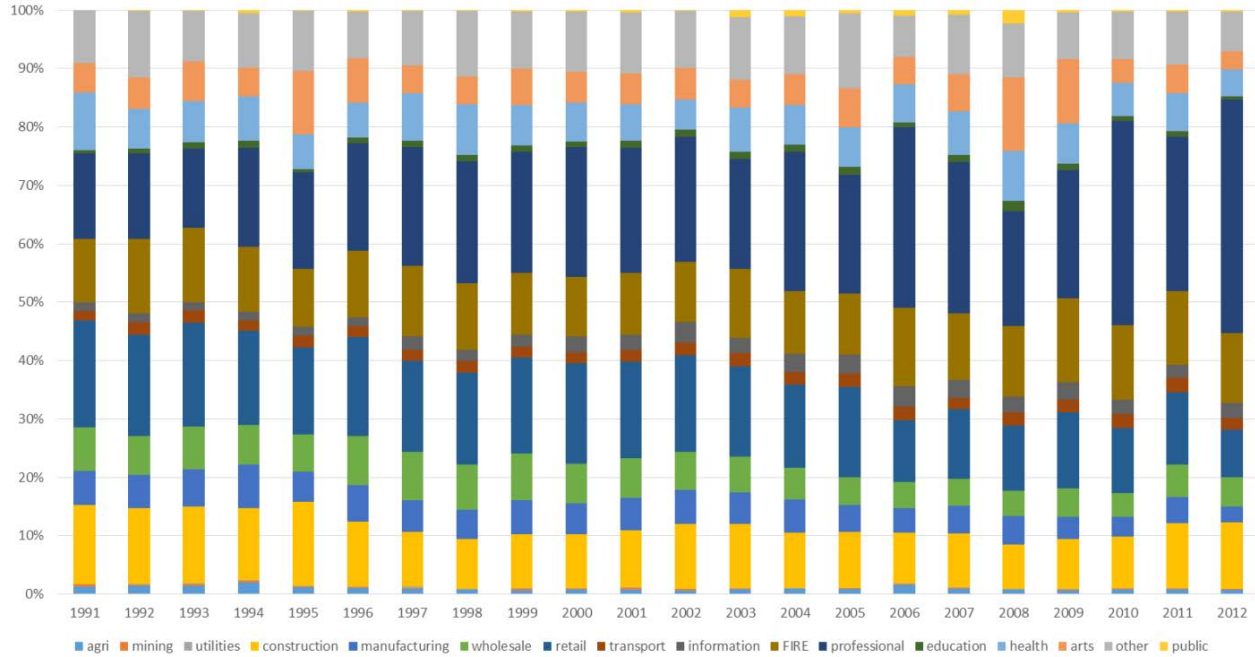
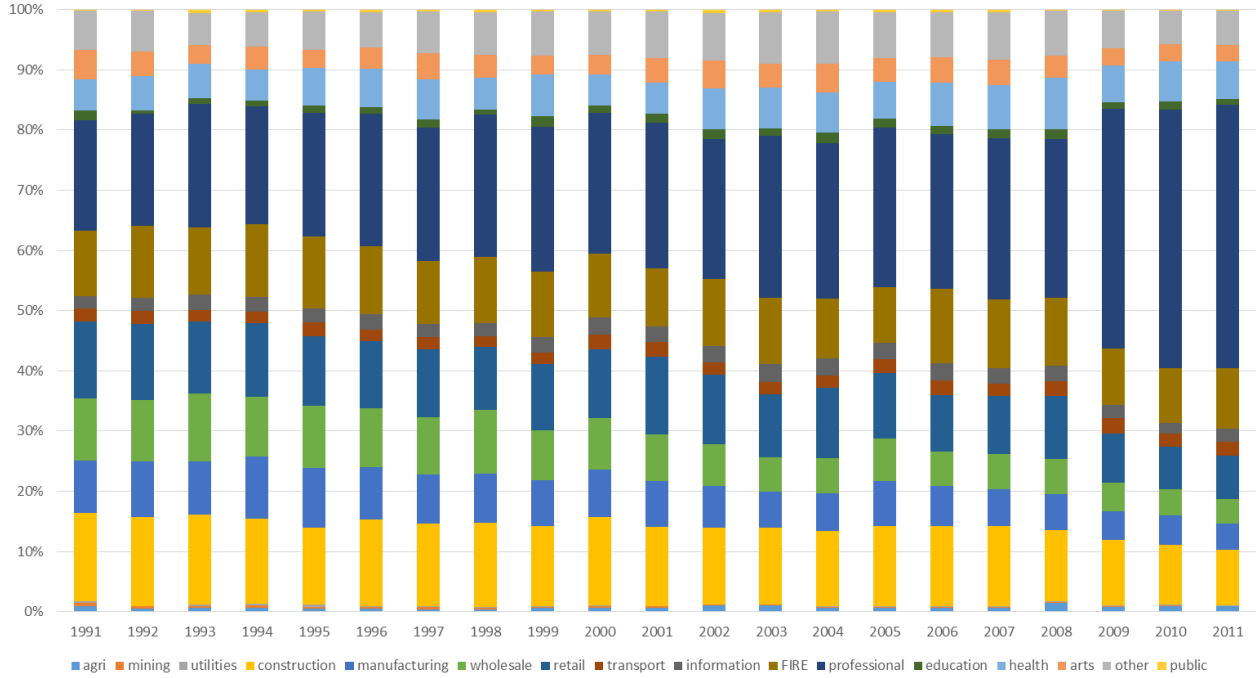


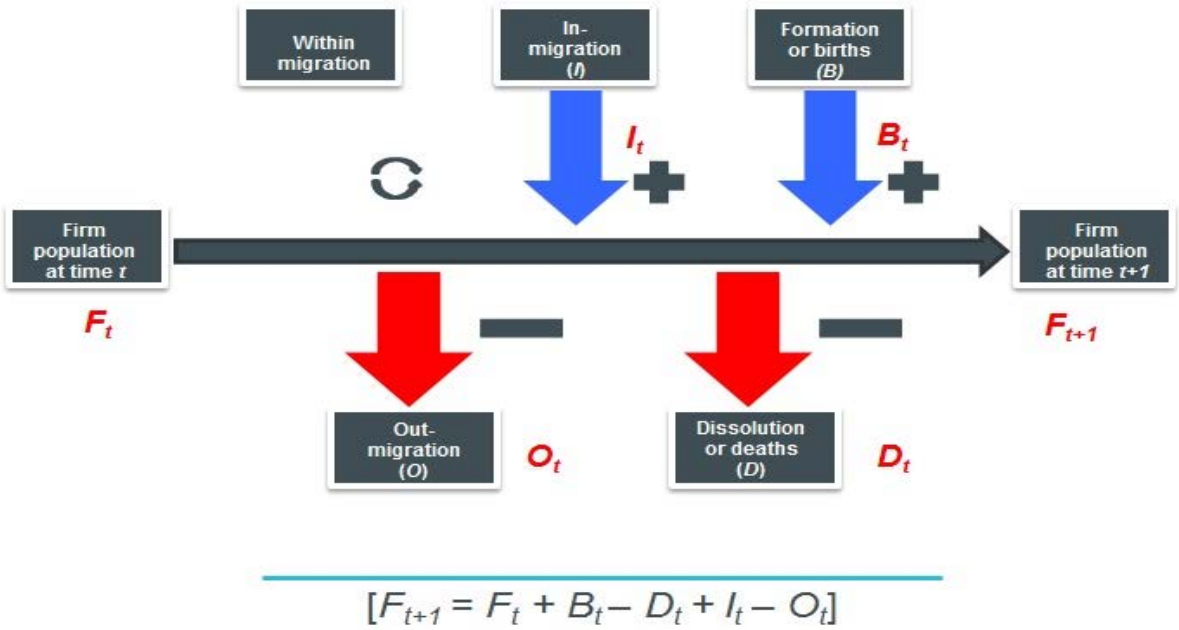
Figure 5-4 Composition of Relocations by Industry Types from 1991 to 2011



5.2 Model Specification and Estimation Results

The basic concept of a firm synthesizer model is to capture the evolution of firm population in time and space. It involves modeling a series of events related to firm births, deaths, and migration as depicted in Figure 5-5.

Figure 5-5 Overview of Firm Evolution



In order to analyze these firm dynamics related to birth, death, mobility and growth, we use a series of regression models to predict the probability of observing each of the firm event. For this purpose, we divide this discussion into adopted modeling structure and their respective firm event within the MAG and PAG region. Furthermore, we also note that all the model estimation results are based from the NETS observed for years 2007 to 2011.

5.2.1 Firm Dissolution and Relocation

To capture the firm behavioral dynamics related to dissolution and relocation, we use binary probit panel regression model to predict the probability of firm being dissolved (or closed) or relocating from location *l* at time *t*. The model formulation for these events is illustrated in the equation below.

$$y_{it}^{e*} = \alpha' f_{it} + \beta' s_{it} + \gamma' z_{it} + \varepsilon_{it} \quad (1)$$

Where,

y_{it}^{e*} is the latent measure of observing an event *e*, in this case *e* takes the events of dissolution (or closure) and relocation respectively.

f_{it} – is a vector of explanatory variables describing firm's internal attributes *i* as related to employment, age and dummy variable of subsidiary and change in employment for time *t*.

s_{it} – is a vector of dummy variables indicating if the firm *i* is of a specific industry type as defined by the two-digit NAICS code as observed for time *t*.

z_{it} – is a vector of explanatory variables describing firm's *i* external attributes related to zonal level information of employment by two-digit NAICS and student enrollment numbers as observed for time *t*.

ε_{it} – random error term assumed to have a standard logistic distribution.

a, β, γ – vector of parameters estimated using maximum likelihood method.

5.2.2 Firm Growth

We predict employment size as a proxy to understand the trajectory of firm growth in the region. For this purpose, log-linear model specification to estimate employment size of the firm *i* for time *t* is illustrated as in equation 1.

$$\log(e_{it}) = \alpha' \log(e_{i,t-1}) + \beta' f_{it} + \gamma' s_{it} + \theta' z_{it} + \varepsilon_{it} \quad (2)$$

Where,

e_{it} is the employment size (or number of employees) of the firm *i* for time period *t* and $e_{i,t-1}$ is the employment size of the firm *i* observed for time period *t-1*

f_{it} – is a vector of explanatory variables describing firm's internal attributes *i* as related to age and dummy variable of subsidiary for time *t*.

s_{it} – is a vector of dummy variables indicating if the firm *i* is of a specific industry type as defined by the two-digit NAICS code as observed for time *t*.

$\tilde{z}_{it}$ – is a vector of explanatory variables describing firm's *i* external attributes related to zonal level information of employment by two-digit NAICS and student enrollment numbers as observed for time *t*.

ε_{it}– random error term assumed to be normal distributed.

a, β, γ, θ – vector of parameters estimated using ordinary least squares (OLS) technique.

5.2.3 Firm Location Choice

Location choice model is a long-term decision for newborn and migrating firms in the region. We formulate a Multinomial Logit Model based on Random Utility Model framework as proposed by McFadden (1978).

Given previous location, with traffic analysis zones (TAZs) as choice alternatives *i,* for each firm *n,* this utility U_{in} for each alternative $i=1, 2,\ldots, j$ for an individual firm *n* is given by the function form as shown below

$$U_{in} = V_{in} + \varepsilon_{in} = X_n A_{in}\beta_i + A_{in}\alpha_i + \varepsilon_{in} \quad (3)$$

Where the systematic part $V_{in} = X_n A_{in}\beta_i + A_{in}a_i$ accounts *for* location choice attributes A_i and an interaction between location and firm characteristics X_n, the coefficients *a* and *β* are estimated using maximum likelihood estimation techniques.

The variables included in the specified model are attributes of location choices (or TAZs) A_i such as – employment by NAICS industry type, population density and miles of freeways, arterials and collector roads. A selection of these location attributes also are interacted with individual firm characteristics $X_n A_i$ to capture observed sources of heterogeneity across firms for their response to location attributes.

5.2.4 Model Estimation Results

In this section, we summarize the model estimation results and inference for key lifecycle events of growth, dissolution, relocation and location choice behavior of firms in the study region of MAG and PAG. Presented in from Table 5-5 through Table 5-10 are estimation results along with corresponding coefficient estimates and goodness of fit measures.

For instance in the growth model (shown in Table 5-5), from coefficient estimates related to age and percentage change in employment support the hypothesis that – older and mature firms tend to grown slow and reach saturation. In addition, path dependency also triggers growth, with lagged employment (previous year employment) having a positive impact. Although, the negative coefficient across all industry types (by NAICS two digit) implies slower growth, there certainly exist significant differences in magnitude for each of the industry type. For example, firms in healthcare, food and accommodation and professional services have higher magnitude of growth when compared with other industry types in the region.

As shown in Table 5-6, firm death model estimation results suggest that probability of firm failure (or dissolution or death or closure) is primarily determined by firm internal characteristics. For instance, mature (or older) and larger firms have higher likelihood of survival in the region. Older and larger firms are expensive to close down while having significant ramifications in the region's economy. Additionally, variation of firm closures is significantly different across the industry types, as some sectors are more

dynamic than others, e.g., construction and food and accommodation services have higher likelihood of failure in the region.

Unlike other firm lifecycle events models estimated earlier, birth of new firms is a complex process. As a result, we simulate firm birth based on the approach to quantify number of births to the existing population of firms, which leads to the concept of firm birth rate. This approach assumes that firm birth rate follows a negative binomial distribution, with mean μ and γ representing the dispersion parameter. For this effort, we estimate firm birth rates individually by each industry type as defined by NAICS two-digit and three-digit NAICS codes for manufacturing and wholesale trade only. The parameters are estimated using likelihood techniques as fitted for observed firm birth rates using NETS data from 1991 to 2011. The parameter estimates are tabulated in detail as shown in Table 5-10.

Table 5-5 Firm Growth Model Estimation Results (Calibrated)

Variable	Type	Coefficient	t-statistic
Intercept	Constant	0.15	71.579
Log of previous year's employment	Numeric	0.974	5223.180
Is subsidiary	Numeric	0.039	17.551
Percent of change in employment from 2 years ago to last year	Numeric	-1.76E-04	-2.562
Age of Company	Numeric	-0.001	-48.032
Agriculture	Binary	-0.035	-16.442
Construction	Binary	-0.013	-10.261
Manufacturing II	Binary	-0.026	-11.301
Wholesale	Binary	-0.025	-16.031
Retail I	Binary	-0.025	-18.098
Retail II	Binary	-0.026	-17.034
Transportation	Binary	-0.013	-7.785
Transportation – Postal	Binary	-0.019	-6.581
Information	Binary	-0.021	-12.454
Finance and Insurance	Binary	-0.039	-29.457
Real Estate	Binary	-0.030	-22.376
Professional/Scientific/Technical	Binary	-0.024	-20.104
Management	Binary	-0.051	-24.497
Administrative/Support/Waste Management	Binary	-0.015	-12.947
Health Care	Binary	-0.009	-6.770
Arts/Recreation	Binary	-0.033	-18.967
Accommodation/Food	Binary	-0.018	-10.579

Variable	Type	Coefficient	t-statistic
Other Services, (except Public)	Binary	-0.022	-16.534
TAZ Emp: Construction	Numeric	4.75E-05	15.54
TAZ Emp: Manufacturing II	Numeric	8.34E-05	6.86
TAZ Emp: Manufacturing III	Numeric	4.31E-06	2.28
TAZ Emp: Wholesale	Numeric	3.46E-05	9.59
TAZ Emp: Retail I	Numeric	1.63E-05	5.96
TAZ Emp: Retail II	Numeric	4.35E-05	8.62
TAZ Emp: Transportation	Numeric	8.83E-06	1.84
TAZ Emp: Transportation – Postal	Numeric	1.01E-04	3.37
TAZ Emp: Information	Numeric	6.04E-05	6.21
TAZ Emp: Finance and Insurance	Numeric	1.37E-05	6.38
TAZ Emp: Real Estate	Numeric	4.06E-05	6.96
TAZ Emp: Professional/Scientific/Technical	Numeric	7.92E-06	5.98
TAZ Emp: Administrative/Support/Waste Management	Numeric	2.84E-05	11.07
TAZ Emp: Arts/Recreation	Numeric	8.34E-05	5.86
TAZ Emp: Accommodation/Food	Numeric	3.77E-05	8.34
TAZ Emp: Other Services, (except Public)	Numeric	4.53E-05	5.93
TAZ Emp: Public Administration	Numeric	1.14E-05	6.03
Miles of freeway	Numeric	5.97E-04	5.35
Wholesalers, NonDurable	Binary	4.17E-03	2.35
Textiles	Binary	-2.77E-02	-5.91
Leather/Apparel Manu	Binary	-2.89E-02	-5.67
Petro/Chem Manu	Binary	2.88E-03	0.69
Collector miles, Retail II	Numeric	-1.75E-03	-2.27
Manu III, NAICS 332, 333, 334, 336, 339	Binary	-2.00E-02	-11.83
Population density	Numeric	-1.22E-06	-18.59
R-squared			**0.958**
Number of Observations			**382,114**

Table 5-6 Firm Dissolution Model Estimation Results (Calibrated)

Variable	Type	Coefficient	t-statistic
Intercept	Constant	-1.7	-254.3309
Log of employment	Numeric	-0.1779	-49.5325
Is subsidiary	Binary	0.4003	9.2958
Age of Company	Numeric	-0.0351	-84.7079
Population	Numeric	-2.55E-05	-19.3233
Construction	Binary	0.1178	11.0502
Manufacturing II (NAICS 32)	Binary	0.2201	7.6338
Manufacturing III (NAICS 33)	Binary	0.0523	1.7310
Wholesale	Binary	0.2111	10.8551
Retail I (NAICS 44)	Binary	0.1763	13.7703
Retail II (NAICS 45)	Binary	0.1178	7.4524
Real Estate	Binary	-0.0842	-6.3941
Management	Binary	-0.1756	-5.8484
Administrative/Support/Waste Management	Binary	-0.2640	-30.0011
Educational	Binary	-0.3361	-10.4238
Health Care	Binary	-0.0344	-2.6236
Arts/Recreation	Binary	0.0348	1.5852
Accommodation/Food	Binary	0.3800	21.2171
Textile products (NAICS 314)	Binary	-0.0639	-0.6378
Apparel Manu (NAICS 315)	Binary	0.1222	1.2389
Wholesalers, NonDurable (NAICS 424)	Binary	-0.0186	-0.6600
Percent of change in employment	Numeric	-0.0001	-0.0483
Emp Change: Retail I/II	Numeric	-0.0254	-1.4252
Emp Change: Professional/Scientific/Technical	Numeric	-0.0469	-2.1933
TAZ Emp: Wholesale	Numeric	-4.62E-05	-0.7495
TAZ Emp: Administrative/Support/Waste Management	Numeric	0.0006	18.4410
TAZ Emp: Other Services, (except Public)	Numeric	0.0005	4.8869
Miles of collector	Numeric	-0.0297	-10.2402
Fwy miles, Wholesale/Retail	Numeric	0.0053	1.1990
Manu: Metal/Electronics/Furniture (NAICS 332)	Binary	0.0733	1.8120
Pseudo R squared			**0.0168**
Number of Observations			**1,469,213**

Table 5-7 Firm Relocation Model Estimation Results (Calibrated)

Variable	Type	Coefficient	t-statistic
Intercept	Constant	-5.107	-18.355
Log of employment	Numeric	0.079	5.454
Age of Company	Numeric	-0.034	-16.998
Agriculture	Binary	1.483	3.592
Utility	Binary	1.530	2.608
Construction	Binary	1.794	4.692
Wholesale	Binary	1.768	4.601
Information	Binary	1.800	4.628
Finance and Insurance	Binary	1.482	3.867
Real Estate	Binary	1.400	3.645
Professional/Scientific/Technical	Binary	1.870	4.909
Management	Binary	0.964	2.291
Administrative/Support/Waste Management	Binary	1.666	4.378
Educational	Binary	1.223	3.033
Health Care	Binary	1.500	3.917
All Manufacturing (NAICS 31-33)	Binary	1.874	4.860
Retail I or II (NAICS 44-45)	Binary	1.276	3.336
Transportation and Warehousing (NAICS 48-49)	Binary	1.742	4.482
Arts/Food/OtherServices (NAICS 71, 72, 81)	Binary	1.213	3.168
Emp Change: Transportation – Postal	Numeric	-0.009	-0.034
Emp Change: Administrative/Support/Waste Management	Numeric	0.003	0.166
Emp Change: Public admin	Numeric	-0.910	-0.244
TAZ Emp: Construction	Numeric	6.18E-04	3.48
TAZ Emp: Manufacturing II	Numeric	6.75E-04	0.98
TAZ Emp: Manufacturing III	Numeric	-9.67E-05	-0.62
TAZ Emp: Other Services, (except Public)	Numeric	6.13E-04	0.97
Miles of collector	Numeric	-3.39E-02	-2.32
Miles of freeway	Numeric	2.23E-02	2.66
Pseudo R squared			**0.01**
Number of Observations			**1,331,852**

Table 5-8 Firm Move-In Location Choice Model Estimation Results (Calibrated)

Variable	Type	Coefficient	t-statistic
Arterial miles, Trans/Wholesale	Numeric	0.0096	1.30
Miles of collector	Numeric	0.0247	2.50
Miles of freeway	Numeric	-0.1030	-9.10
Fwy miles, Retail	Numeric	0.0919	2.70
Employment Density	Numeric	-2.10E-04	-30.50
Population Density, Food/Accom.	Numeric	-8.80E-05	-2.70
Population Density, Retail	Numeric	-1.10E-04	-4.70
Population Density	Numeric	4.10E-07	0.10
Has Univ., Health	Numeric	-0.057	-0.20
Has Univ., Arts/Rec	Numeric	1.130	3.70
Has Service TAZ emp - Service only	Numeric	4	N/A
Log Service TAZ emp - Service only	Numeric	0.1	N/A
Size Variables			
Pop, Food/Accom.	Numeric	-0.048	-0.10
Pop, Retail	Numeric	-0.290	-1.10
Agri/Nat. Resc.. TAZ emp., Agri/Nat. Resc. only	Numeric	3.220	7.40
Food/Accom. TAZ emp., Food/Accom. only	Numeric	2.870	4.70
Professional TAZ emp., Professional only	Numeric	2.890	21.40
Retail TAZ emp., Retail only	Numeric	1.240	1.70
Trans/Wholesale TAZ emp., Trans/Wholesale only	Numeric	2.210	16.40
Util/Construction TAZ emp., Util/Construction only	Numeric	2.320	7.90
Pseudo R squared			**0.036**
Number of Observations			**4,518**

Table 5-9 Firm Move-Within Location Choice Model Estimation Results (Calibrated)

Variable	Type	coefficient	t-statistic
Miles of Arterial		-0.011	-2.40
Arterial miles, Trans/Wholesale		0.014	1.90
Miles of collector		-0.069	-8.40
Miles of freeway		-0.013	-3.00
Fwy miles, Manuf.		0.043	2.70
Population Density, Manuf.		-5.20E-05	-4.40
Population Density, Professional		-1.60E-05	-4.10
Population Density, Retail		-3.60E-05	-4.60
Population Density, Trans/Wholesale		-5.60E-05	-11.00
Population Density, Util/Construction		-7.30E-05	-10.10
Agri/Natural Resources. TAZ emp. dens., Agri/Nat. Rsrc. only		-5.30E-05	-2.00
Edu/Health TAZ emp. dens., Edu/Health only		-1.50E-05	-5.00
Food/Accom. TAZ emp. density., Food/Accom. only		2.70E-06	1.50
Manuf. TAZ emp. dens., Manuf. only		-1.10E-05	-2.40
Professional TAZ emp. dens., Professional only		-6.80E-06	-6.20
Retail TAZ emp. dens., Retail only		6.60E-07	0.40
Trans/Wholesale TAZ emp. dens., Trans/Wholesale only		-2.70E-05	-8.50
Util/Construction TAZ emp. dens., Util/Construction only		-1.50E-05	-4.50
Has Univ., Accom.		-0.381	-1.80
K-12 enroll, Edu		-0.00027	-3.90
Size Variables			
Population		-0.025	-0.50
Agri/Nat. Rsrc. TAZ emp., Agri/Nat. Rsrc. only		4.920	14.60
Edu/Health TAZ emp., Edu/Health only		3.340	35.40
Food/Accom. TAZ emp., Food/Accom. only		2.620	12.60
Manuf. TAZ emp., Manuf. only		3.430	27.70
Professional TAZ emp., Professional only		3.950	66.70
Retail TAZ emp., Retail only		2.480	19.80
Trans/Wholesale TAZ emp., Trans/Wholesale only		2.720	31.20
Util/Construction TAZ emp., Util/Construction only		2.900	28.80
Pseudo R squared			**0.082**
Number of Observations			**25,526**

Table 5-10 Firm Birth Rate Estimates by Industry Type

NAICS	Mean	Dispersion Parameter
11	295.59	1.72
21	25.82	11.34
22	25.18	3.00
23	3001.05	2.83
44	2079.50	4.57
45	1446.55	3.97
48	637.32	1.86
49	138.36	3.49
51	816.50	2.40
52	2188.45	1.62
53	1998.27	2.07
54	4594.91	2.13
55	359.09	0.47
56	7528.55	0.79
61	328.77	5.44
62	2102.68	2.53
71	714.41	5.01
72	735.77	6.80
81	2988.45	2.47
92	125.41	2.27
311	75.09	2.92
312	7.91	3.43
313	10.41	16.80
314	27.68	5.01
315	24.36	24.16
316	10.23	1.44
321	45.05	4.74
322	9.55	6.59
323	125.27	7.74
324	6.82	8.60
325	44.95	8.68
326	19.68	10.36
327	50.09	7.68
331	17.45	2.39
332	94.64	12.63
333	71.95	15.96
334	88.00	11.26
335	28.95	6.29
336	37.41	17.83
337	59.45	7.39
339	222.77	3.97
423	934.73	5.04
424	561.55	2.10

5.3 Simulation Validation Results

5.3.1 Application on Validation Sample

The portion of the NETS data which was left out of estimation was used to do the initial validation the model. This smaller set comprised 20 percent of the firms. They were randomly selected to avoid introducing bias into either set. By removing these firms from the estimation, the simulation could be tested against independent data that had not been a part of the model's creation. This has the built in advantage of comparing data sets with the exact same variables and methods of collecting information.

The primary validation exercise involved looking at totals across NAICS industries for several variables: firms, employment, deaths, average employment per firm and moves. The simulation was seeded with the 2007 validation set firms and run for 4 years, up to 2011. Analysis showed reasonable agreement between the simulation output – 2011 simulated firms – and the NETS 2011 firms. A number of runs were done to tease out variances due to the randomness inherent to such a model.

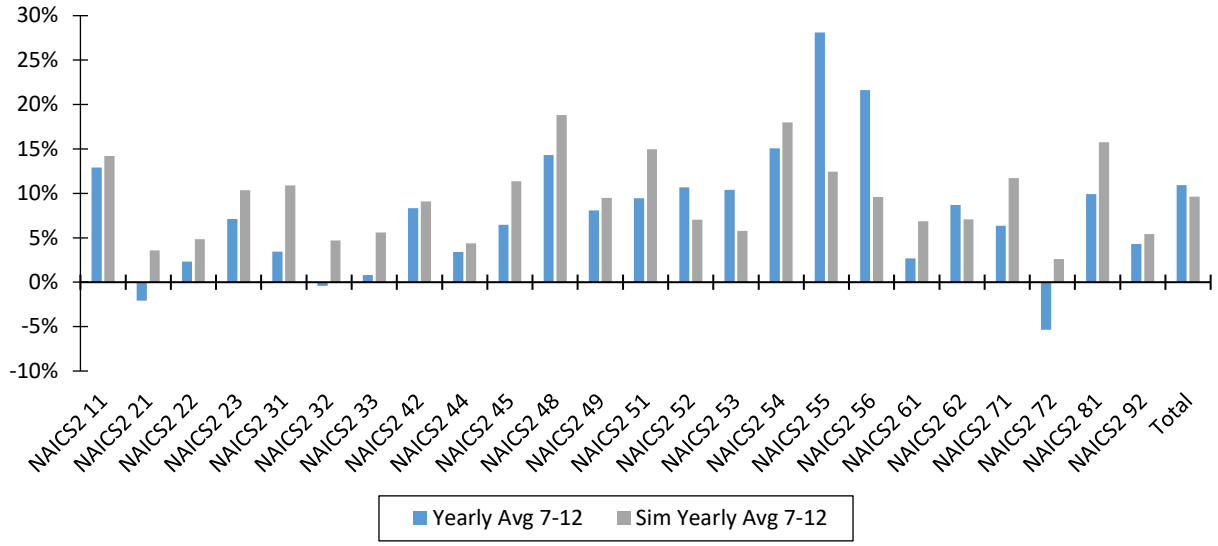
Overall, the simulation performed well. Small adjustments were made to the submodels' constants to match the observed more closely. Table 5-11 shows the results of the comparisons. The deviation for the number of firms and employment is calculated by the difference in the rates of change over the period. For born, died, and moved, they are rates of occurrence within the population. For example, the average rate of moving per year (moved/firms) is almost identical between the simulation and the observed data.

Table 5-11 Summary of 20 Percent Sample Validation Results

Observed	Year	Firms	Born	Died	Moved	Employment
	2007	59,574	–	–	–	523,492
	2008	69,513	12,640	2,701	1,249	537,425
	2009	66,995	13,786	16,304	1,739	563,071
	2010	71,162	9,027	4,860	1,685	534,760
	2011	88,558	28,328	10,932	1,721	571,373
Simulation	**Year**	**Firms**	**Born**	**Died**	**Moved**	**Employment**
	2007	56,780	–	–	–	492,645
	2008	64,952	13,152	5,995	1,298	495,397
	2009	67,740	8,841	7,207	1,463	486,939
	2010	75,433	13,987	7,498	1,544	500,305
	2011	83,518	15,300	8,552	1,681	503,675
	Deviation	**-1%**	**-4%**	**-2%**	**0%**	**-6%**

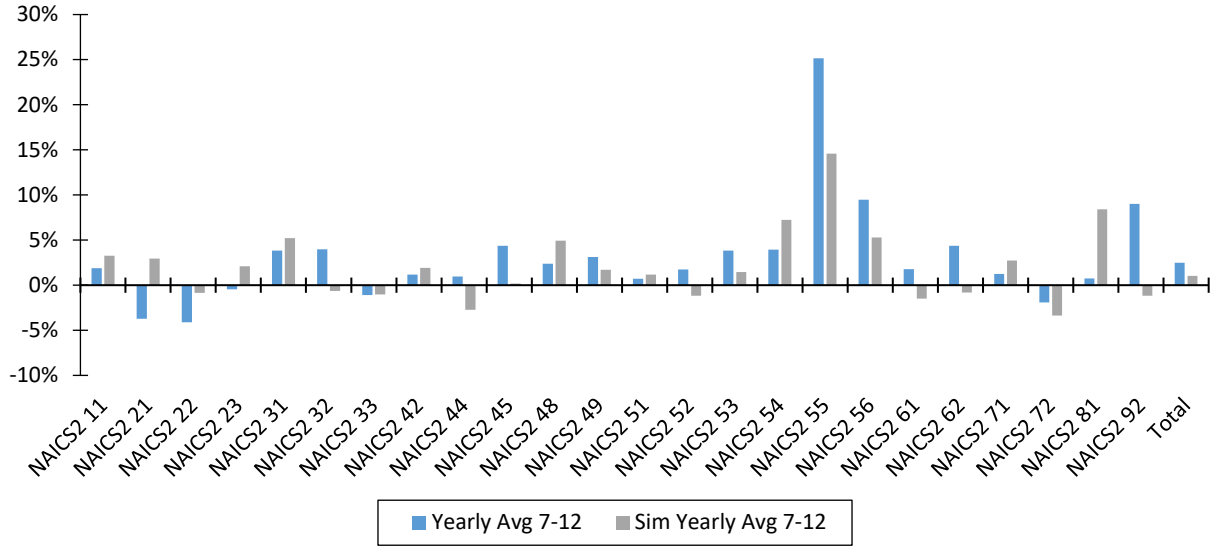
Figure 5-6 shows the change in number of firms from one year to the next as an average over the simulation period. The simulation follows the variations across industries, though its growth is less prone to outliers than the volatile observed economy is. The largest average growth is seen in NAICS 55 (management of companies and enterprises) which showed great volatility in the simulation as well as the model. Very few firms in the region were classified under NAICS 55 in the NETS data until recently.

Figure 5-6 Average of Yearly Change in Number of Firms



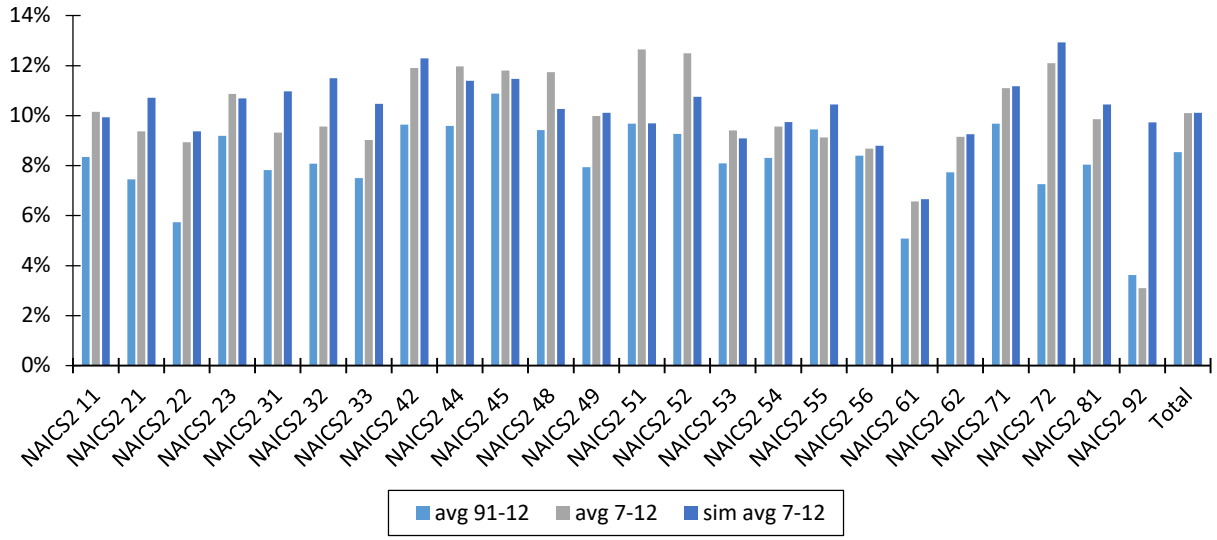
The change in employment, calculated as number of companies same as above, shows that the model is reflecting the difference between firm number growth and employment growth. The latter remained under 5 percent for all but 3 of the industries, though the former was above 5 percent for most. Again the explosive growth of NAICS 55 could not be matched, but the trends in other industries were followed, as seen below.

Figure 5-7 Average of Yearly Change in Employment



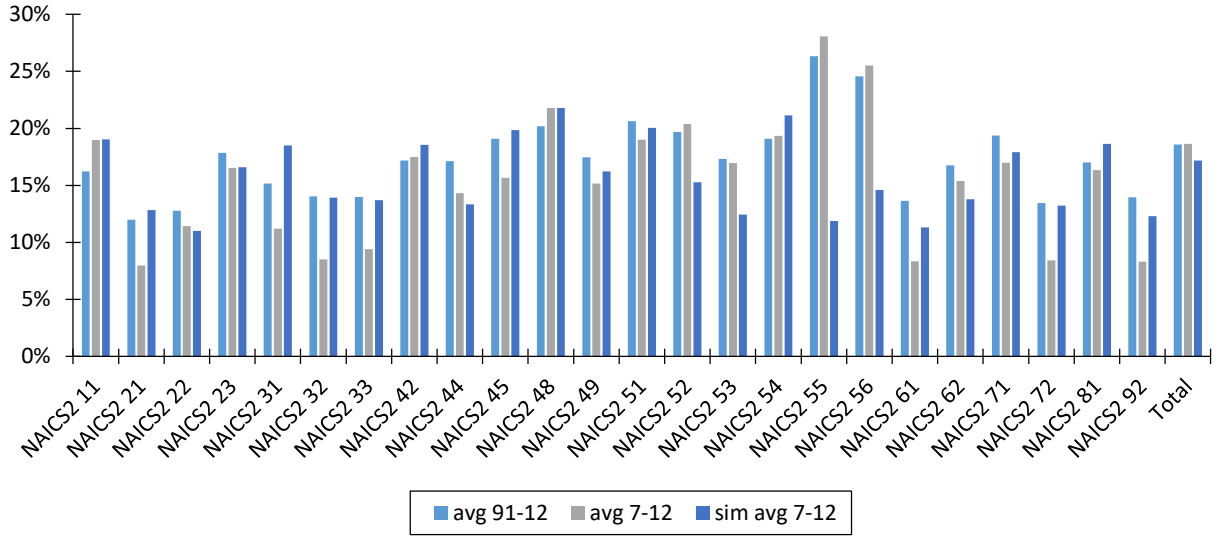
The "death rate" or the rate at which firms ceased to exist, in Figure 5-7, tracked well with averages during the period of simulation. These were generally higher than those experienced over the whole 21 year period of the NETS data. This reflects the use of estimation data from that later period.

Figure 5-8 Average of Rate of Death of Establishments (Per Year)



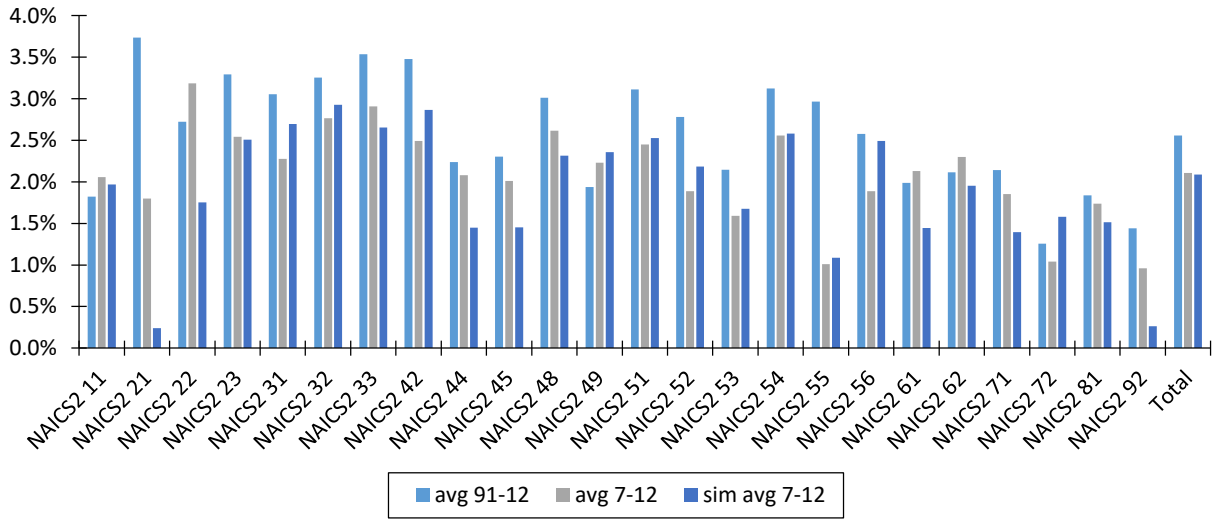
The birth rates varied between 5 percent and 30 percent with an average around 18 percent. The birth rates, unlike the other submodels, were estimated using the full 21 year period. This shows that the simulation maintains a higher number of births though the rate was unusually low due to the recent economic conditions.

Figure 5-9 Average of Rate of Birth of Establishments (Per Year)



A small percentage of firms move each year. This can cause some variances in rates in industries with few total firms, as in NAICS 21, 22, and, particularly in 1991 to 2012, NAICS 55. The regional total, as well as the majority of types, maintain an accurate number of moves.

Figure 5-10 Average of Rate of Choice to Move (Per Year)



Aggregate Level Validation Results

To further test the model's predictive accuracy we compare the simulation results across multiple datasets. We compare these results for firms with 5 or more employees only by industry types across each of the counties in the region based on the data availability. The simulation results here are compared across three different data sets: 1) 2015 InfoGroup; 2) 2012 NETS; and 3) 2012 Maricopa Employer Database.

In the following figures we summarize series of these results by comparing firm simulation results across the three databases. We compare these for two types of metrics: a) total number of firms in a county by industry type; and b) total employment in a county by industry type.

Overall, the results shown in Figure 5-11 through Figure 5-20 indicate that firm simulation results are very consistent across all the counties and by industry types. However, there are differences in a few industries e.g., NAICS 55. The differences in the regional total were low – under 2 percent for employment and 4 percent for firms. Analyzing the trends in the number of firms versus that of employment indicates both follow the same patterns of share of population and employees per firm. Notably the discrepancies between the two are very similar to those between the InfoGroup data and the NETS data, suggesting that the model is accurately following its estimation input.

Figure 5-11 Aggregate Level Validation - InfoGroup - Firm Counts

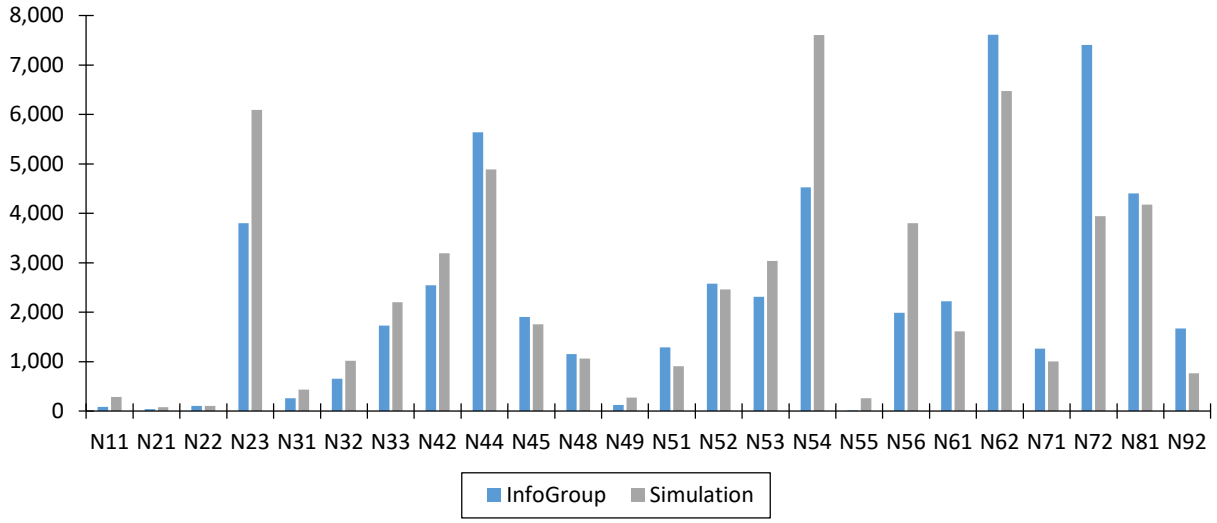


Figure 5-12 Aggregate Level Validation - InfoGroup - Employment Size

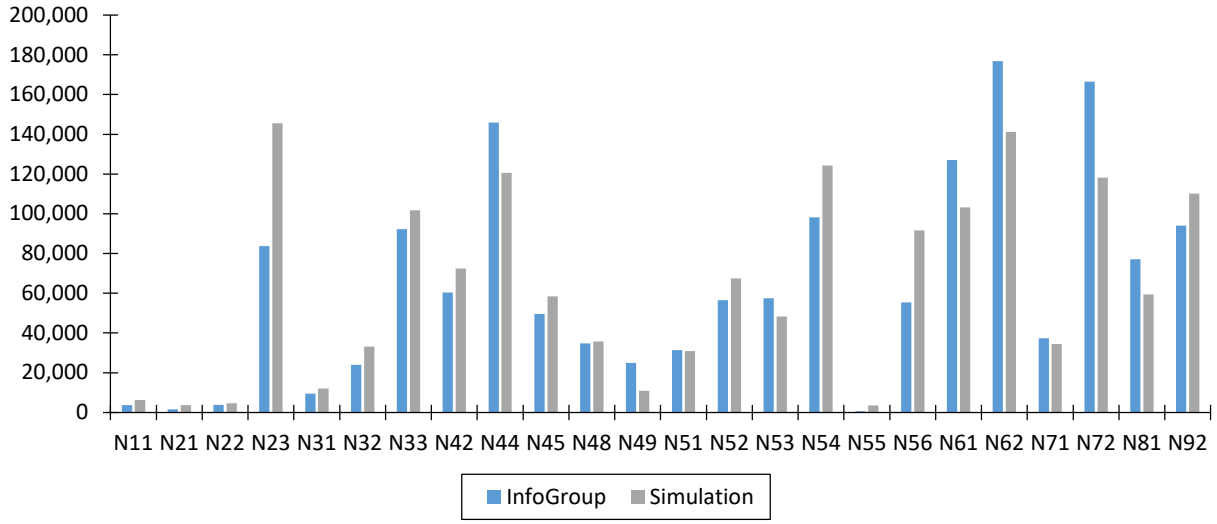


Figure 5-13 Aggregate Level Validation - Maricopa County Employer Database - Firm Counts

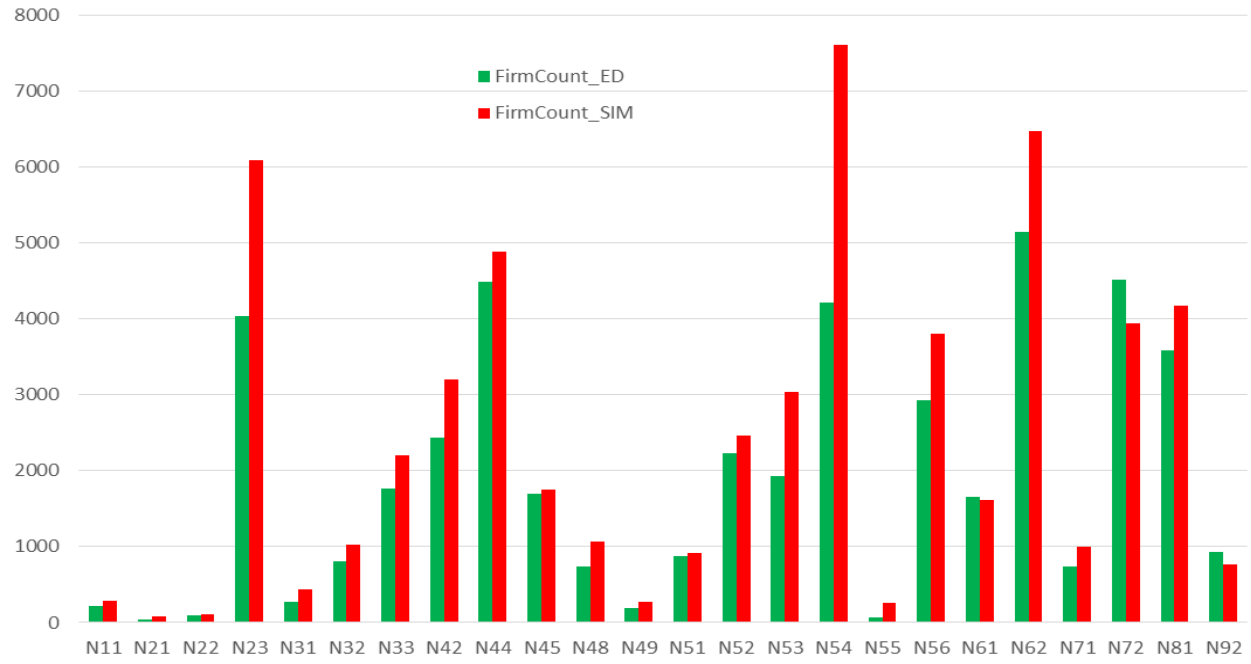


Figure 5-14 Aggregate Level Validation - Maricopa County Employer Database - Employment Size

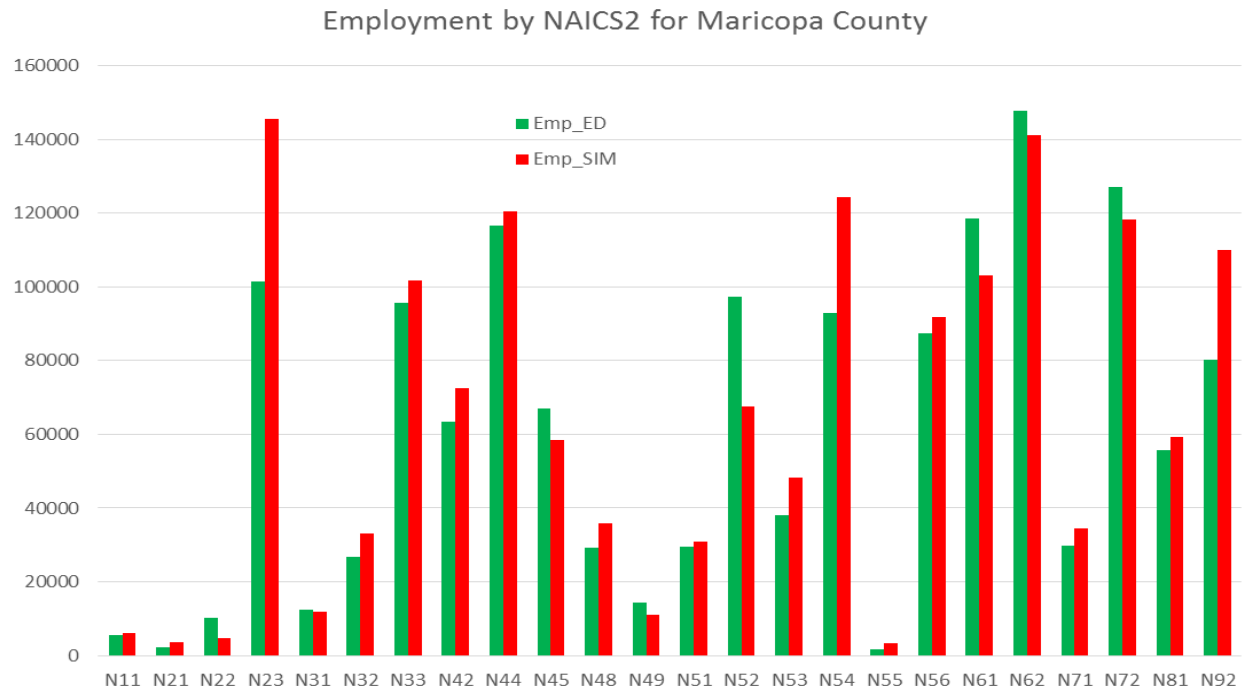


Figure 5-15 Aggregate Level Validation - NETS Database - Firm Counts (Maricopa County)

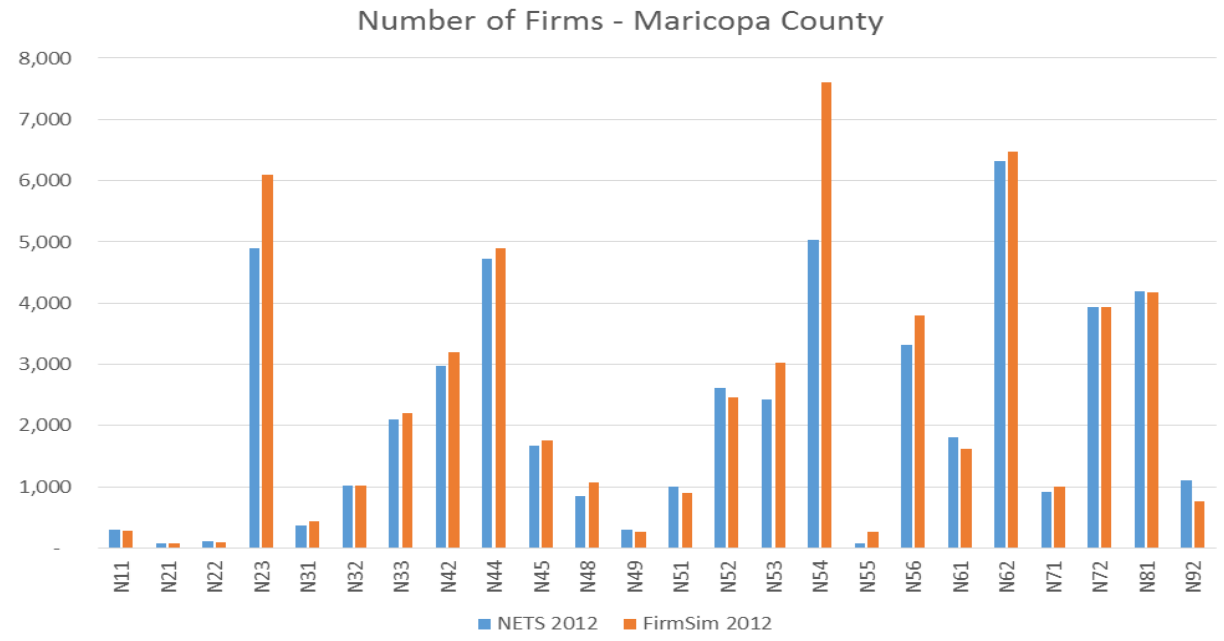


Figure 5-16 Aggregate Level Validation - NETS Database - Employment Size (Maricopa County)

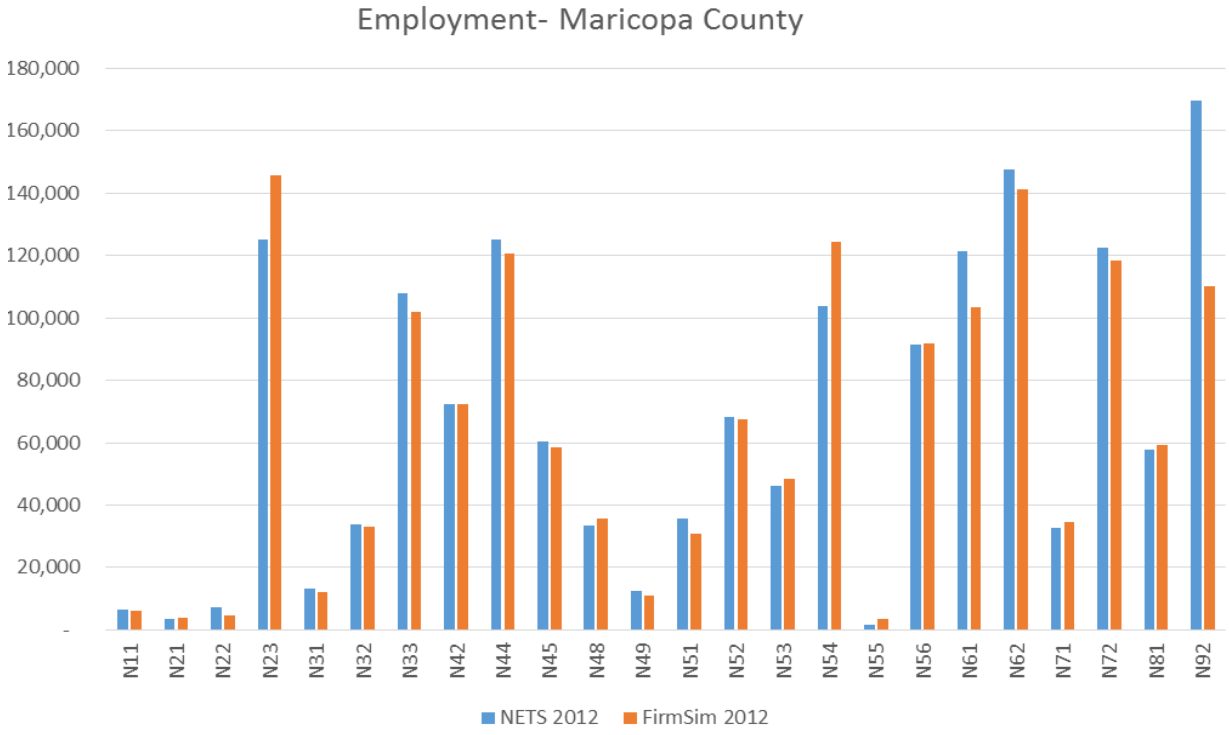


Figure 5-17 Aggregate Level Validation - NETS Database - Firm Counts (Pinal County)

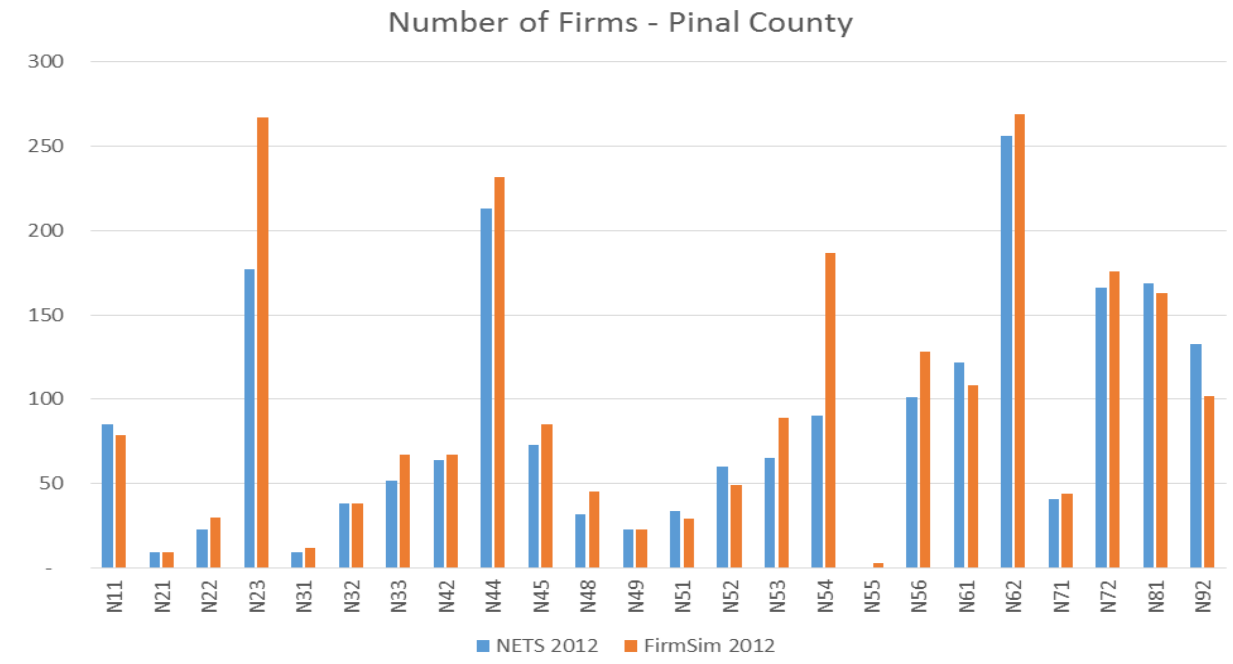


Figure 5-18 Aggregate Level Validation - NETS Database - Employment Size (Pinal County)

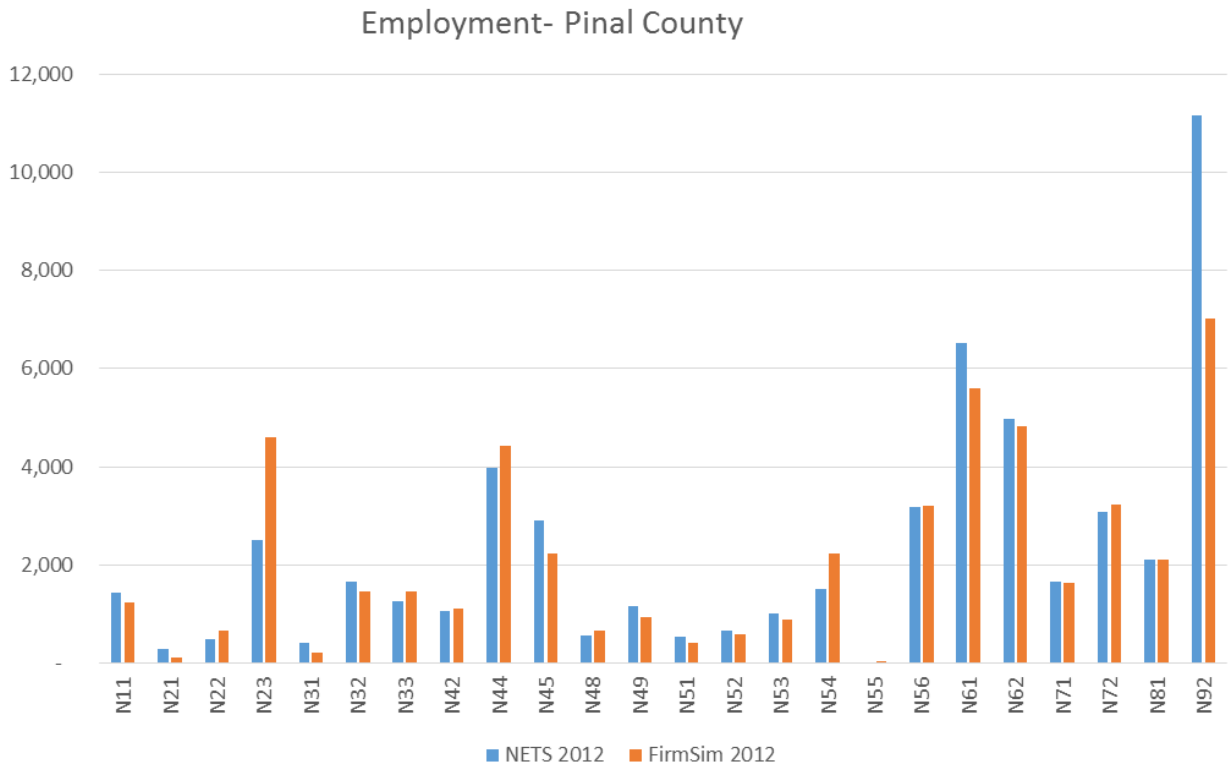


Figure 5-19 Aggregate Level Validation - NETS Database - Firm Counts (Pima County)

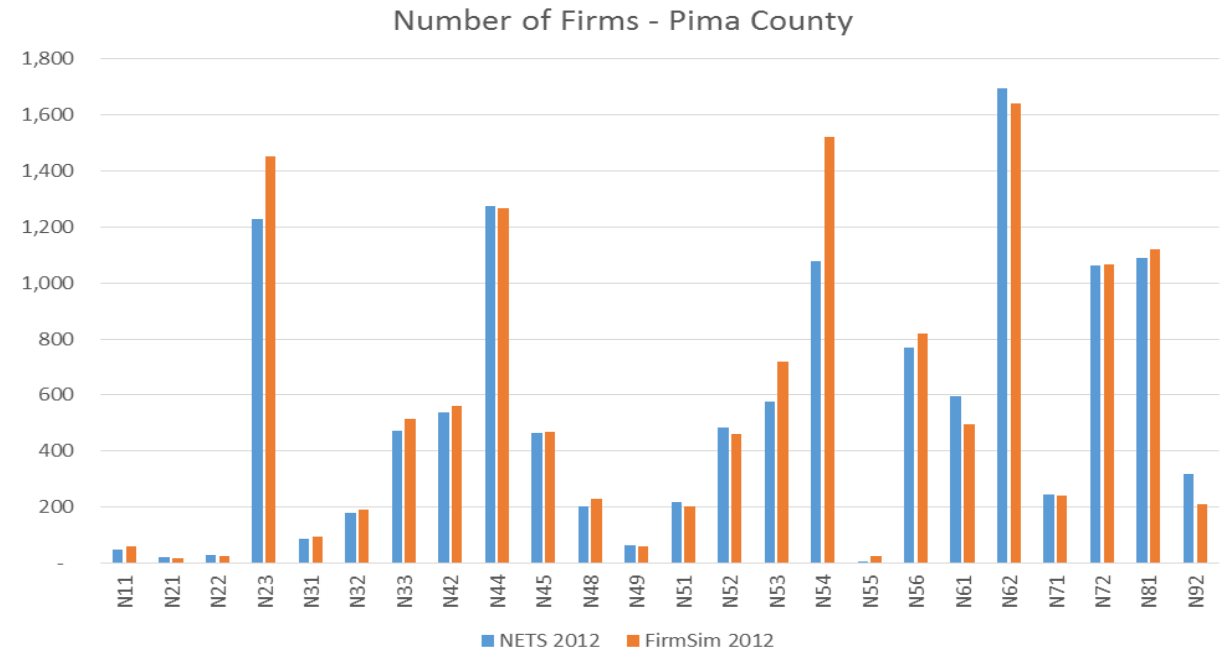
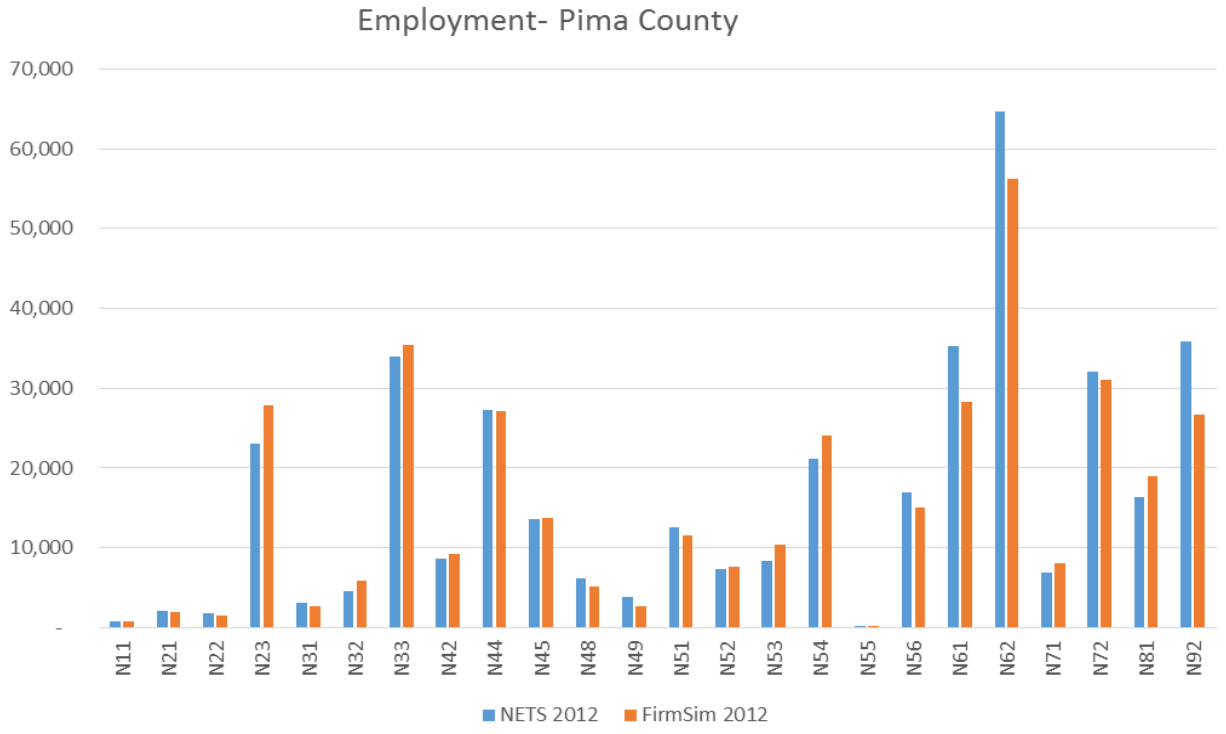


Figure 5-20 Aggregate Level Validation - NETS Database - Employment Size (Pima County)



6 Supply Chain Model

Consultant Technical Lead –Zahra Pourabdollahi (RS&H), Mark Ehlen (Mark Ehlen Consulting) and Monique Stinson (MIT)

MAG Technical Lead – Haidong Zhu, Kyunghwi Jeon, Sreevatsa Nippani, and Daehyun You

This chapter outlines the supply chain model framework and its components that was developed by RS&H team. It also describes the agents, industry classes, geographical coverage, zone system, and supply chain markets included in the model.

6.1 Model Framework

The developed agent-based supply chain and freight transport model uses disaggregate behavior-based logistics and transportation choice models to simulate commodity flows at firm level. The model considers firms or business establishments as individual decision-making agents in the freight transportation system. It assumes that logistics and supply chain decisions are made by these agents and replicates their decision-making behavior in the freight market. These logistics decisions include supplier selection, mode choice, and shipment size configuration.

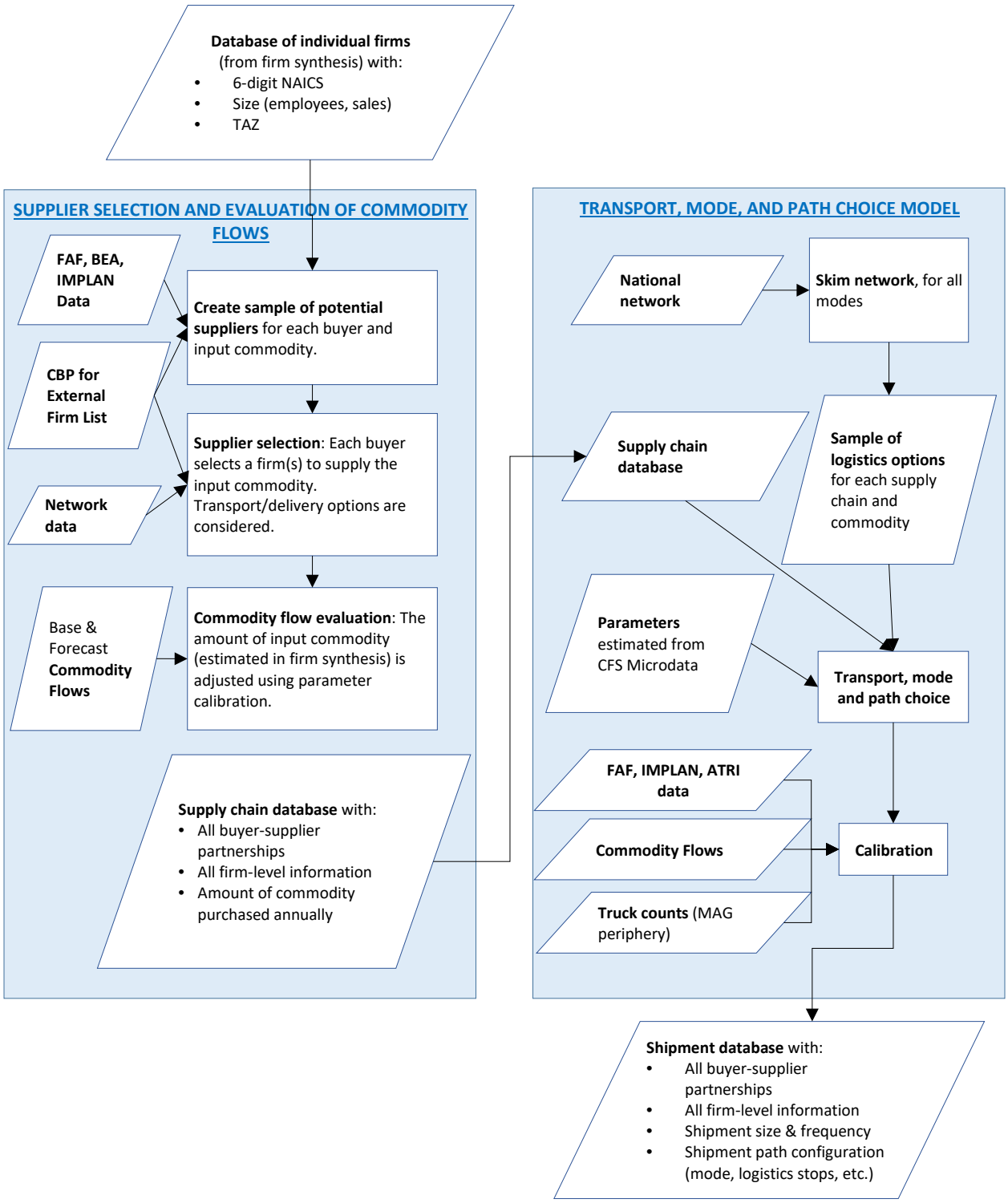
The framework comprises two main modeling components (i) supplier selection model and evaluation of commodity flows and (ii) transport, mode and path choice model. Figure 6-1 outlines the model framework and its main components. It has a sequential structure in which each model component feeds into the next module and outputs of one model is used as input in the next model component. The supplier selection and evaluation of commodity flows model comes first and determines B-to-B (business to business) supply chains. The transport, mode, and path choice model uses outputs from supplier selection model and determine other logistics choices for each supply chain in the next step.

The Supplier Selection and Evaluation of Commodity Flows consists of two main steps. The first step is the freight generation process in which total annual commodity productions and consumptions are determined at firm level. The national and regional economic activity data and socio-economic factors are used in a systematic procedure to estimate the commodity input-output rates per employee by industry class. This step generates a list of buyer and supplier firms with the type and annual amount of commodity purchased or supplied.

The supplier and buyer agents then enter the second step in which business partnerships are evaluated and B-to-B supply chains are formed between agents. The results of this model determines total annual commodity that is traded between each pair of supplier-buyer firms. The cutting-edge ACE approach was utilized for modeling the supplier selection behavior and forming supply chains in the study area. ACE approach is an evolutionary economic method that combines behavioral theories with computer science to model economic processes such as supply chain formation. In the ACE approach, economic processes are modeled as dynamic systems of interacting agents and pre-determined rules and assumptions are utilized to simulate interactions between these agents. The model in this project was developed based on ACE approach to capture both behavioral and economical aspects of supplier selection process for different markets. In addition, multiple criteria, such as cost, distance and reliability, were incorporated into the model structure which represents decision makers' preferences in selecting

business partners. The methodology and data for development of the Supplier Selection and Evaluation of Commodity Flows are discussed in detail in Section 6.2.

Figure 6-1 MAG Behavior-Based Supply Chain and Freight Transportation Framework



Once the supply chains are formed in the Supplier Selection and Evaluation of Commodity Flows model, the total annual commodity flows between firms are used as input to the Transport, Mode, and Path Choice model in which logistics choices of mode and shipment size are analyzed for the commodity flows to, from, and within the Arizona Sun Corridor Megaregion. The Nested Logit structure is utilized to jointly model mode and shipment size choice. Five modes of transportation, including truck, rail, air, pipeline, and parcel (such as, U.S. Postal Service, UPS, and other couriers) are included in the model. The continuous weight variable is classified into four categories of shipment sizes and included in the model structure. The model is highly disaggregated and developed using the first generation of Commodity Flow Survey Public Use Microdata (CFS PUM). The data is an invaluable source that provides detailed shipment information for approximately 4.5 million individual shipments obtained from business establishments. The model converts commodity flows into shipments between supplier – buyer. The output of the model is records of shipments with their main characteristics including supplier, buyer, commodity type, total annual flow, mode choice, shipment size, rail yard used for trans-shipment (for rail shipments), and external highway station (for external truck shipments). The methodology and data for development of the Transport, Mode, and Path Choice model are discussed in detail in Section 6.3.

6.1.1 Geographical Coverage and Zone System

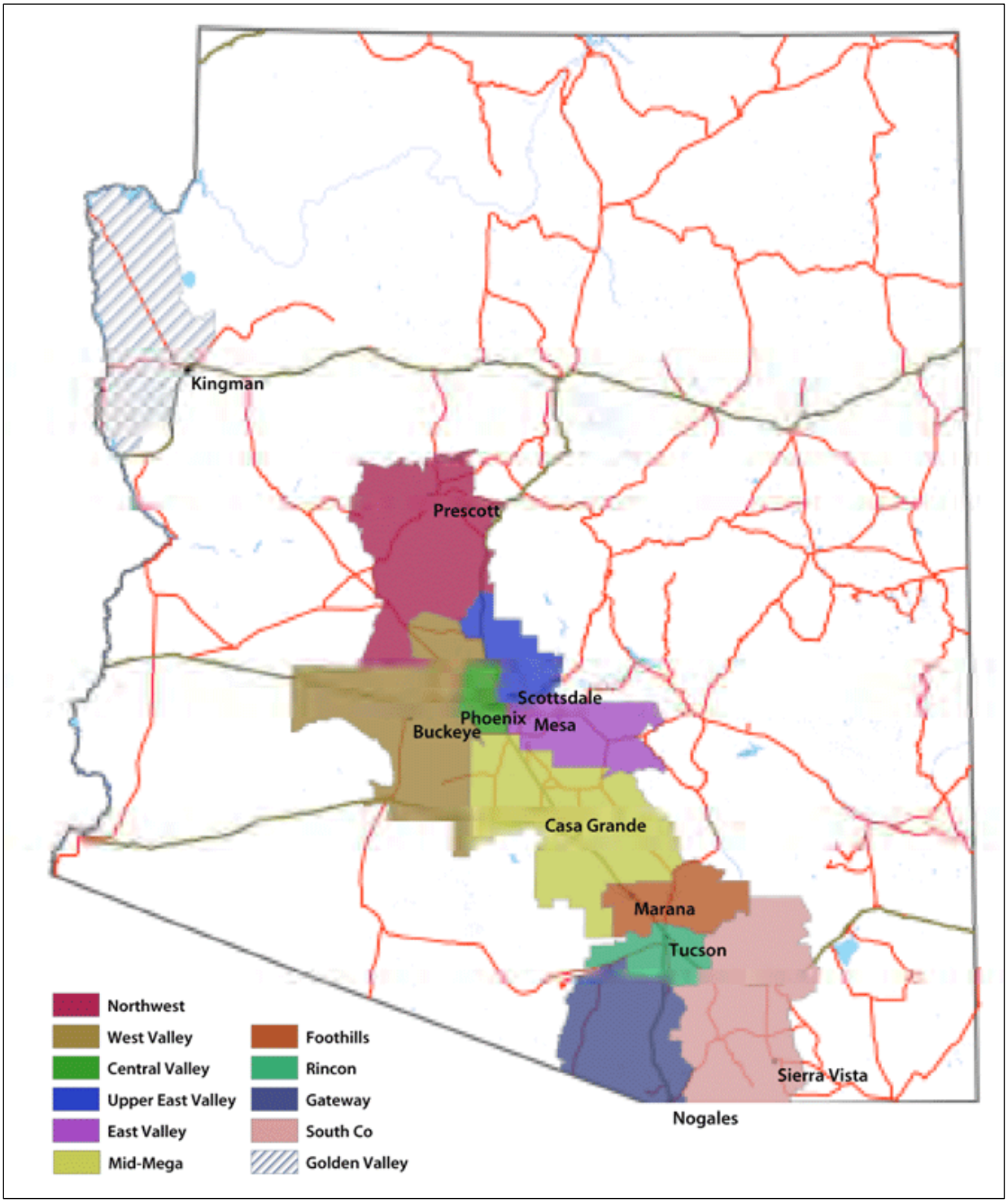
The model simulates freight shipments to, from, and within the Arizona Sun Corridor Megaregion. Therefore, it is considered a regional model which covers external commodity flows to and from the region at disaggregate level of firm-to-firm. The Sun Corridor Megaregion is equivalent to Indiana in both size and population. It is among the fastest growing megaregions in the country and a freight gateway to the international market. The Sun Corridor covers the Phoenix and Tucson Metropolitan Areas which are located in Maricopa, Pima and Pinal Counties and home to more than 80 percent of the state's residents. Figure 6-2 demonstrates the Sun Corridor Megaregion.

The model focuses on regional commodity flows to/from/within MAG/PAG region. However, in order to replicate the real world supply chains that originate from or end in external regions, it was necessary to include all agents (firms) in external regions for the development and application of the model. Therefore, a variable zone system was used for the framework which comprises finer zones in the Sun Corridor region for more detailed analysis and larger zones for the external areas to Sun Corridor where loss of precision is acceptable. The zone system is comprised of Traffic Analysis Zone (TAZ) level in the MAG/PAG region, counties in the rest of Arizona and FAF zones for the rest of the country. Table 6-1 shows zone numbers and range used in the model. Figure 6-3 presents the zone system used in this project.

Table 6-1 Zone System and Zone Numbers in the Model

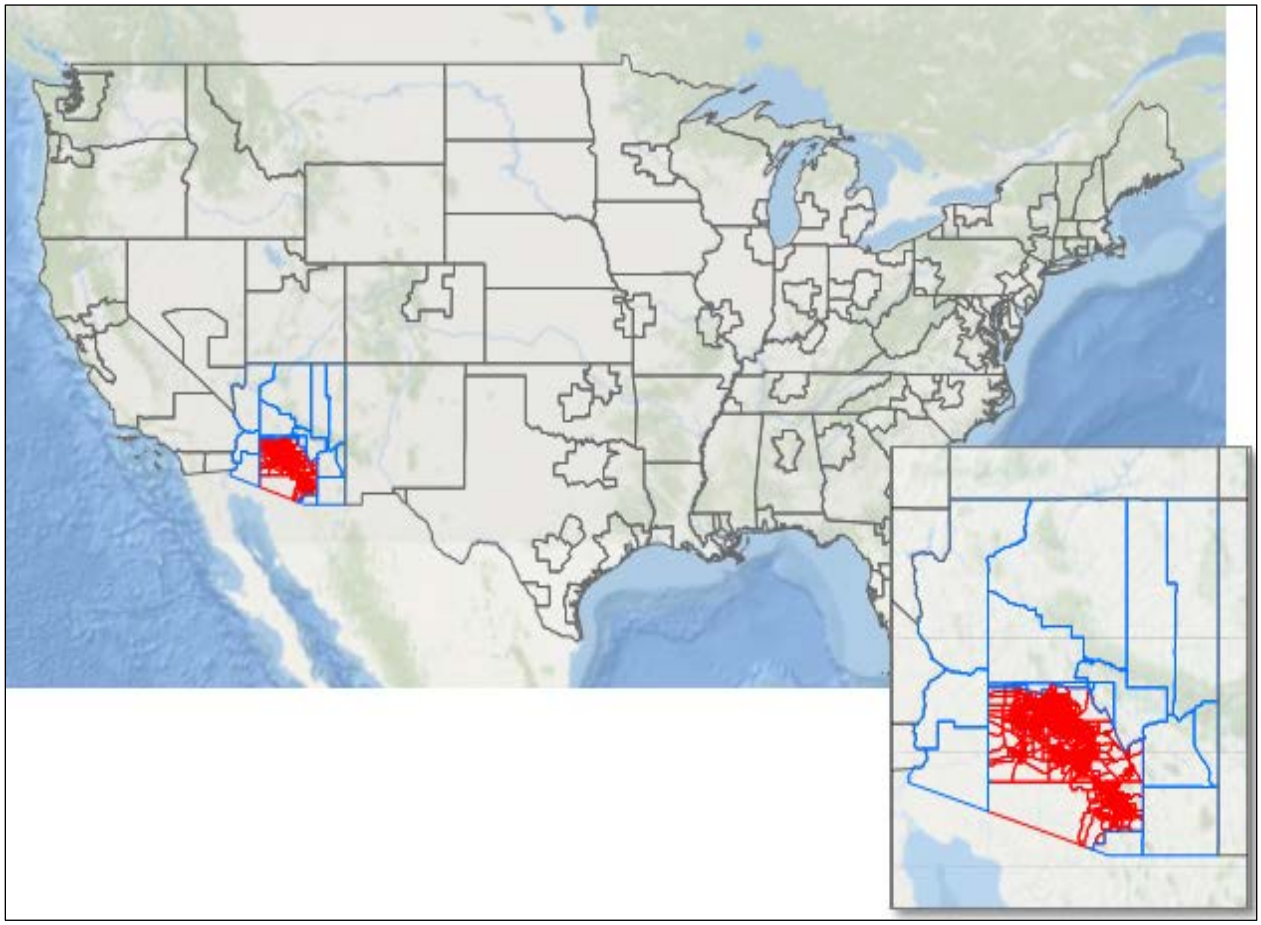
Model Zone Number	Corresponding FAF-Zone	Region
101-3140	41	TAZs in MAG modeling area
5001-6104	42	TAZs in PAG modeling area
9000-9012	49	Rest of AZ (Counties)
9100-9220	Other FAF zones (11,12,…)	Rest of country (FAF zones)

Figure 6-2 Sun Corridor Megaregion[27]



[27] Megapolitan: Arizona's Sun Corridor, Morrison Institute for Public Policy, Arizona State University: https://morrisoninstitute.asu.edu/products/megapolitan-arizonas-sun-corridor

Figure 6-3 Model Zone System



The model replicates domestic commodity flows for the region. This includes all the domestic flows that start and end in the country and the domestic leg of all international flows that has an origin or end in the MAG/PAG region. The exclusion of the international leg of the import/export flows is due to the lack of data on foreign firms and supplier selection processes of the international trade market. Data on import/export flows to/from the MAG/PAG region and more detailed information of international trade market behavior is required to correctly model the international flows in the behavior-based model framework.

The framework has a very flexible structure regarding the geographical scale. It can perform analysis at regional scale or can be tailored to be used at the national or sub-area scale. While using the framework at nationwide scale require some aggregation in order to deal with computational issues, performing analysis at sub-area scale might require more detailed data.

6.1.2 Decision Making Agents

Firms or business establishments are the decision-making agents in the model. It is assumed that logistics choices and supply chain decisions are made by these agents. The underlying assumption of the ACE based approach is that economic processes such as supply chain formation are result of interactions between decision-making agents based on behavioral rules and assumption. Before replicating supply

chains and freight transportation choices, decision making agents need to be enumerated or synthesized. In this model, decision making agents are all business establishments in the country that can interact with any firm in the MAG/PAG region.

For the MAG/PAG region, information on the business establishments was obtained from Firm Synthesizer Model. For the external regions, the County Business Patterns (CBP) data provided by U.S. Census Bureau was analyzed to obtain economic information on firms. In total, 8,500,754 business establishments were considered as agents in the model. This translates into 117,989,784 paid employees. Since considering all these individual agents in the simulation process was computationally cumbersome, a straightforward method was used to partition these individual firms into firm clusters and reduce the number of decision-making agents. In this process, all firms in the same zone, with the same industry class, and same employment size are grouped into a cluster of decision makers who are assumed to have similar behavior regarding supply chain decision making process. Table 6-2 summarizes MAG/PAG establishment and employment data for all industry classes considered in this model for the base year of 2015. The 2-digit North American Industry Classification (NAICS) system is used for determining the industry class.

It should be noted that not all these agents are freight producer or consumer. The supplier and buyer agents were extracted from this list of firms in the supplier selection model. The process to identify supplier and buyer agents is discussed in Section 6.2.

Table 6-2 MAG/PAG Total Number of Business Establishments Enumerated in the Model

NAICS	DESCRIPTION	EMPLOYMENT	ESTABLISHMENTS
11	Agriculture, Forestry, Fishing and Hunting	15,506	5,052
21	Mining, Quarrying, and Oil and Gas Extraction	5,533	432
22	Utilities	8,141	557
23	Construction	193,354	48,038
31-33	Manufacturing	187,468	16,053
42	Wholesale Trade	111,312	18,452
44-45	Retail Trade	288,057	46,665
48-49	Transportation and Warehousing	66,407	9,190
51	Information	54,148	9,899
52	Finance and Insurance	112,249	23,975
53	Real Estate and Rental and Leasing	96,916	25,476
54	Professional, Scientific, and Technical Services	199,234	50,644
55	Management of Companies and Enterprises	9,723	4,214
56	Administrative and Support and Waste Management and Remediation Services	248,484	88,629
61	Educational Services	191,967	17,821
62	Health Care and Social Assistance	223,814	28,275
71	Arts, Entertainment, and Recreation	48,174	8,284
72	Accommodation and Food Services	142,008	8,935
81	Other Services (except Public Administration)	134,471	37,733
92	Public Administration	106,603	1,738
TOTAL		2,347,864	450,062

6.1.3 Supply Chain Markets

Forty two classes of commodities based on the Standard Classification of Transported Goods (SCTG) were considered in this model (see Table 6-3). Each class of commodity is considered as an independent economic market for which supply chains were determined.

Table 6-3 SCTG Classes Considered in the Model

SCTG	COMMODITY DESCRIPTION
1	Live animals and live fish
2	Cereal grains
3	Other agricultural products
4	Animal feed and products of animal origin, n.e.c.
5	Meat, fish, seafood, and their preparations
6	Milled grain products and preparations, bakery products
7	Other prepared foodstuffs and fats and oils
8	Alcoholic beverages
9	Tobacco products
10	Monumental or building stone
11	Natural sands
12	Gravel and crushed stone
13	Nonmetallic minerals n.e.c.
14	Metallic ores and concentrates
15	Coal
16	Crude Petroleum
17	Gasoline and aviation turbine fuel
18	Fuel oils
19	Coal and petroleum products, n.e.c.
20	Basic chemicals
21	Pharmaceutical products
22	Fertilizers
23	Chemical products and preparations, n.e.c.
24	Plastics and rubber
25	Logs and other wood in the rough
26	Wood products
27	Pulp, newsprint, paper, and paperboard
28	Paper or paperboard articles
29	Printed products
30	Textiles, leather, and articles of textiles or leather
31	Nonmetallic mineral products
32	Base metal in primary or semi-finished forms and in finished basic shapes
33	Articles of base metal
34	Machinery
35	Electronic and other electrical equipment and components and office equipment
36	Motorized and other vehicles (including parts)
37	Transportation equipment, n.e.c.
38	Precision instruments and apparatus
39	Furniture, mattresses and mattress supports, lamps, lighting fittings, and illuminated signs
40	Miscellaneous manufactured products
41	Waste and Scrap
43	Mixed Freight

6.2 Supplier Selection and Evaluation of Commodity Flows

This section describes the Supplier Selection Model (SSM) that has been developed for the MAG/PAG SHRP2-C20 agent-based supply chain model. The MAG project requires a supplier selection model that allocates commodities from those firms that produce them to those that need them. Firms include those

within the MAG/PAG megaregion and outside the region. The approach described herein is an agent-based computation economics method that captures how buyers will make discrete choices about who to purchase from, and suppliers make indirect decisions about to whom to sell.

This section describes:

- The SSM components and how they interact with each other in the model stream;
- The input data, model development methodology and some of the preliminary results.

In this section, the focus has been on developing the algorithms, inputs, and coding processes. The model has been applied domestically for a single commodity for development and testing purposes. The inputs and coding have been expanded to cover all model commodities and domestic markets in calibration and application step.

6.2.1 Data Sources

The Supplier Selection Model has been developed using proprietary and publicly available data, supplier selection criteria from the literature, and multimodal skims provided by MAG project team. Data sources used in development include:

- Firm Synthesizer output was used to classify supplier and buyer agents in the Megaregion area;
- Employment data from County Business Patterns (CBP) was used to generate firms external to the Sun Corridor Megaregion and to estimate industry-specific input and output factors (see below);
- IMPLAN regional make-use tables were used to calculate regional production and consumption rates by industry class per employee for all firms in the Megaregion
- Bureau of Economic Analysis (BEA) Input-Output (IO) Accounts tables, which contain Make and Use amounts of exchanged commodities between industries (at the six-digit NAICS level) were used to calculate amount of commodity produced and consumed on average per employee in each industry class for firms outside the Mega region;
- National average and region – specific commodity prices derived from the FAF4.1 and Arizona-based sample of CFS-Microdata;
- Two Commodity-Industry Crosswalks that provide mapping between industry classes (6-digit NAICS and BEA IO industry classes) and made commodities (2-digit Standard Classification of Transported Goods or SCTG). These crosswalks were developed by the project team using the detailed description of SCTG and NAICS codes provided by U.S. Census Bureau and a preliminary commodity-industry bridge used in FAF data;
- BEA correspondence table that links BEA IO industry classes to NAICS 6-digit industry classes;

- Supplier selection criteria and importance weights derived from the results of a freight establishment survey and a system of ordered logit models developed in a recent study [28];
- Aggregated inbound and outbound commodity flows at FAF zone level derived from FAF4.1 data and used to calibrate estimated production and consumption amounts at firm-level.
- Great Circle Distance (GCD) matrix calculated using the zoning system in the Geographical Information System (GIS) software.

Calibration of the Supplier Selection model will follow in Section 6.4. The calibration stage has utilized both aggregate and disaggregate data sources, including FAF and CFS Microdata which will be discussed later.

6.2.2 Foundational Principles of the Model

The following guiding principles were used in developing the model and in selecting the appropriate algorithms for the SSM. These are:

- The flows of commodities from supplier to buyer need to reasonably represent the real-world flows, in particular, each particular class of buying firm has potential classes of supplier firm, reflecting a "best" supplier for each buyer;
- A buying firm needs to select a single supplier from which to purchase their commodity. As a result, each buying firm can only buy from firms that have supply greater than or equal to the buying firm's demand; [29]
- Buying firms have a number of quantitative and potentially qualitative properties that influence their choice of supplier;
- Some supplier firms will have more demand for their commodity than supply, and will therefore have to turn down some potential buyers; and
- To the extent that real supplier markets are Pareto efficient (each buyer can't be made better off without making other buyers worse off), the SSM allocations between suppliers and buyers are "optimal."

6.2.3 Market Clearing Mechanisms

In modern supply chain management, multiple factors are taken into account in supplier evaluation and selection process. Based on findings from several studies [30, 31], price/cost is not the most widely used

[28] Pourabdollahi, Z., Karimi, B. and K. Mohammadian, A Multi-Criteria Supplier Evaluation Model in a Supply Chain. Submitted to the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
[29] This assumption is required as a primary condition for the selected matching algorithm. This would also decrease the computational complexity of the problem by reducing the size of matching sets. More detailed commodity-industry and industry-industry crosswalks need to be developed and the market clearing algorithm need to be modified to relax this condition and allow buyers to outsource from multiple suppliers (for purchasing a single commodity).
[30] De Boer, L., Labro, E. and P. Morlacchi, A review of methods supporting supplier selection, European Journal of Purchasing & Supply Management, Vol. 7, 2001, pp. 75–89.
[31] Ho, W., Xu, X. and P. K. Dey, Multi-criteria decision making approaches for supplier evaluation and selection: A literature review, European Journal of Operational Research, Vol. 202, 2010, pp. 16–24.

criterion in supplier selection models. Instead, it takes the third place after quality and delivery which are the most frequently used factors in selecting suppliers. Solving supplier selection problem requires the application of a market-clearing mechanism. The MAG SSM uses a modeling process that is founded on economic, market-clearing mechanism.

In economics, market-clearing mechanisms are defined as processes of allocating a good or service from a set of potential sellers to a set of potential buyers in a market. Various analytical and heuristic approaches incorporating multiple criteria have been used for modeling supplier selection process. Multi-objective optimization (MOP) [32, 33], analytic hierarchy process (AHP) [34, 35], data envelopment analysis (DEA) [36, 37, 38] and simple multi-attribute rating technique (SMART) [39, 40] are among the most common multi-criteria methods used to evaluate and select the best suppliers. Most of these approaches are based on traditional theories in which the behavioral elements of supplier selection process are overlooked, and preferences and beliefs of decision-making agents are not captured.

Based on the reviewed literature, applied market-clearing mechanisms to supplier selection problem can be classified into at least four different approaches: econometrics, optimization, auctions, and hybrid. This section describes each method and its potential application to the MAG SSM.

6.2.3.1 Econometrics Approaches

Econometrics approaches use data to estimate the likely flow of commodities between suppliers and buyers. The Industry Make & Use Tables provided by BEA, FAF Commodity Flow Data, and regional flows estimated by regional industry input-output accounts, computable general equilibrium, and hybrid models as are examples of estimated supplier-buyer purchases and shipments, for both goods and services that have been developed by compiling survey or census data and estimating "true" flows through regression analysis.

However, these models would be difficult to implement in the MAG model: First, each buyer has to ultimately select one supplier – and in subsequent implementations to select the same supplier. Econometrics models such as Probit and Logit models are a form of "probability of selection" model, not actual deterministic selection of one supplier. Second, the set of buyers' selections must also be constrained by (i) total demand must be less than or equal to total supply, and (ii) each supplier cannot

[32] Haleh, H. and A.Hamidi, A fuzzy MCDM model for allocating orders to suppliers in a supply chain under uncertainty over a multi-period time horizon. Expert Systems with Applications, Vol. 38 (8), 2011, pp. 9076–9083.
[33] Yu, M., Goh, M. and H. Lin, Fuzzy multi-objective vendor selection under lean procurement. European Journal of Operational Research, Vol. 219(2), 2012, pp. 305–311.
[34] Levary, R. R. Using the analytic hierarchy process to rank foreign suppliers based on supply risks. Computers and Industrial Engineering, Vol. 55(2), 2008, pp.535–542.
[35] Mafakheri, F., Breton, M., and A. Ghoniem, Supplier selection-order allocation: A two-stage multiple criteria dynamic programming approach. International Journal of Production Economics, Vol. 132(1), 2011, pp. 52–57.
[36] Wu, T., and J. Blackhurst, Supplier evaluation and selection: An augmented DEA approach. International Journal of Production Research, Vol. 47(16), 2009, pp.4593–4608.
[37] Falagario, M., Sciancalepore, F., Costantino, N. and R. Pietroforte, Using a DEAcross efficiency approach in public procurement tenders. European Journal of Operational Research, Vol. 218(2), 2012, pp. 523–529.
[38] Saen, R. F., Developing a new data envelopment analysis methodology for supplier selection in the presence of both undesirable outputs and imprecise data. International Journal of Advanced Manufacturing Technology, Vol. 51(9–12), 2010, pp. 1243–1250.
[39] Huang, S.H. and H. Keska, Comprehensive and configurable metrics for supplier selection. International Journal of Production Economics, Vol. 105 (2), 2007, pp. 510–523.
[40] Chou, S., and Y. Chang, A decision support system for supplier selection based on a strategy-aligned fuzzy SMART approach. Expert Systems with Applications, Vol. 34(4), 2008, pp. 2241–2253.

sell more than they make. Constrained econometric estimation, where theoretically possible, is computationally complex and makes software verification difficult.

6.2.3.2 Operations Research Approaches

Operations research approaches are constrained optimizations, where an objective function is maximized or minimized subject to a number of conditions that constrain the function's potential solution space. In the MAG SSM, an operations research approach could be, for example, "minimize the difference between the buyers' and suppliers' preferred transaction partners, subject to each supplier not selling more than they produce." But the optimization formulation would be complex: the optimization would be computing over discrete-choice space, where each buyer has a ranked set of scores for potential suppliers and each supplier has many ranked sets of scores for each combination of potential buyers.

6.2.3.3 Auction Approaches

The third potential approach is using auction algorithms. Auctions are a common Agent-based Computational Economics (ACE) approach for allocating supply and demand over many buying and supplying firms, for example in financial-instrument and electricity spot markets. In auctions, buyers and supplier "bid" in their desired amount bought and sold, and a market-clearing mechanism allocates the market commodity from suppliers to buyers, based on the particular market rules. Some auction markets use an algorithm to "clear" the market, while others have specific market makers who themselves buy and sell the commodity, profiting from the difference between sales price and purchase price. Different auction methods are effective, in particular, when buyers and suppliers are not necessarily willing to reveal their true preferences of the value of the purchase or sale.

6.2.3.4 Hybrid Approaches

Finally, the hybrid approaches combine two or more of other approaches to leverage advantages of each individual approach and compensate for their shortcomings. The hybrid optimization and econometrics ACE approach is a recent example of application of hybrid methods for supplier selection problem [41].

Each of these approaches have their advantages and shortcomings for application to this supplier selection problem. For example, an operation research or hybrid model formulation would be very complex for application for such a big problem, and the econometrics models such as Logit and Probit models limit the application capabilities due to implementation of "probability of selection", not actual deterministic selection of one supplier. Various modeling approaches were evaluated to select the most appropriate approach for application to supplier selection problem in this project. Next section describes the selected method and the cutting-edge market-clearing mechanism that was designed and applied.

6.2.4 Modeling Approach

Each of the reviewed approaches has its advantages and disadvantages, but most become increasingly complex as the number of buying/selling firms become larger. Given that the primary goals of this study are to ensure market-based economic efficiencies, an algorithm that is based on ACE approach and results in optimal and stable market allocations is selected. Agent-based computational economics or "ACE is the computational study of economies modeled as evolving systems of autonomous interacting

[41] Pourabdollahi, Z., Karimi, B., Mohammadian, A., and Kawamura, K. A Hybrid Agent-based Computational Economics and Optimization Approach for Supplier Selection Problem, the 95th Annual Transportation Research Board, January 2016, Washington D.C.

agents" [42]. The initial conditions and interaction rules are specified by the modeler, however, the model is designed to evolve over time as its constituent agents interact with each other and learn from previous interactions.

Economic markets comprise large numbers of interacting agents that are involved in the decision making processes. These micro-level interactions result in establishment of macroeconomic regulations and conditions such as trading protocols and monetary policies which in turn feed back into the determination of micro-level interactions. This creates a complex dynamic system with recurrent feedback loops connecting agent behaviors, interaction networks, and regional/social welfare outcomes [43]. Traditional supplier selection approaches lack the means to model this two-way feedback between micro-level and macro-level interactions quantitatively and accurately. These models heavily rely on macro-level externally imposed conditions such as fixed decision rules, common knowledge assumptions, and market equilibrium constraints. However, the ACE approach captures face-to-face interactions among economic agents and models complex economic decision making processes at micro-level.

The modeling framework in this study is a regional-market model of buyers and suppliers that uses an ACE market-clearing algorithm. The process is based on the assumption that buyers and suppliers in a market access each other randomly, that is, without particular order, on an as-needed basis where buyers make orders when they get 'low' on on-site inventories.

The supplier-selection algorithm, chosen for implementation, is the Noble Prize-winning Roth-Peranson [44] hospital/residents algorithm (RPA). The RPA is a form of auction, and has been proven to be an optimal and stable market-clearing mechanism. The algorithm most prominent use is in the National Residents Matching Program which sets the basis for receiving the Nobel Prize in Economics. As described in detail below, the RPA method uses a set of constructs and parameter assumptions that are very close to those required by the MAG supplier selection model. In the original RPA hospital/residents algorithm, each resident ranks each hospital, and each hospital ranks each resident. The market-clearing mechanism then assigns each resident to each hospital in a way that no other allocation is better. In this supplier selection model, residents are represented by buyers and hospitals are represented by suppliers. The next section describes the algorithm structure.

The RPA is ideal for the MAG SSM, largely because it only requires ordinal rankings of suppliers and buyers, not quantities and conditions to be met. In other words, the model only needs an ordinal ranking of who each buyer would like to purchase from, and an ordinal ranking of whom each supplier would like to sell to. In the MAG implementation, the ranking is formed using a quantitative Supplier Selection Score for each supplier and buyer, which is comprised of different criteria including prices, distance, capacity and other factors. Using this mechanism, we can ensure that (i) each buyer's set of potential suppliers are only those with the right requirements and (ii) each potential supplier generates a supply that can meet or exceed the buyer's demand.

[42] Tesfatsion, L., Agent-Based Computational Economics. ISU Economics, 2003. Accessed in June 2016 at http://www2.econ.iastate.edu/tesfatsi/
[43] Tesfatsion, L., Agent-Based Computational Economics: A Constructive Approach To Economic Theory. Economics Department, Iowa State University, 2005. Accessed in June 2016 at http://www2.econ.iastate.edu/tesfatsi/
[44] Roth, A. and E. Peranson (1999). "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design". The American Economic Review 89 (4): pp. 756–757.

6.2.5 Model Design

The model only requires an ordinal ranking of suppliers, and an ordinal ranking of buyers. The ranking is formed using a quantitative Supplier-Buyer Selection Score (Utility) for each supplier and buyer, which is calculated based on important evaluating criteria including the different supplier prices, distance, reliability and delivery. The supplier-buyer score system and ranking procedure is adopted from literature [45]. Table 6-4 presents the ranking parameters and their relative importance weights that were considered in this model for ranking suppliers. The annual purchase amount (demand) is the only parameter in evaluating buyers and ranking them from suppliers' perspective. It is assumed that suppliers prefer to sell to buyers with larger demand.

Table 6-4 Average Importance Weights of Supplier Evaluation Criteria[43]

Criterion	Description	Proxy	Data Source	Average weight
Cost (Price)	Region specific price of traded commodity	-	FAF	4.13
Reliability	Reliability of supplier in providing the commodity	Annual Production/Supply Capacity	Synthesized Firms & County Business Patterns	4.34
Delivery	Accessibility to different modes of transportation	Travel time of different modes of transportation	Network Skims	4.36
Distance	Distance between supplier and buyer	Great Circle Distance	GIS shapefile of Zone system	3.15

The algorithm is based on three main steps:

1. Each buyer calculates a score (utility) for each potential supplier, based on important supplier characteristics such as cost, distance, supplier size, available modes of transportation. This buyer then ordinally ranks each supplier from best to worst score. Each potential supplier in the set must meet the basic requirements of the buyer, including having supply of the commodity greater than or equal to the buyer's demand for the commodity. The evaluating criteria and their average importance weights are presented in Table 6-4.

2. Each supplier calculates a score for each potential buyer, based on buyer's total annual demand. While this buyer ranking by suppliers is not as critical as the supplier ranking by buyers, a ranking is needed for the algorithm. This supplier then ordinally ranks each buyer from best to worst.

3. The market-clearing algorithm iteratively matches buyers to suppliers, ensuring that all economic conditions are met and the market-clearing mechanism has allocative efficiency and stability.

It should be noted that each commodity represents an individual market which is allocated independently. Therefore, the market clearing algorithm is based on the commodity rather than industries. Moreover, the weights of supplier evaluation criteria (Table 6-4) do not alter from one market to another one.

[45] Pourabdollahi, Z., Karimi, B. and K. Mohammadian, A Multi-Criteria Supplier Evaluation Model in a Supply Chain. Submitted to the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

6.2.5.1 Buyers

Each buyer has a set $s \in S$ of potential suppliers and associated rankings for each $\{r_s\}$ ordered such that $r_1 > r_2 > \ldots > r_s > \ldots r_S$. Each potential supplier in the set must meet the basic requirements of the buyer.

The score of each potential supplier (utility accrued by buyer) is calculated using the following formulation:

$$S_{ij} = \sum_{c=1}^{C} W_c \times M_{ci} \quad (1)$$

Where:

S_{ij} : the utility accrued by buyer j when selecting potential supplier i

W_c : the importance weight of criterion c, $c \in (1, \ldots C)$ obtained from Table 6-4

M_{ci} : A normalization value for the measure for supplier i under criterion c

Based on this formulation, each potential supplier will have a score that depends on its measures (production capacity, distance, employment size, and available transportation options) under evaluation criteria.

6.2.5.2 Suppliers

Likewise, each supplier has a set $b \in B$ of potential buyers and associated rankings $\{r_b\}$ ordered such that $r_1 > r_2 > \ldots > r_b > \ldots r_B$. Each potential buyer in the set must meet the basic requirements of the supplier.

6.2.5.3 Market-Clearing Mechanism

The RPA clears the market as follows:

Step 1: The algorithm goes through the list of buyers. For each buyer, it attempts to assign the buyer to its top-priority (r1) supplier. If the supplier has capacity to sell, the selecting buyer gets assigned to the supplier.

Step 2: If the supplier does not have available capacity, the RPA looks to see if the supplier prefers this buyer to others already on its list. If it does prefer it, then it will free up capacity by dropping off the lowest priority buyer that was assigned to it. The selecting buyer is then assigned to the supplier. If the supplier does not prefer the selecting buyer, the next most preferred supplier becomes the top-priority supplier in the algorithm and Step 2 is repeated for the new potential supplier.

Step 3: If a buyer was dropped, then it becomes the selecting buyer in Step 1, selecting a new supplier (and potentially dropping another buyer).

Termination conditions: (i) all buyers have been through Steps 1 and 2, thus, all buyers have ranked their potential suppliers and selected the best one; and (ii) all dropped buyers in Step 3 have been allocated which means these dropped buyers have gone through step 1 and 2 again to find their best supplier.

6.2.6 Model Application and Preliminary Results

The model is coded in the Java object-oriented programming software and implemented in the MAG behavior-based supply chain and freight transportation model using R programming structure. The model replicates supplier selection process at the individual firm level and form firm-to-firm supply chains for all firms in the study area. This section describes the model implementation steps and

preliminary application results. The coded model includes two step, the first step is the "Freight Generation" procedure in which firm-level commodity production and consumptions were calculated, supplier and buyer agents were determined and freight markets were formed. The second step is the application of SSM to each market and determination of annual commodity flows from firm to firm.

6.2.6.1 Freight Generation Procedure

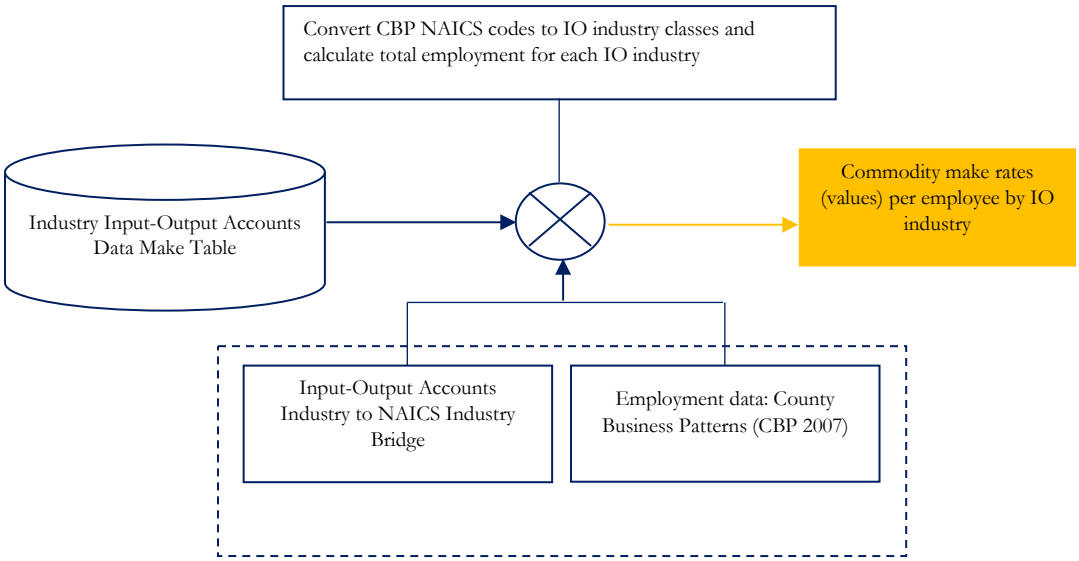
The main input datasets for development and implementation of the model included a complete list of firms (agents) in the country and the commodity Make & Use tables by industry type. These two datasets were processed and combined to enumerate supplier and buyer firms in each market and determine their important characteristics that are influential in the decision making process of selecting suppliers and forming supply chains.

The 2007 Benchmark Industry Input-Output Accounts provided by Bureau of Economic Analysis (BEA) and the regional make –use tables provided by IMPLAN were used to develop commodity make and use (or production and consumption) rates per employee by industry. For this purpose, first, a commodity-industry crosswalk is developed which maps each NAICS industry codes to the commodity type that is generated by the industry. The developed crosswalk is called crosswalk "A". Moreover, a second industry-commodity crosswalk, named crosswalk "B", is developed by the project team which provides a bridge between IO industry code and produced commodity(ies). Crosswalk "B" is developed based on crosswalk "A" and the correspondence table between NAICS and IO industry provided by BEA. These crosswalks are used to determine the type of commodity made and used by each IO and NAICS industry class. Then, the 2007 CBP data and the BEA correspondence table between NAICS and IO industry were used to determine the number of employees for each IO industry class.

In the next step, the calculated total employment for IO industry classes is joined with the make and use tables, and the total monetary make and use amounts are divided by total employment for each IO industry class to obtain average make and use rates per employee. The firm list is merged with the make and use rate tables using common IO industry code. The correspondence table between IO industry codes and NAICS codes was used to convert NAICS industry to IO codes for the firm data. Finally, number of employees of each firm and unite value of commodities provided by FAF4.1 are used to calculate total amount (weights) of commodities made and used by each firm. After that produced-consumed amounts were estimated using the national BEA IO accounts tables, the Arizona-based production and consumption rates derived from IMPLAN data were used to modify commodity production-consumption amounts in the MAG-PAG region. Figure 6-4 and Figure 6-5 present the steps to calculate the production-consumption rates per employee and calculate total commodity made and used by firms.

Figure 6-4 Firm Level Production Estimation Procedure

(a) Development of Make Rates per Employee for IO Industries



(b) Estimation of Made Commodities for Synthesized Firms

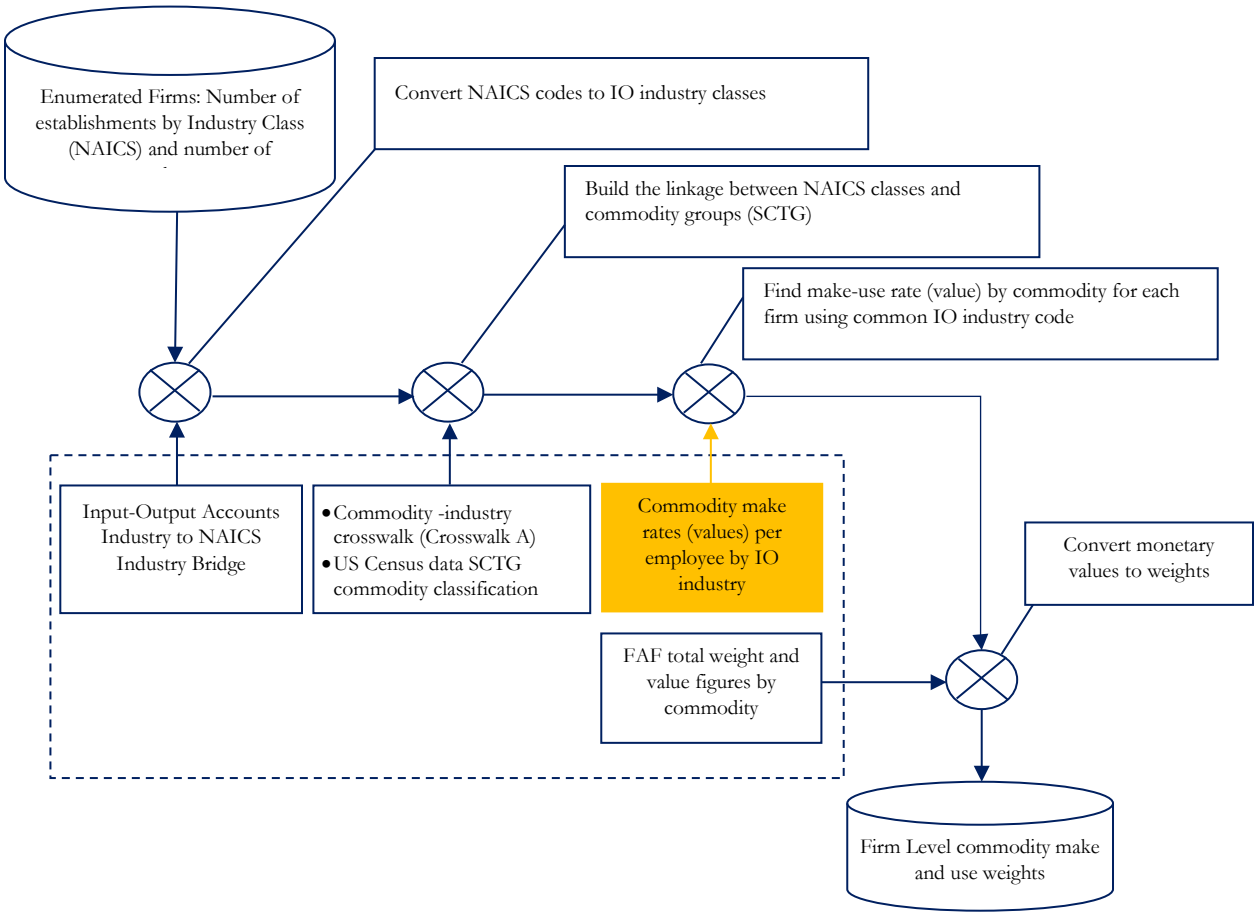
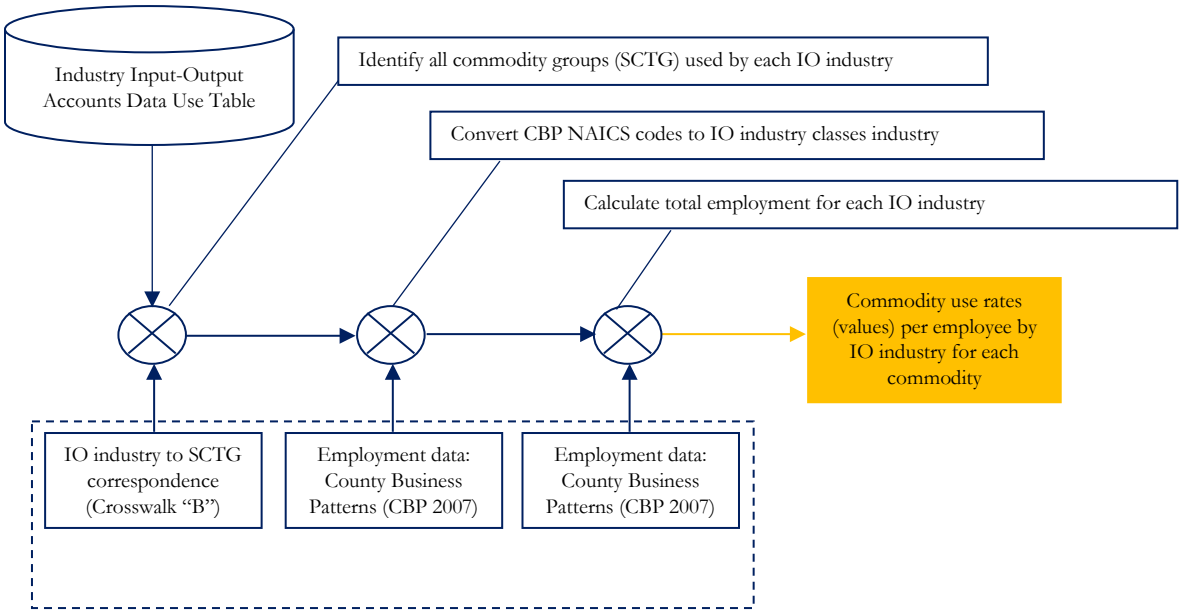
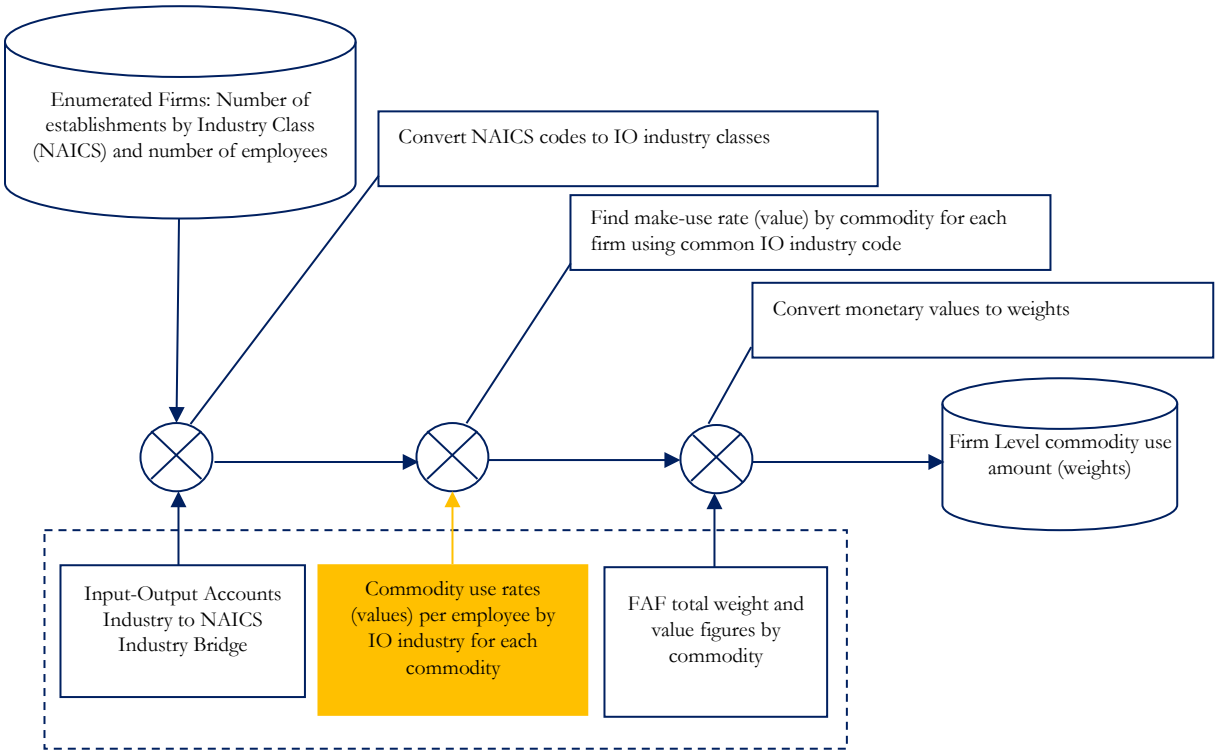


Figure 6-5 Firm Level Consumption Estimation Procedure

(a) Development of Use Rates per Employee for IO Industries



(b) Estimation of Used Commodities for Synthesized Firms



6.2.6.2 Important Note Regarding NAICS Coding Differences

Several input data sources use NAICS system for classification of industries. Major differences exists in NAICS coding in these input datasets. These differences were treated carefully when enumerating supplier/buyer firms and calculating firm-level input-output rates. The first issue was the inconsistency of the NAICS codes across different datasets used in this model. The NAICS coding system has changed from 2007 to 2012. More than 200 industries in the 2007 coding system have either changed to another NAICS code or are aggregated or disaggregated with another industry codes. The input-output accounts provide the bridge between IO industry codes and NAICS codes for the year 2007. The 2013 County Business Patterns data which were used for enumeration of external firms uses a mixture of 2012 and 2007 NAICS codes. Finally, the 2012 CFS Microdata uses only the 2012 NAICS system. Since all these datasets needed to be joined based on the industry class, the development of the crosswalks between IO industry and NAICS codes and between NAICS industry and SCTG codes has been performed with special care to provide the correspondence system between these datasets and to prevent the exclusion of any information in the joining process.

6.2.6.3 Application of SSM to Markets

Once the type and amount of input and output commodities for each firm is estimated, the complete lists of supplier and buyer firms for each market will be generated. The firms then enter the supplier selection algorithm where suppliers are matched with buyers and amounts of commodities traded are estimated.

The developed model was applied and tested for a small market to test the performance of the model. For the initial development, the model only considers two factors in evaluating suppliers: distance and employee size of supplier (which is a proxy for reliability). Once the performance of the model was tested and approved, the model was extended to cover all markets and include all evaluating criteria in the supplier selection process. The final results of model application are presented in Section 6.4 which shows results of calibrated model. The following images show the results of test run for SCTG 22 (Fertilizers).

6.2.6.4 Software Platform and Computation

The enumeration of firms, calculation of production and consumption rates and preparation of the supplier and buyer firms data were developed in R. The supplier selection algorithm was developed in Java and will be invoked and run within the R script. The object-oriented programming languages such as Java are more appropriate for development of agent-based models due to providing the flexibility to define objects and attributes.

The RPA is particularly well suited for speed, and has a flexible structure that can be easily adjusted later if and where necessary. The algorithm itself is not limited in the numbers of potential buyers and suppliers in a market, but the agent-based nature of the model may introduce relatively longer run times as the number of agents increase. The runtime complexity of this algorithm is of O(nb x ns) order where nb and ns are the number of buyers and number of suppliers, respectively. The algorithm, while efficient and concise in terms of coding, can be slowed by the complexity of the problem. Therefore, parallel threading technique was used to spread the processing across parallel threads and reduce run times. It is also worth noting that the solution, stability, and replicability of the algorithm is largely dependent on the structure of buyer and supplier rankings. Repeated solutions could differ slightly from each other if buyer rankings are, for example, nearly identical between suppliers.

6.2.6.5 Data Issues

As with many ACE algorithms, the data must be sufficient and comprehensive in order to ensure that markets clear. In this model, detailed information that sheds light on the decision-making process of supplier selection and identifies the most important criteria affecting the choice of suppliers in supply chains is necessary for designing the ranking procedure and interaction system in the model. It is highly recommended that supplier selection information be gathered later in an establishment survey. The survey questions should be focused on gathering information on the supplier selection process, such as the scoring process by firms of different sizes and industries.

Another data issue can happen where the supplier selection model is over structured with data. For example, in the case that there is sufficient commodity from sellers to ensure that buyers can meet all demand, but limitations on transportation modes may create a situation where there is a lack of net supply to some buyers. Software implementation of the RPA will require results data sets that can ferret out lack of market clearing due to over-structuring of the algorithm.

6.3 Transport, Mode, and Path Choice Model

This section documents the Transport, Mode, and Path Choice model (called Path Choice model hereafter) that has been developed for the MAG/PAG SHRP2-C20 agent-based freight model. The model evaluates mode and shipment size options jointly for each buyer-supplier partnership.

6.3.1 Overview and Data Sources

A disaggregate joint model of mode choice and shipment size was estimated using the Commodity Flow Survey (CFS) 2012 Microdata sample as the primary source. This dataset comprises shipper-based[46] commodity shipments throughout the U.S. in 2012.

In addition, the following sources were used in Path Choice:

- Multimodal transportation levels of service (developed by MAG): travel times, distances, logistics handling nodes (e.g., intermodal terminals and air cargo terminals), and number of rail tracks used (i.e., by track owner);
- Transportation costs for Truck (Less-than-truckload and Truckload), Rail, Air and Parcel. This information is based primarily on Bureau of Transportation Statistics data[47] and supplemented by rates looked up online for this project (See Table 6-5).

[46] The CFS Microdata only contains information about suppliers. There is no information on the buyer. Since a supplier selection model needs information on both buyer and supplier, the data are not appropriate for estimating a Supplier Selection model. The Microdata can, however, be used in calibration and validation of the Supplier Selection model.
[47] http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_03_21.html

Table 6-5 Transportation costs used in Path Choice model development

Assumed Rates Per Ton-Mile	Year		Sources
Mode	2012	2015	
Air	$1.354	N/A	RITA
Class I rail	$0.0395	N/A	RITA
LTL>2,000 lbs. & FTL	$0.18*	$0.20	http://www.freightcenter.com/freight/rail-freight.aspx
LTL, 150-1,500 lbs.	$0.50**	$0.55	http://www.freightcenter.com/freight/rail-freight.aspx
Parcel: <150 lbs.	$2.83***	$3.15	FedEx freight priority website U-Ship website
Parcel/Air, 150+ lbs.	$2.09	N/A	Calculated using an average of Air ($1.354) and Parcel ($2.83) rates
Cass Information Systems*	Truck rates: 2015 rate is 10% higher than 2012 rate		
Cass Information Systems**	IMX rate: 2015 rate is same as 2012 rate		
Assumption*: 50-100 Lbs. Parcel Rates Are Also 10% Less Expensive In 2012**			

Costs were calculated based on routed distance and great circle distance as reported in the CFS Microdata dataset. For application, the skim distances will be used. Table 6-6 summarizes the assumptions that were used to derive costs for each mode. These include ton-mile costs, minimum costs per shipment, and type of distance (routed vs. great circle). The source of the ton-mile costs are listed in Table 6-5. Minimum costs were also established to better reflect realistic shipping cost structures and to avoid computation of trivial costs for short-distance or lightweight shipments. A $10 minimum was used for truck and parcel/air shipments that are less than 150 lbs. A $15 minimum was set for truck and parcel/air shipments weighing 150 lbs. or more. A $200 minimum cost was used for rail.

Table 6-6 Summary of Cost Assumptions

Alternative(S)	Distance**[48]	Cost Per Ton-Mile	Minimum Cost ($)
Rail	Routed Distance	$0.039	200
Truck 1 (<150 lbs.)	Great Circle Distance	$2.83	10
Truck 3 (150-1,499 lbs.)	Great Circle Distance	$0.50	15
Truck 4-5 (1,500+ lbs.)	Great Circle Distance	$0.18	15
Parcel or Air 1 (<150 lbs.)	Great Circle Distance	$2.83	10
Parcel or Air 3-4 (150+ lbs.)	Great Circle Distance	$2.09	15

Table 6-7 shows the assumptions that were used in calculating travel times. For truck and parcel/air trips, the calculation methodology was different for external trips and trips within the Sun Corridor. In general, an assumed terminal time was added to account for loading, unloading, and other terminal activities. The terminal times were based on judgment. Routed distances as reported in the CFS Microdata were used except for external parcel/air trips, for which great circle distance was used.

Travel times were calculated as follows. For internal trips (trips within the MAG/PAG megaregion). Routed shipment distance as reported in the Microdata was used in the travel time calculations. Truck

[48] Source: as reported in CFS Microdata

travel time was calculated as: 4 hours + Routed Distance (mi.) / 20 mph. The extra four-hour period was added to account for loading and unloading. Trips that reportedly used Parcel or Air were assumed to have much faster service: 1 hour + Routed Distance (mi.) / 40 mph.

The following assumptions regarding travel speeds were used in estimating travel times for external (or non-local) trips. A default speed of 22 mph was assumed for rail. Truck speeds were assumed to be 65 mph within a 650-mile driving distance, which is approximately the maximum driving range in a 10-hour day. Distances between 650 and 1300 miles were assumed to be covered in 34 hours (which is two 10-hour driving days plus an overnight 14-hour rest break) for an average speed of 38 mph. Average trucking speeds for distances beyond this range were assumed to be 32 mph, which was derived using similar assumptions. Parcel/Air shipments are assumed to travel at 250 mph on the line-haul portion with some terminal time at each end, which is approximately the average of truck (65 mph) and air (500 mph) line-haul speeds Truck-based parcel shipments typically travel faster than other types of truck shipments since the major parcel carriers (e.g., UPS) typically use multiple distribution points on a route to facilitate smoother and faster flow of goods.

Using these speed assumptions, travel time by mode is calculated as shown in Table 6-7.

Table 6-7 Summary of Travel Time Assumptions

GEO-GRAPHY	MODE	Distance[49]	NOTE	TRAVEL TIME	SPEED (MPH)
ALL TRIPS	Rail	Routed Distance	A maximum of 4 trackage changes was allowed	12 hours + Distance/Speed + 12 hours*Number of Trackage Changes	22
INTRA-SUN CORRIDOR	Truck	Routed Distance	If Microdata distance>200 miles, distance was constrained to equal 100 miles	4 hours + Distance/Speed	20
	Parcel or Air	Routed Distance		1 hour + Distance/Speed	40
EXTERNAL	Truck, 0-650 miles	Routed Distance		16 hours + Distance/Speed	65
	Truck, 651-1,299 miles	Routed Distance		16 hours + Distance/Speed	38
	Truck, 1,300+ miles	Routed Distance		16 hours + Distance/Speed	32
	Parcel or Air	Great Circle Distance	If calculated time exceeds truck time, time is constrained to equal 75 percent of truck time	12 hours + Distance/Speed	250

The rail track penalties capture activities related to switching tracks, waiting for priority trains to clear, and undergoing handling at logistics points such as breakbulk yards. The external travel time calculation

[49] Source: as reported in CFS Microdata

process occasionally permits truck to be faster than Parcel/Air, which in reality is not the case (although parcels are largely transported by truck, their associated charges and levels of service are different from shipping services that are specifically labeled as truck). When this happens, the Parcel/Air time is recalculated as 75% of the Truck time.

6.3.2 CFS Microdata: Descriptive Statistics

This section describes salient characteristics of the Microdata sample that was used as the primary data source for estimation of the path choice model.

The full dataset contains a sample of around 4.5 million shipment records. Approximately 100,000 records are based on shipments that begin and/or end in the State of Arizona. These Arizona-based shipments were used for the model development process. Table 6-8 shows the number of Arizona-based shipment observations by reported mode. Key observations are:

- The sample contains approximately 63,844 truck-only shipments, 662 rail-based (carload or intermodal) shipments, and 36,790 parcel or air shipments.
- To facilitate a straightforward estimation and validation process, similar modes are grouped together as follows. All truck shipments are combined into one mode. Rail-based shipments are grouped together into a single rail mode. Air and parcel are grouped together due to similarities in shipping patterns: shipments by these modes are typically relatively high-value and smaller in size than shipments by other modes. Furthermore, shippers may purchase high-speed service (i.e., air service) while in reality the goods are moved by truck.
- Shipments with insufficient modal information are not used.
- After removing shipments with insufficient modal information, 101,476 observations remained for use in the model estimation process.

Table 6-8 Arizona-Based Shipment Records in the CFS 2012 Microdata Sample

MODE	USE IN MODEL ESTIMATION AND APPLICATION	N
TRUCK	Yes (Combined To Make One Truck-Only Mode)	1
FOR-HIRE TRUCK		35,038
PRIVATE TRUCK		28,805
RAIL	Yes	501
TRUCK AND RAIL	Yes	161
AIR (INCL TRUCK & AIR)	Yes – Combined With Parcel	2,863
PARCEL, USPS, OR COURIER	Yes – Combined With Air	34,107
MODE SUPPRESSED	No - Unknown Mode	5
SINGLE MODE	No - Unknown Mode	221
MULTIPLE MODE	No - Unknown Mode	135
TRUCK AND WATER	No - Water Not Included	1
NON-PARCEL MULTIMODE	No - Unknown Mode	4
ALL SHIPMENT RECORDS		101,842
SHIPMENT RECORDS WITH COMPLETE MODE INFORMATION		101.476

Additional processing was used to remove records with insufficient or unreasonable information. These included observations with:

- Missing commodity code;
- Unknown origin or destination region;
- Missing region on the Arizona end of the trip – that is, only records with known location in Arizona were kept;
- Certain shipments were negligible in number but were creating sample size-related problems in the estimation process:
    - ✓ Air shipments over 15,000 lbs.;
    - ✓ Rail shipments under 1,500 lbs.;
    - ✓ Rail shipments in the high-value category;
    - ✓ Parcel/air shipments over 150 pounds in the low-value category.

Based on these criteria, 775 records were eliminated from the estimation dataset. A total of 100,701 observations remained for model estimation.

Table 6-9 shows the tons and values reported for these shipments by mode using all Arizona-based records. The unweighted values represent the sample statistics and the weighted values represent the expanded [50] figures that represent the estimated totals and means of all Arizona-based shipments. The model estimation is based on the unweighted sample. The calibration and validation effort may use statistics from either the unweighted or expanded data. This table shows that, as expected, the largest shipments occur by rail with mean shipment sizes of 1,365 tons (sample) and 222 tons (weighted). Rail-truck intermodal (IMX) also carries relatively heavy shipments with a mean of 55 tons (sample) and 21 tons (weighted). Mean truck shipments are less than 10 tons and mean parcel and air shipments are less than 200 pounds. Weighted mode shares are comparable to national mode shares reported in the 2012 FAF3 data, which indicates that truck carries about 80-85% of goods by weight nationally and rail about 10-15%.

Table 6-9 also shows that truck, air and parcel modes carry a disproportionately high volume of goods by value (compared to goods by weight). This is demonstrated by the Value per Pound metric, which is the ratio of shipment value to shipment weight. Air and parcel modes especially are used for high-value shipments and rail modes for low-value shipments. Table 6-10 shows the domestic origins and destinations of the Arizona-based shipment records. The shipment records are well distributed across

[50] The weighting factors were developed by the Bureau of Transportation Statistics (BTS) and are included in the Microdata sample dataset. The Microdata Users' Guide contains information on how certain weights were obfuscated to improve confidentiality. However, the Guide does not document how the factors were developed. The factors may be based on the seven-component weighting factors that are discussed in the 2012 Commodity Flow Survey Methodology (Revised 12/03/2014). The RS&H, Inc. team has contacted BTS for more information. The CFS documentation states: "*CFS respondents provided data for a sample of shipments made by their respective establishments in the survey year. For each establishment, we produced an estimate of that establishment's total value of shipments for the entire survey year. To do this, we used four different weights: the shipment weight, the shipment nonresponse weight, the quarter weight, and the quarter nonresponse weight. Three additional weights are then applied to produce estimates representative of the entire universe. These are the establishment-level adjustment weight, the establishment (or first-stage sample) weight, and the nonresponse post-stratification adjustment weight.*"

the U.S. Appendix A shows which states and metropolitan regions are included in each summary region. Export shipments are examined in greater detail later.

Table 6-9 Shipment Weights and Values by Mode: Unweighted (Sample) And Weighted (Expanded) [51]

Mode	N	UNWEIGHTED			WEIGHTED			UNWEIGHTED			WEIGHTED			MEAN VALUE PER POUND (UNWTD.)
		Total Tons	% Total Tons	Mean Tons per Shipment	Total Tons	% Total Tons	Mean Tons per Shipment	Total Value ($1,000)	% Total Value	Mean Value per Shipment ($)	Total Value ($1,000)	% Total Value	Mean Value per Shipment ($)	
MODE SUPPRESSED	5	291	0.03	58	40,267	0.03	6	308	0.02	61,578	61,744	0.02	9,579	0.76
SINGLE MODE	221	7,955	0.72	36	1,339,630	0.88	17	12,181	0.88	55,116	1,092,886	0.42	13,854	84.46
TRUCK	1	0.01	-	0.01	9	0	0.012	0.028	0	28	23	0.00	28	1.22
FOR-HIRE TRUCK	35,038	279,882	25.22	8	85,793,463	56.36	2.2	834,209	60.12	23,809	131,950,822	51.23	3,457	47.56
PRIVATE TRUCK	28,805	124,327	11.2	4	41,601,143	27.33	0.8	160,732	11.58	5,580	52,184,551	20.26	1,040	14.61
RAIL	501	683,915	61.64	1,365	20,410,114	13.41	222	53,791	3.88	107,367	4,726,732	1.84	51,517	0.76
AIR (INCL TRUCK & AIR)	2,863	263	0.02	0.09	0.09	0.04	0.00002	218,358	15.74	76,269	12,885,929	5	2,258	2,672.67
MULTIPLE MODE	135	3,581	0.32	27	27	0.38	0	4,071	0.29	30,152	334,773	0.13	7,380	20.51
PARCEL, USPS, OR COURIER	34,107	297	0.03	0.01	0.01	0.45	0.00023	94,072	6.78	2,758	52,104,027	20.23	292	362.48
TRUCK AND RAIL	161	8,915	0.8	55	1,703,572	1.12	21	9,382	0.68	58,270	2,207,803	0.86	27,360	3.55
TRUCK AND WATER	1	20	-	20	1,403	0	20	29	0.00	29,409	2,073	0.00	29,409	0.74
NON-PARCEL MULTIMODE	4	147	0.01	37	9,891	0.01	28	367	0.03	91,698	33,881	0.01	95,306	1.55

[51] Source: CFS 2012 Microdata summarized by the RS&H, Inc.

Table 6-10 Origin and Destination Regions of Shipments in the Arizona-Based Shipment Dataset [52]

ORIGIN REGION	DESTINATION REGION															
	Sun Corridor - Megaregion Tucson	Remainder of Arizona	California - Southern	California - Other	Nevada	Utah	Colorado	New Mexico	Pacific Northwest	Central-Upper Midwest	Great Lakes	Northeast U.S.	Texas and Oklahoma	Southeast U.S.	Alaska and Hawaii	Total
SUN CORRIDOR - PHOENIX AND TUCSON	26,498	2,407	3,868	1,585	633	407	562	735	1,084	1,127	2,150	2,506	2,617	3,071	76	49,326
REMAINDER OF ARIZONA	1,407	9,076	1,381	889	468	205	321	287	429	479	636	625	665	838	30	17,736
CALIFORNIA - SOUTHERN	4,210	576	-	-	-	-	-	-	-	-	-	-	-	-	-	4,786
CALIFORNIA - OTHER	2,235	436	-	-	-	-	-	-	-	-	-	-	-	-	-	2,671
NEVADA	888	280	-	-	-	-	-	-	-	-	-	-	-	-	-	1,168
UTAH	932	189	-	-	-	-	-	-	-	-	-	-	-	-	-	1,121
COLORADO	639	194	-	-	-	-	-	-	-	-	-	-	-	-	-	833
NEW MEXICO	446	286	-	-	-	-	-	-	-	-	-	-	-	-	-	732
PACIFIC NORTHWEST	1,356	242	-	-	-	-	-	-	-	-	-	-	-	-	-	1,598
CENTRAL-UPPER MIDWEST	2,370	581	-	-	-	-	-	-	-	-	-	-	-	-	-	2,951
GREAT LAKES	4,565	698	-	-	-	-	-	-	-	-	-	-	-	-	-	5,263
NORTHEAST U.S.	4,708	558	-	-	-	-	-	-	-	-	-	-	-	-	-	5,266
TEXAS AND OKLAHOMA	2,169	481	-	-	-	-	-	-	-	-	-	-	-	-	-	2,650
SOUTHEAST U.S.	4,624	769	-	-	-	-	-	-	-	-	-	-	-	-	-	5,393
ALASKA AND HAWAII	65	17	-	-	-	-	-	-	-	-	-	-	-	-	-	82
ARIZONA-UNKNOWN REGION	43	28	36	9	1	4	3	5	10	3	16	13	19	32	23	245
CALIFORNIA-UNKNOWN REGION	17	4	-	-	-	-	-	-	-	-	-	-	-	-	-	21
TOTAL	57,172	16,822	5,285	2,483	1,102	616	886	1,027	1,523	1,609	2,802	3,144	3,301	3,941	129	101,842

FREQUENCY MISSING = 22

[52] Source: CFS 2012 Microdata summarized by the RS&H, Inc.

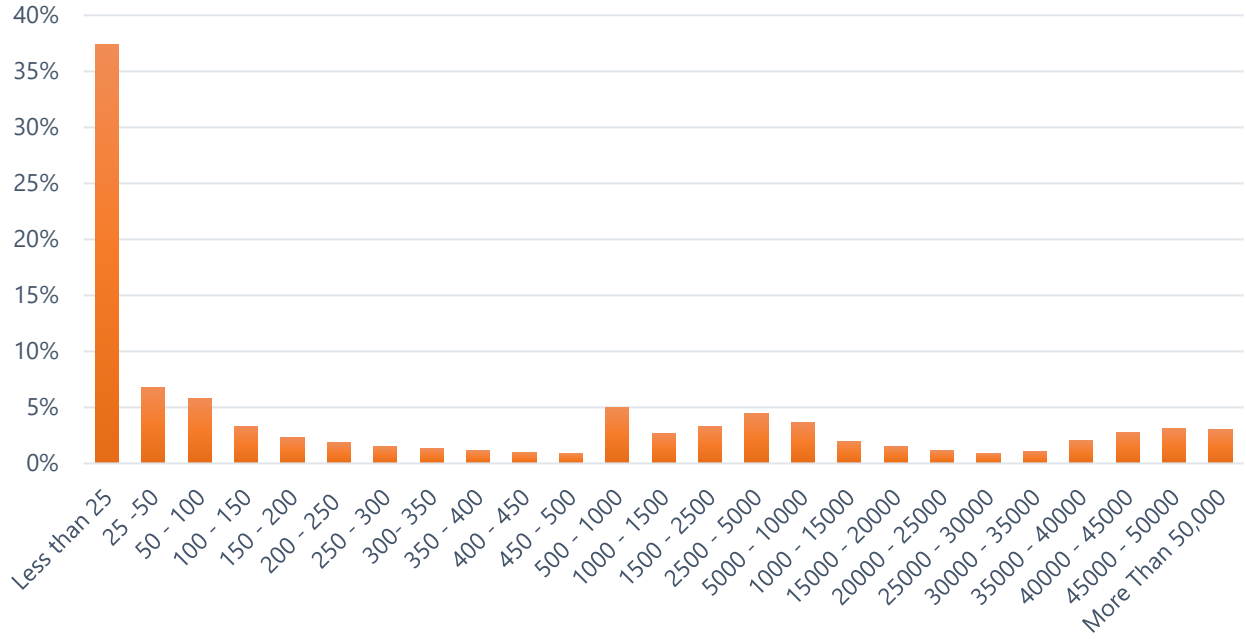
Shipment size is modeled jointly with mode in this Path Choice formulation. Four shipment size categories are used in the estimation:

- Less than 150 pounds;
- 150-1,499 pounds;
- 1,500-34,999 pounds; and
- 35,000+ pounds.

The size thresholds were selected based on several attributes of the estimation dataset. First, Figure 6-6 presents distribution of shipment weights over size categories. As it can be more than 40% of flows are small size shipments of less than 50 pounds and the frequency shipments drops as the shipment size increases. Figure 6-7 shows the percentage of shipments by mode across a range of size categories. This figure demonstrates that while truck shipments are distributed relatively evenly across the categories, rail and parcel/air are heavily skewed to the left and right, respectively. As such, the shipment size thresholds generate a sound sample size for each alternative.

Second, the selected size categories accommodate differential pricing by weight for truck shipments. Figure 6-8 shows that the cost per ton-mile is highest for shipments under 1,500 lbs. and levels off above 1,500 lbs. This supports the use of 1,500 lbs. as a threshold. Third, parcel carriers typically handle shipments that are (or can be) broken down into parcels of 150 pounds or less [53], which supports using 150 pounds as a threshold. Table 6-11 shows the resulting sample sizes in each mode and shipment size category.

Figure 6-6 Number of Shipments by Size Categories



[53] See https://en.wikipedia.org/wiki/Less_than_truckload_shipping

Figure 6-7 Percentages of Shipments by Mode Across Chosen Size Categories

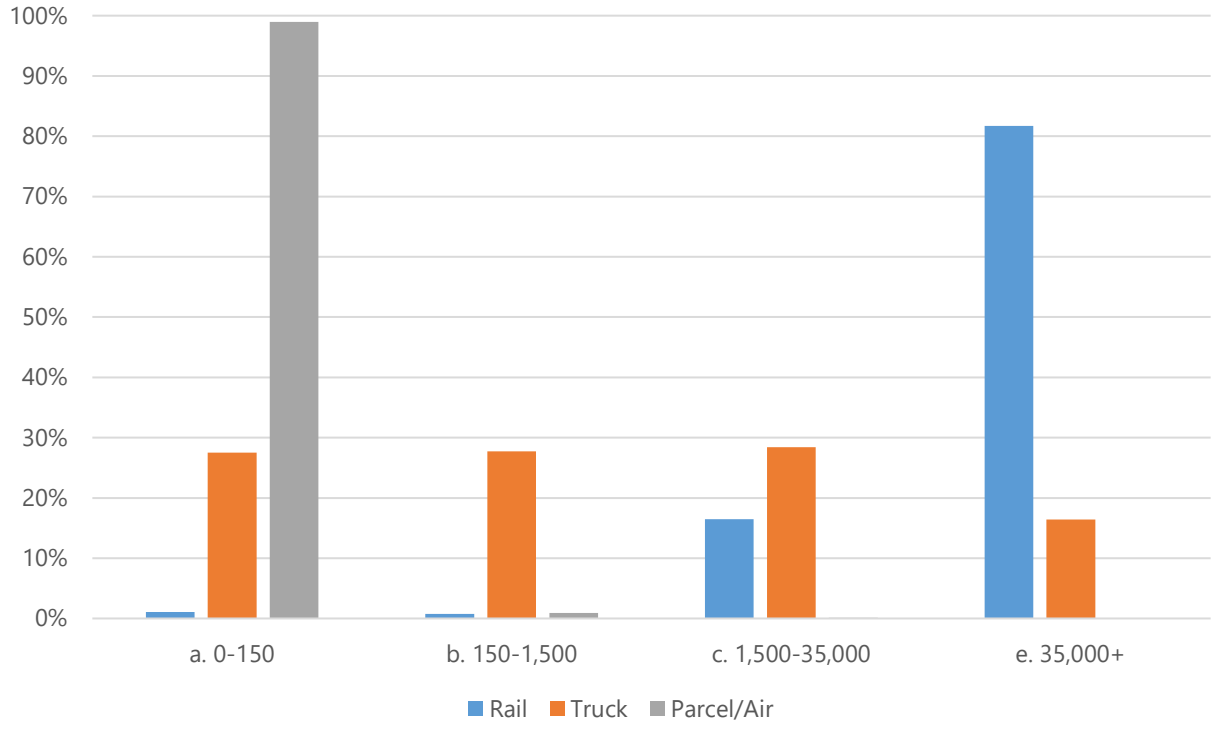
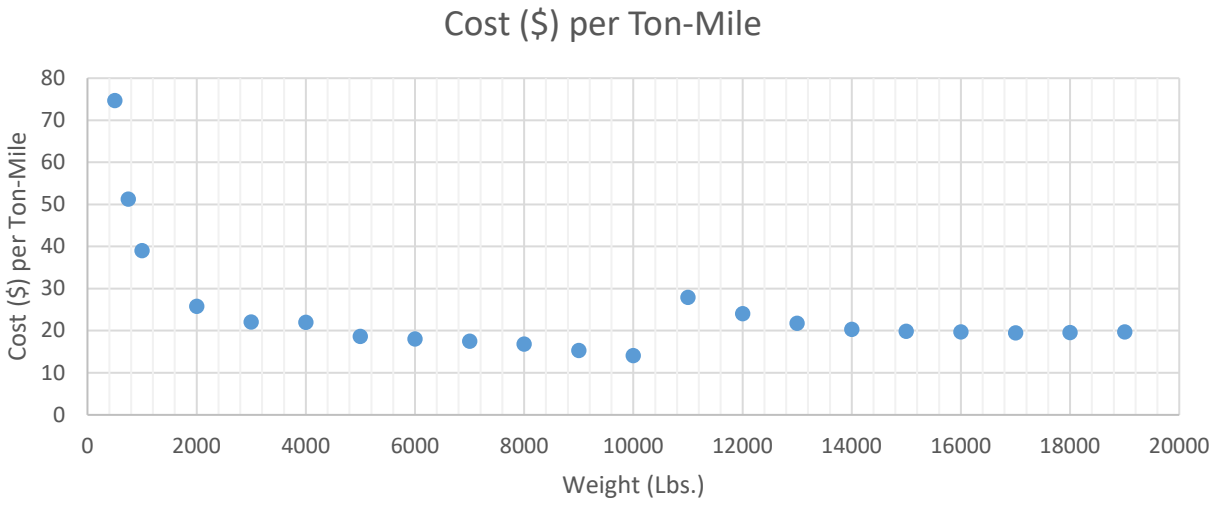


Figure 6-8 Costs per Ton-Mile by Shipment Size [54]



[54] Source: RS&H, Inc. using information from http://www.freightcenter.com/freight/rail-freight.aspx

Table 6-11 Sample Size: Arizona-Based Shipments by Mode and Size Category

MODE CATEGORIES	SHIPMENT SIZE CATEGORIES				
	< 150 lbs.	150-1,499 lbs.	1,500-34,999 lbs.	35,000+ lbs.	All Sizes
RAIL CARLOAD/IMX	-	-	96	518	614
TRUCK	17,506	17,693	17,914	10,137	63,250
PARCEL/AIR	36,479	333	25	-	36,837
ALL MODES	53,985	18,026	18,035	10,655	100,701

The CFS Microdata captured exports to Canada, Mexico and other areas (see Table 6-12). The destination country was specified only for Canada and Mexico. In total, 3,373 Arizona-based export shipments were captured (3.3% of the Arizona-based estimation sample) with 818 shipments destined to Mexico, 513 to Canada, and 2,042 to other countries. In general, robust sample sizes of export shipments were captured for Truck and Parcel/Air. However, no export shipments were captured for the Rail/IMX 1,500-34,999 pound category and a negligible number of observations were recorded for the Rail/IMX 35,000+ pound and Parcel/Air 1,500-34,999 pound categories. The Parcel/Air 150-1,499 pound category also had a limited number of shipments destined to Mexico and Canada.

Table 6-12 Sampled Export Shipments by Alternative and Destination

ALTERNATIVE	EXPORT DESTINATION				% OF EXPORT SHIPMENTS BY ALTERNATIVE	TOTAL SAMPLE SIZE
	Mexico	Canada	Other	Total		
RAIL/IMX, 1,500-34,999 LBS.	-	-	-	-	0.0%	96
RAIL/IMX, 35,000+ LBS.	2	1	-	3	0.6%	518
TRUCK, <150 LBS.	73	58	24	155	0.9%	17,506
TRUCK, 150-1,499 LBS.	125	73	58	256	1.4%	17,693
TRUCK, 1,500-34,999 LBS.	270	90	117	477	2.7%	17,914
TRUCK, 35,000+ LBS.	248	27	71	346	3.4%	10,137
PARCEL/AIR, <150 LBS.	92	258	1,588	1,938	5.3%	36,479
PARCEL/AIR, 150-1,499 LBS.	8	4	170	182	54.7%	333
PARCEL/AIR, 1,500-34,999 LBS.	-	2	14	16	64.0%	25
TOTAL	818	513	2,042	3,373	3.3%	100,701

For analysis purposes, the SCTG commodities were aggregated into four groups: Bulk Goods; Fuels and Fertilizers; Intermediate Products, Foodstuffs and Miscellaneous Goods; and Manufactured Goods (these groupings are discussed in more detail later in this report). Table 6-13 shows the types of commodities by alternative that were captured in the CFS export shipments.

Table 6-13 shows that most of the observed export shipments were in the Intermediate Products, Foodstuffs and Misc. or Manufactured Goods categories. Approximately half of all observed export shipments in the sample are manufactured goods that are shipped using Parcel or Air. Very few Bulk Goods and Fuels and Fertilizers exports were observed in the CFS; these shipments were observed almost exclusively in the Truck 35,000+ pound category.

While the export sample provides sound information, it also has limitations for the Path Choice model:

- Imports are not included in the CFS. The Path Choice model will be applied to domestic leg of the export shipments.
- The CFS coverage of international Rail/IMX shipments was too small to use in model estimation. Adjustments were made during calibration and validation.
- The CFS does not cover all commodities. Excluded commodities may have significant international shipment activity. These include: Live animals and fish; Cereal grains; Other agricultural products; Calcareous monumental or building stone; Crude petroleum; Logs and wood in the rough; and Machinery. This was addressed during calibration and validation.

Table 6-13 Sampled Export Shipments by Alternative and Commodity Category

ALTERNATIVE	COMMODITY CATEGORY				
	Bulk Goods	Fuels And Fertilizers	Intermediate Products, Foodstuffs And Misc.	Manufactured Goods	Total
RAIL/IMX, 35,000+ LBS.	-	-	3	-	3
TRUCK, <150 LBS.	-	-	24	131	155
TRUCK, 150-1,499 LBS.	1	2	72	181	256
TRUCK, 1,500-34,999 LBS.	6	6	163	302	477
TRUCK, 35,000+ LBS.	73	9	141	123	346
PARCEL/AIR, <150 LBS.	1	1	241	1,695	1,938
PARCEL/AIR, 150-1,499 LBS.	-	-	33	149	182
PARCEL/AIR, 1,500-34,999 LBS.	-	-	15	1	16
TOTAL	81	18	692	2,582	3,373

Table 6-14 shows the value-per-pound thresholds used in model estimation. These breakpoints were helpful in distinguishing among differential commodity assignments among modes. The number of individual shipments from the CFS Microdata sample is also shown.

Table 6-14 Value per Pound Categories Used in Model Estimation [55]

MODE	VALUE ($) PER POUND			TOTAL SAMPLE SHIPMENTS	% SAMPLE SHIPMENTS
	Vdlow (<$1/Lb.)	Vdmed ($1-$3.99)	Vdhigh ($4+/Lb.)		
RAIL/IMX	83%	17%	0%	614	0.60%
TRUCK	38%	30%	32%	63,250	63%
AIR/PARCEL	1%	8%	91%	36,837	36%
TOTAL NUMBER OF SHIPMENTS IN ESTIMATION SAMPLE	24,941	22,166	53,594	100,701	100%
% SAMPLE SHIPMENTS BY VALUE	25%	22%	53%	100%	

[55] Source: CFS 2012 Microdata summarized by the RS&H, Inc.

Since the MAG/PAG SHRP-2 C20 model is a disaggregate model, the values-per-pound of individual shipments are used in the Path Choice model development rather than aggregate values-per-pound. However, for comparison to the disaggregate shipment value-per-pound ratios in Table 6-14, aggregate figures were tabulated based on all shipments within each SCTG category to generate the average value-per-pound of goods in each SCTG category (see Table 6-15). Unweighted and weighted averages are shown for all Microdata shipments and for the Arizona-based shipments that are used in the Path Choice model. The average values range from about one cent per pound for Sand and Gravel to roughly $60-$80 per pound for Arizona-based Transportation Equipment (not elsewhere classified) and Precision Instruments.

In addition, a comparison of Table 6-14 and Table 6-15 demonstrates that while only 25 percent of individual shipments are valued at less than one dollar per pound, the averages across all shipments within a SCTG category are typically less than one dollar per pound. The most likely explanation for this is that the average is weighted down by heavier, lower-value shipments within a given SCTG category. The aggregate values in Table 6-15 were used in model calibration and validation.

Table 6-15 Sample Shipments: Average Value per Pound by SCTG Category

SCTG	DESCRIPTION	AVERAGE VALUE PER POUND			
		Unweighted		Weighted	
		Entire Sample	Arizona Sample	Entire Sample	Arizona Sample
1	Animals and Fish (live)	$1.29	$0.77	$1.41	$0.87
2	Cereal Grains (includes seed)	$0.16	$0.14	$0.14	$0.16
3	Agricultural Products (excludes Animal Feed, Cereal Grains, and Forage Products)	$0.33	$0.55	$0.47	$0.81
4	Animal Feed, Eggs, Honey, and Other Products of Animal Origin	$0.31	$0.48	$0.25	$0.49
5	Meat, Poultry, Fish, Seafood, and Their Preparations	$1.75	$2.14	$1.68	$2.18
6	Milled Grain Products and Preparations, and Bakery Products	$0.63	$0.94	$0.67	$0.81
7	Other Prepared Foodstuffs, and Fats and Oils	$0.71	$0.83	$0.56	$0.65
8	Alcoholic Beverages and Denatured Alcohol	$0.97	$0.90	$0.90	$0.93
9	Tobacco Products	$13.77	$15.30	$11.53	$16.89
10	Monumental or Building Stone	$0.18	$0.09	$0.19	$0.12
11	Natural Sands	$0.022	$0.008	$0.009	$0.012
12	Gravel and Crushed Stone (excludes Dolomite and Slate)	$0.006	$0.010	$0.006	$0.004
13	Other Non-Metallic Minerals not elsewhere classified	$0.043	$0.044	$0.046	$0.031
14	Metallic Ores and Concentrates	$0.09	$1.60	$0.17	$1.13
15	Coal	$0.02	$0.02	$0.02	$0.02
17	Gasoline, Aviation Turbine Fuel, and Ethanol (includes Kerosene, and Fuel Alcohols)	$0.45	$0.43	$0.46	$0.39
18	Fuel Oils (includes Diesel, Bunker C, and Biodiesel)	$0.42	$0.49	$0.42	$0.47
19	Other Coal and Petroleum Products, not elsewhere classified	$0.27	$0.34	$0.37	$0.52
20	Basic Chemicals	$0.83	$0.61	$0.47	$0.47
21	Pharmaceutical Products	$43.36	$21.43	$25.25	$29.24
22	Fertilizers	$0.27	$0.27	$0.22	$0.25
23	Other Chemical Products and Preparations	$1.70	$1.45	$1.67	$2.11
24	Plastics and Rubber	$1.33	$1.59	$1.50	$1.83
25	Logs and Other Wood in the Rough	$0.12	$0.10	$0.06	$0.19
26	Wood Products	$0.32	$0.38	$0.24	$0.40
27	Pulp, Newsprint, Paper, and Paperboard	$0.55	$0.54	$0.50	$0.61
28	Paper or Paperboard Articles	$0.84	$0.87	$0.83	$0.91
29	Printed Products	$1.90	$1.75	$2.06	$2.70
30	Textiles, Leather, and Articles of Textiles or Leather	$3.98	$4.10	$5.89	$4.99
31	Non-Metallic Mineral Products	$0.17	$0.16	$0.11	$0.12
32	Base Metal in Primary or Semi-Finished Forms and in Finished Basic Shapes	$1.00	$1.37	$0.76	$1.04
33	Articles of Base Metal	$1.34	$1.43	$1.75	$1.99
34	Machinery	$6.79	$10.01	$5.07	$9.37
35	Electronic and Other Electrical Equipment and Components, and Office Equipment	$10.30	$19.75	$11.20	$9.97
36	Motorized and Other Vehicles (includes parts)	$4.63	$5.90	$3.98	$4.19
37	Transportation Equipment, not elsewhere classified	$59.22	$57.66	$20.04	$81.06
38	Precision Instruments and Apparatus	$8.08	$81.50	$24.11	$66.84
39	Furniture, Mattresses and Mattress Supports, Lamps, Lighting Fittings, and Illuminated Signs	$2.66	$1.99	$2.92	$3.02
40	Miscellaneous Manufactured Products	$4.88	$14.30	$3.51	$9.01
43	Mixed Freight	$2.08	$2.52	$1.84	$2.22
AVERAGE (EXCLUDES WASTE/SCRAP AND UNKNOWN COMMODITIES)		$0.50	$0.56	$0.63	$0.85

For analysis purposes, commodities are distinguished using four groups: Bulk Goods; Fuels and Fertilizers; Intermediate Products, Foodstuffs and Miscellaneous Goods; and Manufactured Goods (Table 6-16). Goods in each of these categories generally are similar in terms of:

- Type of good (e.g., bulk vs. finished product);
- Share of shipments by mode and shipment size; and
- Value per pound.

Table 6-17, Table 6-18 and Table 6-19 show the sample sizes in for each commodity by mode and shipment size category.

Figure 6-9 to Figure 6-12 supplement Table 6-17, Table 6-18 and Table 6-19 by demonstrating the mode and shipment size patterns across the four commodity groupings. For illustration purposes, the mode and shipment size categories are grouped into one Rail category, a small-shipment truck category (Truck shipment size < 1,500 lbs.), a larger-shipment truck category (Truck shipment size 1,500+ lbs.), and a single Parcel/Air category.

As expected, these figures show that Bulk Goods are shipped mostly in larger shipments with relatively high truck and rail use. Truck is typically used for 80% or more of the shipments, although Coal shipments predominantly use Rail. Parcel and Air are used infrequently, presumably for more specialized shipments that are categorized in these SCTG groups.

Mode and shipment size usage patterns for Fuels and Fertilizers shipments are similar to those of Bulk Goods, with heavier Truck shipments and Rail commonly used. However, unlike Bulk goods, the goods in this category use Truck for a greater percentage of small shipments.

The next category (Intermediate Products, Foodstuffs and Miscellaneous Goods), generally consists that have been through some processing or manufacturing but are not considered finished products. Miscellaneous SCTG commodities, such as Animals and Fish, are included here as well. Foods are also included in this category due to the similarities in shipping mode patterns: the goods in this category predominantly use Truck with a lesser amount of Parcel/Air and infrequent Rail use.

Finally, the Manufactured Goods category primarily consists of finished products. Shipments of these goods are typically small and have substantial Parcel/Air usage and Truck usage. Some larger shipments are made using Truck.

Table 6-16 Commodity Categories by SCTG Code

Bulk Goods	Fuels And Fertilizers	Intermediate Products, Foodstuffs And Misc.	Manufactured Goods
2. Cereal Grains (Includes Seed)	17. Gasoline, Aviation Turbine Fuel, and Ethanol	1. Animals and Fish (live)	9. Tobacco Products
4. Animal Feed, Eggs, Honey, And Other Products Of Animal Origin	18. Fuel Oils (includes Diesel, Bunker C, and Biodiesel)	3. Agricultural Products (excludes Animal Feed, Cereal Grains, and Forage Products)	21. Pharmaceutical Products
11. Natural Sands	22. Fertilizers	5. Meat, Poultry, Fish, Seafood, and Their Preparations	23. Other Chemical Products and Preparations
12. Gravel And Crushed Stone		6. Milled Grain and Bakery Products	29. Printed Products
13. Other Non-Metallic Minerals N.E.C.		7. Other Prepared Foodstuffs, and Fats and Oils	30. Textiles, Leather, and Articles of Textiles or Leather
14. Metallic Ores And Concentrates		8. Alcoholic Beverages and Denatured Alcohol	33. Articles of Base Metal
15. Coal		10. Monumental or Building Stone	34. Machinery
25. Logs And Other Wood In The Rough		19. Other Coal and Petroleum Products n.e.c.	35. Electronic and Office Equipment
		20. Basic Chemicals	36. Motorized and Other Vehicles (incl. parts)
		24. Plastics and Rubber	37. Transportation Equipment, n.e.c.
		26. Wood Products	38. Precision Instruments and Apparatus
		27. Pulp, Newsprint, Paper, and Paperboard	39. Furniture, Mattresses (and Supports) Lighting and Illuminated Signs
		28. Paper or Paperboard Articles	40. Misc. Manufactured Products
		31. Non-Metallic Mineral Products	43. Mixed Freight
		32. Base Metal in Forms and Basic Shapes	

Table 6-17 Number of Shipments by SCTG and Mode/Shipment Size Category (Sample)

COMMODITY CATEGORY	SCTG	MODE AND SHIPMENT SIZE CATEGORY								
		Rail Carload / IMX		Truck				Parcel / Air		
		1,500-34,999 Lbs.	35,000+ Lbs.	<150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.	35,000+ Lbs.	<150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.
BULK GOODS	2. Cereal Grains (Includes Seed)	-	6	-	5	10	232	1	-	-
	4. Animal Feed, Eggs, Honey, And Other Products Of Animal Origin	-	17	5	21	21	611	36	-	-
	11. Natural Sands	-	-	2	21	51	546	-	-	-
	12. Gravel And Crushed Stone (Excludes Dolomite And Slate)	1	30	26	22	97	1,768	-	-	-
	13. Other Non-Metallic Minerals Not Elsewhere Classified	-	13	2	15	13	268	18	-	-
	14. Metallic Ores And Concentrates	-	31	1	1	7	301	11	-	-
	15. Coal	-	35	-	-	-	9	-	-	-
	25. Logs And Other Wood In The Rough	-	-	-	11	20	60	1	-	-
	Subtotal	1	132	36	96	219	3,795	67	-	-
FUELS AND FERTILIZERS	17. Gasoline, Aviation Turbine Fuel, And Ethanol (Includes Kerosene, And Fuel Alcohols)	-	18	18	23	126	211	-	-	-
	18. Fuel Oils (Includes Diesel, Bunker C, And Biodiesel)	-	19	11	83	301	414	-	-	-
	22. Fertilizers	-	42	27	88	155	142	34	-	-
	Subtotal	-	79	56	194	582	767	34	-	-

Table 6-18 Number of Shipments by SCTG and Mode/Shipment Size Category (Sample)

COMMODITY CATEGORY	SCTG	MODE AND SHIPMENT SIZE CATEGORY								
		Rail Carload / IMX		Truck				Parcel / Air		
		1,500-34,999 Lbs.	35,000+ Lbs.	<150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.	35,000+ Lbs.	<150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.
INTERMEDIATE PRODUCTS, FOODSTUFFS AND MISC.	1. Animals And Fish (Live)	-	-	1	-	-	3	2	-	-
	3. Agricultural Products	-	1	558	259	345	365	151	-	-
	5. Meat, Poultry, Fish, Seafood, And Preparations	-	1	125	196	203	223	14	1	-
	6. Milled Grain And Bakery Products	2	15	114	126	290	258	21	-	-
	7. Other Prepared Foodstuffs, And Fats And Oils	-	24	251	823	631	1,665	543	1	-
	8. Alcoholic Beverages And Denatured Alcohol	16	17	480	1,255	533	178	44	-	-
	10. Monumental Or Building Stone	-	-	8	42	25	111	-	-	-
	19. Other Coal And Petroleum Products, N.E.C.	-	17	468	1,686	486	809	29	-	-
	20. Basic Chemicals	-	39	246	393	247	591	638	14	17
	24. Plastics And Rubber	48	25	1,057	1,476	1,528	783	1,555	21	7
	26. Wood Products	4	41	228	641	554	781	187	-	-
	27. Pulp, Newsprint, Paper, And Paperboard	2	23	69	136	82	309	84	12	-
	28. Paper Or Paperboard Articles	2	2	201	379	338	599	448	2	-
	31. Non-Metallic Mineral Products	1	46	341	421	406	1,388	403	-	-
	32. Base Metal In Forms And Shapes	8	53	583	966	660	896	475	6	-
	Subtotal	83	304	4,730	8,799	6,328	8,959	4,594	57	24

Table 6-19 Number of Shipments by SCTG and Mode/Shipment Size Category (Sample)

COMMODITY CATEGORY	SCTG	MODE AND SHIPMENT SIZE CATEGORY								
		Rail Carload / IMX		Truck				Parcel / Air		
		1,500-34,999 Lbs.	35,000 + Lbs.	<150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.	35,000 + Lbs.	<150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.
MANUFACTURED GOODS	9. Tobacco Products	-	-	6	27	5	4	38	-	-
	21. Pharmaceutical Products	-	-	623	235	79	91	1,325	6	-
	23. Other Chemical Products And Preparations	2	1	760	746	399	270	1,001	10	1
	29. Printed Products	-	-	1,381	600	189	109	2,650	6	-
	30. Textiles, Leather, And Articles Of Textiles Or Leather	1	2	593	565	206	192	3,293	13	-
	33. Articles Of Base Metal	1	-	1,194	1,070	573	558	1,758	18	-
	34. Machinery	1	-	1,173	945	314	148	2,435	22	-
	35. Electronic And Other Electrical Equipment And Components, And Office Equipment	-	-	2,222	1,115	348	149	6,125	90	-
	36. Motorized And Other Vehicles (Includes Parts)	2	-	453	225	225	299	1,119	3	-
	37. Transportation Equipment, Not Elsewhere Classified	-	-	251	162	48	55	2,492	26	-
	38. Precision Instruments And Apparatus	-	-	376	214	44	16	2,541	28	-
	39. Furniture, Mattresses Lighting, And Illuminated Signs	2	-	561	739	351	135	442	2	-
	40. Misc. Manufactured Products	-	-	695	810	313	352	4,928	49	-
	43. Mixed Freight	3	-	2,396	1,151	1,163	766	1,637	3	-
	Subtotal	12	3	12,684	8,604	4,257	3,144	31,784	276	1
TOTAL SHIPMENTS (SUM = 100,701)		96	518	17,506	17,693	11,386	16,665	36,479	333	25

Figure 6-9 Mode and Shipment Size Share for Bulk Goods

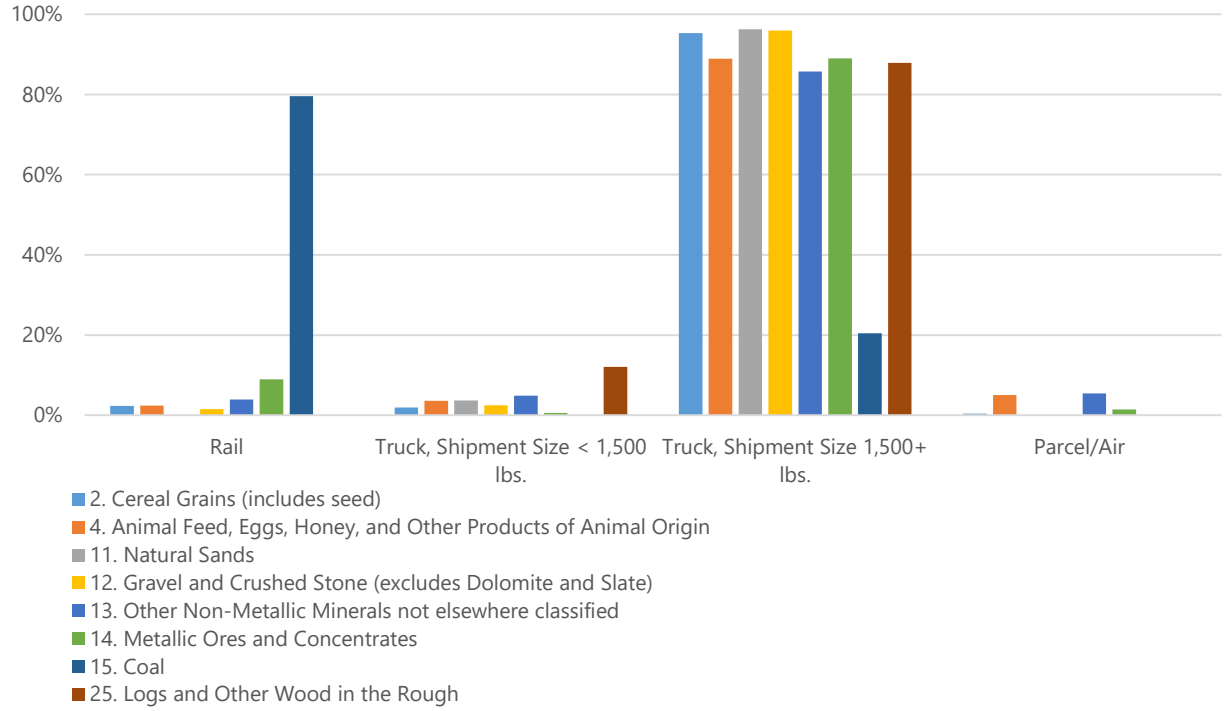


Figure 6-10 Mode and Shipment Size Share for Fuel and Fertilizer

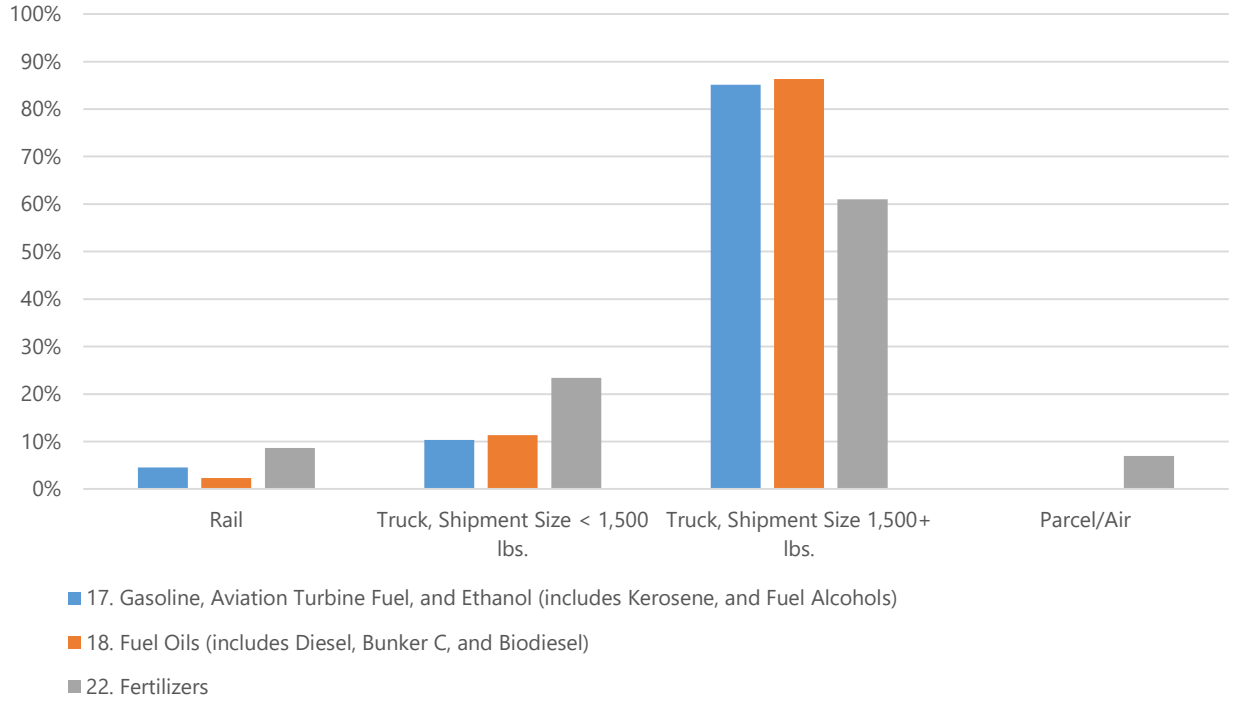


Figure 6-11 Mode and Shipment Size Share for Intermediate Products, Foodstuffs and Misc.

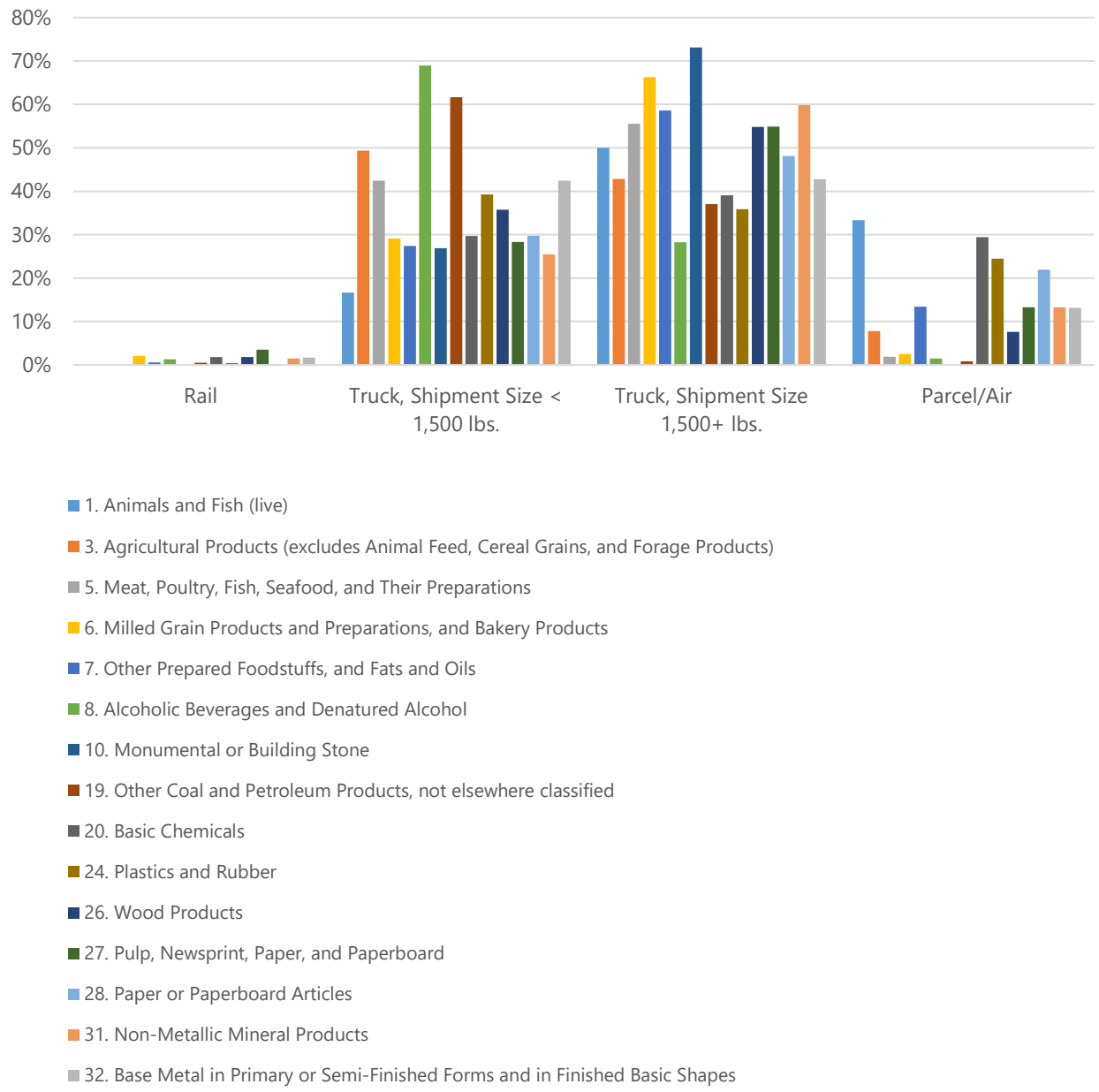


Figure 6-12 Mode and Shipment Size Share for Manufactured Goods

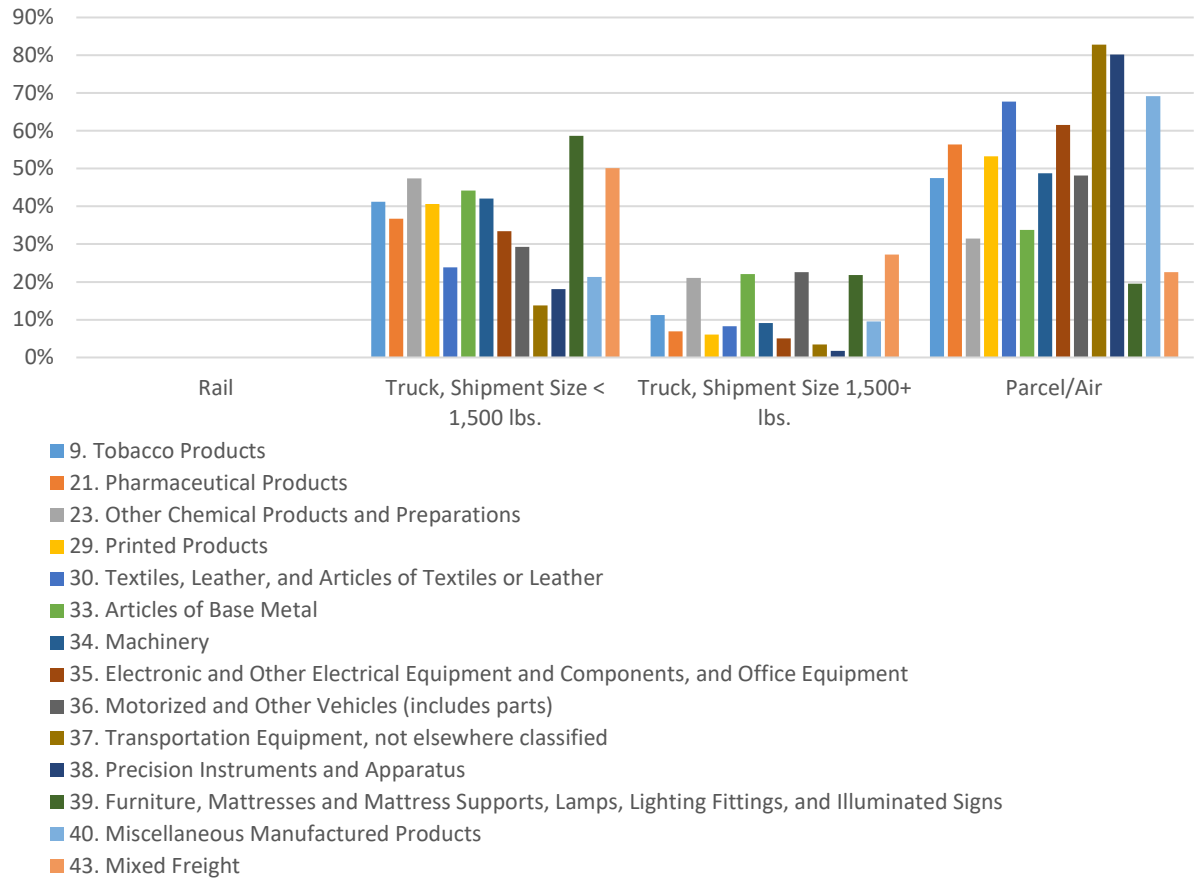


Table 6-20 shows the 10th percentile, median and 90th percentile Value per Pound and Great Circle Distance among shipments in each commodity category. Bulk Goods typically have the lowest value per pound followed by Fuels and Fertilizers, Intermediate Products, Foodstuffs and Misc., and Manufactured Goods. Manufactured Goods tend to have the longest shipment distances, which likely contributes to the high use of Parcel/Air. Many Bulk Goods and Fuels and Fertilizers shipments are relatively short-distance, which illustrates why truck is commonly used for transporting these goods.

Table 6-20 Value per Pound and Great Circle Distance

COMMODITY CATEGORY	N	VALUE PER POUND			GREAT CIRCLE DISTANCE (MILES)		
		10th Pctl	Median	90th Pctl	10th Pctl	Median	90th Pctl
BULK GOODS	4,346	$0.003	$0.010	$0.98	6	30	594
FUELS AND FERTILIZERS	1,712	$0.26	$0.48	$0.93	5	50	611
INTERMEDIATE PRODUCTS, FOODSTUFFS AND MISC.	33,878	$0.19	$1.17	$17.75	6	198	1,559
MANUFACTURED GOODS	60,765	$1.30	$14.26	$294.00	7	625	1,966
ALL GOODS	100,701	$0.32	$4.91	$144.00	6	367	1,855

6.3.3 Path Choice Model Estimation

The path choice model is a joint model where mode and shipment size decisions are evaluated simultaneously. Information on annual flows, including the origin, destination, and SCTG code of the shipment is obtained from the Supplier Selection model. The Path Choice model evaluates options for breaking down the annual flows into individual shipments that are generated regularly throughout a typical year. The CFS Microdata were used to develop nine mode and shipment size categories for the model (Table 6-21). In addition, major pipeline flows were simulated in the model using the observed flow patterns in FAF 4.1 in the calibration step.

Table 6-21 Mode and Shipment Size Categories Used in Path Choice model

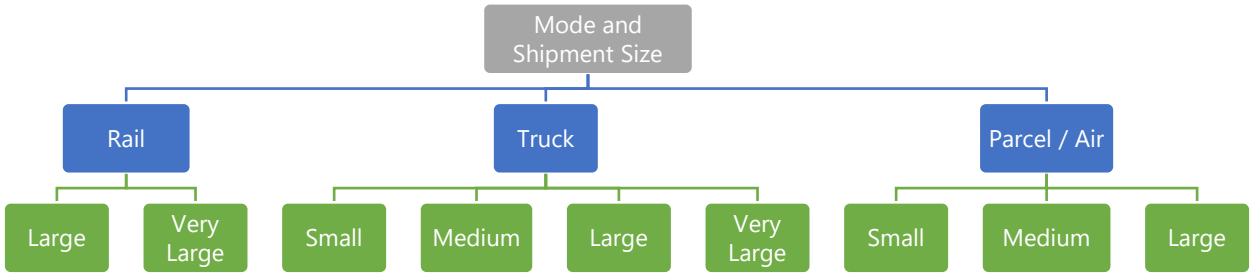
MODE CATEGORIES	SHIPMENT SIZE CATEGORIES			
	< 150 Lbs.	150-1,499 Lbs.	1,500-34,999 Lbs.	35,000+ Lbs.
RAIL CARLOAD/IMX	Not Included	Not Included	Modeled	Modeled
TRUCK	Modeled	Modeled	Modeled	Modeled
PARCEL/AIR	Modeled	Modeled	Modeled	Not Included

The recommended model is a Nested Logit model where mode and shipment size are modeled jointly. Numerous nesting structures were tested, with some grouping alternatives together based primarily on mode and others based more on shipment size. Nesting structures were assessed statistically [56] based on the estimated nesting coefficients (also called logsum parameters or dissimilarity parameters). Nesting coefficients that are above 1.0 or below zero indicate that the model is to be rejected because these values are inconsistent with the theoretical derivation of the model. Nesting coefficients of 1.0 means that the alternatives in the next are not correlated. Similarly, nesting coefficients that are relatively close to 1.0 indicate that the alternatives in the nest are weakly correlated. Relatively low nesting coefficients indicate that the alternatives are indeed highly correlated and that it is therefore logical for the alternatives to be grouped together in the same nest. Based on this rationale, the best nesting structures (those with relatively low nesting coefficients) were structures in which mode was used to group alternatives.

Figure 6-13 shows the best nesting structure that resulted from the analysis. In this structure, the four truck-based alternatives, the three Parcel/Air alternatives, and the two rail-based options are in their own mode-specific nests.

[56] An online reference of discrete choice models and their statistical properties can be found in Koppelman, F. and C. Bhat. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models (2006). Accessible online at http://www.caee.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf.

Figure 6-13 Nesting Structure of Path Choice Model [57]



Based on this structure, as explanatory variables such as travel costs change, shippers are more likely to select a different shipment size category within the same mode grouping and less likely to switch modes. The Theta values for these nests are between 0.3 and 0.4, which lends strong support to this nesting structure (see Table 6-22). Table 6-22 shows the summary statistics of the estimated model, in which 100,701 shipment observations were used. The model fit is reasonably good: Rho2 with respect to the constants-only model is 0.318, and Rho2 with respect to zero is 0.513.

Table 6-22 Path Choice model: Summary Statistics

Attribute	Value
Observations	100,701
Final Log (L)	-107854.1
D.O.F.	39
Rho²(0)	0.513
Rho²(C)	0.318
Theta (Truck Nest)	0.298 *(T-Statistic: 68.9)*
Theta (Parcel/Air Nest)	0.345 *(T-Statistic: 54.9)*
Theta (Rail Nest)	0.395 *(T-Statistic: 44.4)*

Table 6-23 shows the estimated model parameters. The parameter estimates correspond to the values of β in the following utility equation:

$$V_{in} = \beta_1 X_{in}$$

where

i = index of alternatives,

n = index of shipments,

V_{in} = the estimated utility of alternative i for shipment n,

β_i = a vector of estimated parameters corresponding to alternative i, and

$\mathbf{X}_{in}$ = a vector of independent variables corresponding to alternative i and shipment n.

[57] Small: <150 pounds; Medium: 150-1,499 pounds; Large: 1,500-34,999 pounds; Very Large: 35,000+ pounds.

The base category is Truck shipments that are less than 150 lbs. The parameter estimates in Table 6-23 display the signs and magnitudes that are expected. The parameter estimates can be interpreted as follows.

With other conditions remaining the same, medium-sized Truck shipping is the most preferred shipping option followed by small, large and very large Truck shipments. After truck, small Parcel/Air shipments present the next most preferred option followed by larger Parcel/Air shipments. Rail options are the least preferred options. Compared to other types of goods, shipments of Bulk Goods, Fuels and Fertilizers are relatively likely to use higher-capacity transport options including very large (35,000+ lbs.) rail and truck options and other truck options. As expected, small-shipment Truck and Parcel/Air are generally less likely to be used for shipping these goods. These results are reasonable since these types of goods are typically lower-value and therefore do not require faster, more reliable modes.

Manufactured goods, which tend to be higher-value and therefore more time-sensitive, are more likely to use Parcel/Air, small-shipment Truck and other Truck options. As expected, the estimated parameters suggest that these goods are very unlikely to be transported using Rail. Very large Truck shipments are preferred for goods that are destined for export to Mexico. Other Truck options are also utilized for these shipments. Small Parcel/Air is less preferable for these shipments. As discussed earlier, CFS Microdata sample sizes suggest that Rail and medium or large Parcel/Air shipping is infrequently used; however, due to limited sample size, parameters could not be estimated for these categories.

The small-shipment Truck option is preferred for exports that are destined to Canada followed by Medium Truck. Larger-shipment Truck and small Parcel/Air shipments are less preferable. Due to sample size limitations, Rail and larger Parcel/Air shipments appear to be least preferred (at least for the types of industries and commodities that are represented in the CFS). Parcel/Air shipping is highly preferred for export shipments that are destined neither to Mexico nor Canada. Small and medium Parcel/Air shipping is especially preferred for high-value goods. Truck is also used for these shipments, most likely for goods that are transported to seaports then shipped overseas to their final destinations.

Table 6-23 Path Choice Model Parameter Estimates [58]

ALTERNATIVE	VARIABLES & ESTIMATED PARAMETERS														
	ASC	Commodity Categories				Export				Transportation Cost (2012 $)			Time (Hours)		
								Other							
		Bulk Goods	Fuels And Fertilizers	Inter-Mediate Products, Foodstuffs And Misc.	Manufactured Goods	Mex	Can	All Goods	High Value	Low Value***	Medium Value	High Value	Low Value	Medium Value	High Value
RAIL-LARGE	-15.8*	N/A	N/A	0	-30.7*	N/A	N/A	N/A	N/A	-0.00020 *	-0.0057 *	-0.0063 *	0	-0.130 *	-0.170*
RAIL-VERY LARGE	-14.6*	8.47*	9.75*	0		N/A	N/A	N/A	N/A						
TRUCK-SMALL	0	0	0	0	0	0	0	0	0						
TRUCK-MEDIUM	0.287*	0	1.15*	0	-1.14*	0	-0.536*	0	0						
TRUCK -LARGE	-0.106*	2.63*	2.96*	0	-1.98*	0	-1.59**	0	0						
TRUCK-VERY LARGE	-0.777*	4.90*		0	-3.10*	1.10*		0.173**	0						
PARCEL/AIR-SMALL	-5.60*	0	0	0	3.47*	-2.73*	-1.62*	6.30*	2.63*						
PARCEL/AIR-MEDIUM	-11.1*	N/A	N/A	0	3.37*	N/A	N/A	10.6*	1.42*						
PARCEL/AIR-LARGE	-11.6*	N/A	N/A	0	-1.20**	N/A	N/A		0						

[58] *indicates 98% or higher statistical significance; ** indicates 75% or higher. *** Low-value is used for goods that are valued at less than $1/lb.; Medium for $1 to less than $4/lb.; and High for $4+/lb.

Cost and time parameter estimates were developed in a generic, non-alternative-specific fashion. Based on the estimated parameters, values of time can be calculated for low-, medium-, and high-value goods. The calculated values are:

- Low-value goods: Not time-sensitive
- Medium-value goods: $22.81/hour
- High-value goods: $26.98/hour

These results are consistent with recent research that was conducted by the Port Authority of New York/New Jersey [59] (see Table 6-24). In this study, stated preference information was collected from freight-related establishments that ship goods to or from the New York metropolitan area. This study indicates that the median value of time for these shipments is about $29/hour. Excluding the top three values, which likely represent small tonnages overall, the average value of time is about $32/hour. The results from the Path Choice model are reasonably close to these values, particularly since goods going to/from the New York region are probably more expensive than goods moving into or out of the Sun Corridor.

Table 6-24 Values-of-Time for Freight Shipments in a New York-Based Stated Preference Survey [59]

VALUE OF TIME (PER HOUR)	COST<$900			COST>=$900		
	Time>= 25 Hr	Time 12-25 Hr	Time<12 Hr	Time>=25 Hr	Time 12-25 Hr	Time<12 Hr
AGRICULTURAL	N/A	$7.69	$29.63	N/A	$15.09	$58.18
METAL AND MINING	N/A	$15.47	$33.68	$9.05	$35.00	$76.19
CONSTRUCTION	$5.23	$17.91	$37.21	$13.64	$46.67	$96.97
CHEMICAL	$7.07	$18.91	$34.78	$16.67	$44.62	$82.05
WOOD AND PAPER	$5.05	$15.05	$29.36	$9.82	$29.29	$57.14
ELECTRONIC	$7.68	$18.69	$32.32	$16.52	$40.22	$69.57
TRANSPORTATION & UTILITY	N/A	$19.17	$53.33	$8.57	$164.29	$457.14
WHOLESALE & RETAIL	$5.44	$21.47	$47.06	$24.67	$97.33	$213.33

6.3.4 Model Application and Preliminary Results

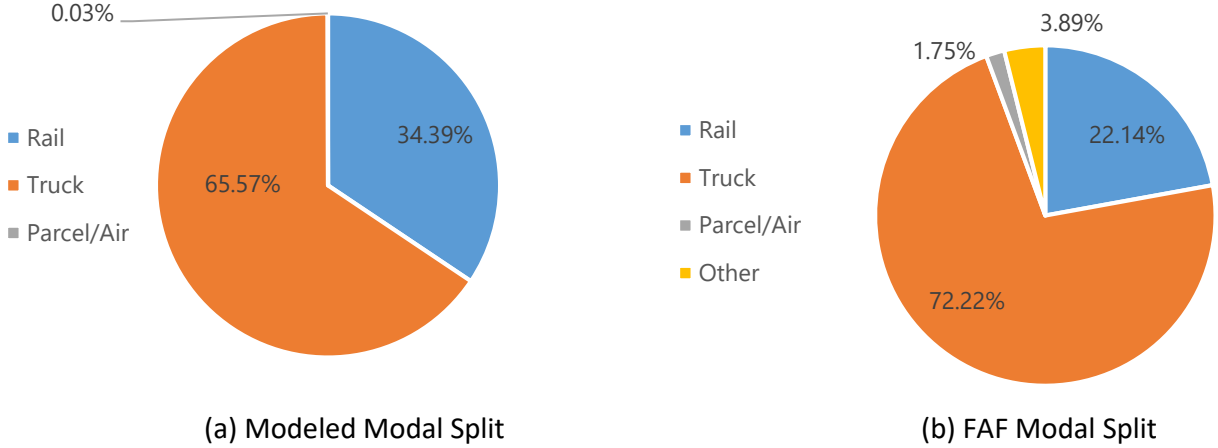
The primary objective of this model was to develop a joint mode choice and shipment size model for freight transportation, focusing on commodity flow in Phoenix-Tucson metro area. The developed model is implemented into MAG's behavior-based supply chain model and employed to simulate commodity flows at the very disaggregate level of firm-to-firm. The preliminary results showed that the model simulates 95% of commodity flows (tonnage) presented in FAF data at disaggregate level between firms. Forty two classes of commodity types, based on the SCTG code, are considered in this simulation.

The presented mode and shipment size choice model in this chapter is then used to replicate logistics choices for individual shipments between firms. Results of simulation are used to validate proposed models. The relative percentages of total tonnage of commodities transported by each of four

[59] Source: Komanduri, A., S. Musti, and K. Proussaloglou. Modeling Values of Time to Support Freight Decision-Making: Results from a Stated Preference Survey in New York. Submitted to the Transportation Research Board (TRB) Model Applications Conference on November 16, 2011

considered modes are compared with the relative values obtained from publicly available FAF data for the MAG/PAG region sample. Figure 6-14 compares the modal split according to FAF3.5 and the model results.

Figure 6-14 Modal Split in Preliminary Results: Modeled VS. FAF3.5



It should be noted that at the time that the model was estimated and preliminary results were obtained, the latest version of FAF data for the year 2015 was provided by FAF 3.5. Later in the calibration step, the FAF 3.5 data was replaced with the latest version (FAF4.1) for validating the model results. The next section explains the model updates and calibration process to improve model accuracy of both supplier selection and path choice model.

6.3.4.1 Data Issues

Similar to any behavior-based logistics choice model, development and application of this model required detailed and disaggregate data. While CFS-Microdata was used as the primary data source for estimation of the model, several other data sources were collected and used for the development and application of the model including the transportation cost and time information by mode, great circle and routed distance matrix between all zones by mode, and external stations used for external shipments to/from outside regions for truck mode. The model sheds light on the mode choice and shipment size selection behaviors for freight flows. However, due to the lack of data the shipping chain choice cannot be determined for truck shipments. The shipping chain choice refer to the number of intermediate stops within the supply chain and distinguishes direct supply chain from indirect chains. This choice model can be developed once the data on the intermediate logistics nodes (consolidation centers, distribution centers, etc.) is available.

6.4 Calibration Process and Application Results

This section presents the calibration process performed on the model and some of the application results of the MAG/PAG SHRP2-C20 agent-based supply chain and freight transportation model.

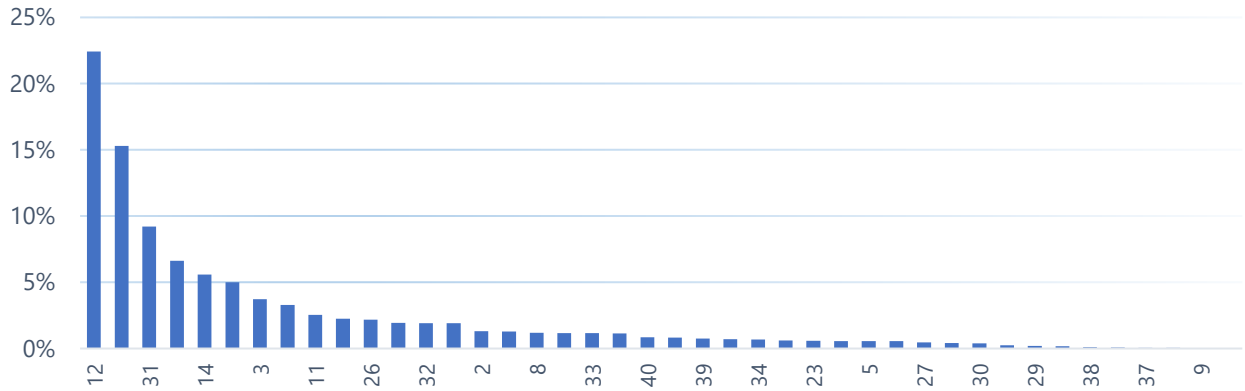
6.4.1 Overview and Data Sources

The calibration of the model included calibration of the supplier selection model and calibration of the mode choice model. FAF4.1 data is the main data source for calibration of both models. FAF4.1 commodity flow patterns to, from, and within MAG/PAG region were used for calibration of the supplier selection model and the modal split in the FAF 4.1 data was used for calibration of the mode

choice model. The regional TRANSEARCH data sample was also considered and analyzed for the model calibration. However, it was not selected for the final calibration due to several inconsistencies that were found between TRANSEARCH data and FAF data.

According to FAF4.1 2015 estimates, 186,423 kilotons of commodities was transported to, from, and within MAG/PAG region which valued more than 306,288 million dollars. Figure 6-15 shows percentage of commodity flows by type (SCTG code) for the region.

Figure 6-15 MAG/PAG Commodity Flow by SCTG Type Based on FAF4.1



Based on the FAF4.1 data, truck is the dominant mode of transportation in the region. Apart from truck, the other modes of shipment are rail, air, pipeline and parcel. Table 6-25 presents commodity flows by mode and by direction based on FAF4.1 data. As it was mentioned in Section 6.3, only truck, rail, air, pipeline and parcel shipments are considered and modeled in this project. Due to lack of data, unknown modes are excluded from the study.

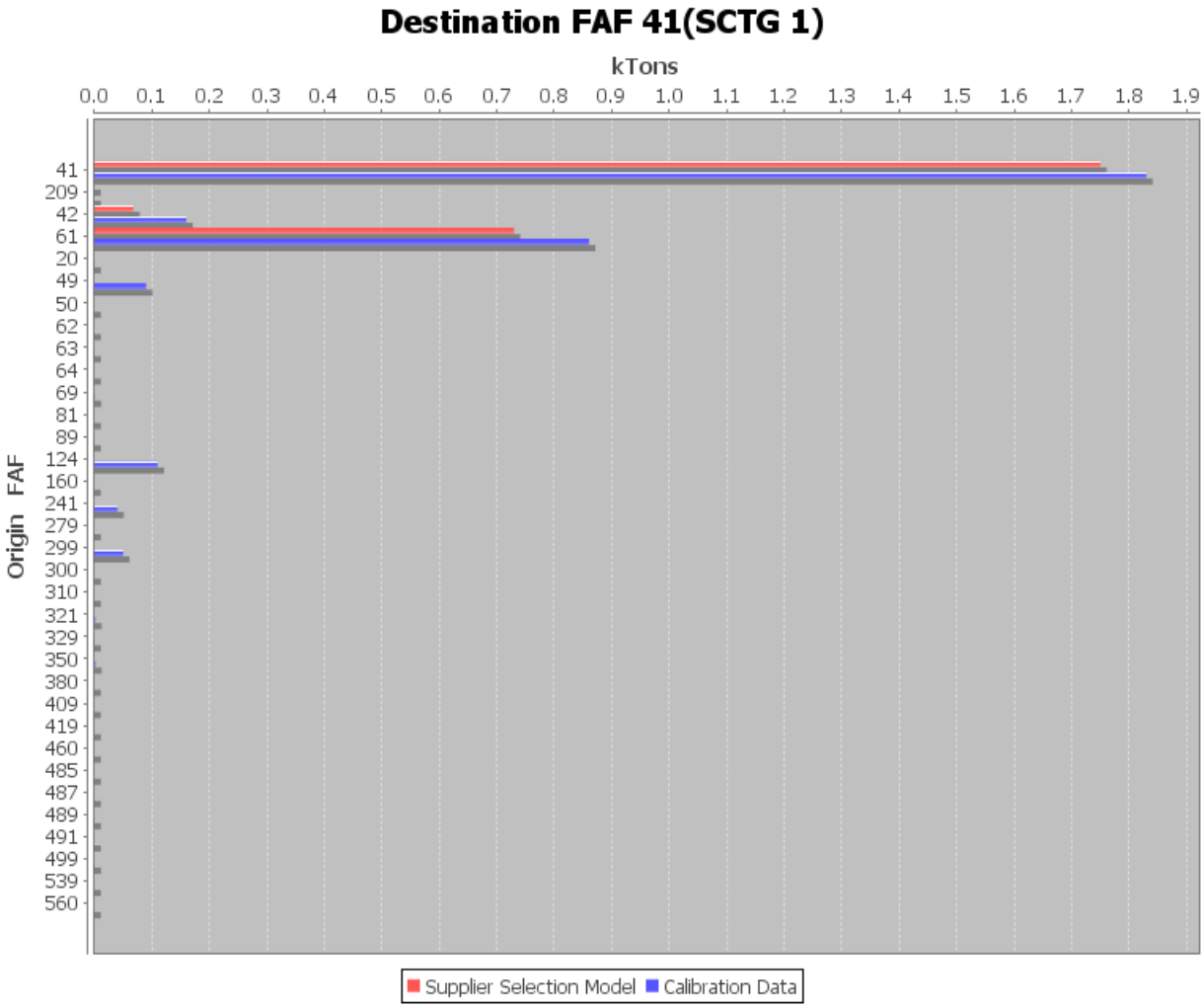
Table 6-25 MAG/PAG Modal Split by Kiloton in FAF4.1

DIRECTION	TRUCK	RAIL	AIR	PARCEL	PIPELINE	UNKNOWN	TOTAL
II (INTERNAL)	99,506	5,165	2	128	9	27	104,836
EI (INBOUND)	25,666	7,658	68	2,374	24,261	12	60,038
IE (OUTBOUND)	16,881	2,148	69	574	1,861	14	21,548
TOTAL	142,052	14,971	139	3,076	26,132	53	186,423

6.4.2 Calibration Process and Application Results

The FAF4.1 data was used to calibrate commodity flow patterns to, from, and within MAG/PAG region. The firm-level supplier selection model outputs were aggregated to FAF zone level for the calibration purpose. OD Flows for each commodity were plotted and compared against the FAF flows (see Figure 6-16) and for those markets that showed difference, the Supplier Selection Model and Market Clearing Algorithm was tuned to provide a better calibration. For the SSM, the calibration was focused on tuning the formation of potential suppliers and buyers sets in order to increase the probability of replicating similar flows as observed in FAF data. The comparison and calibration was performed for all SCTG groups.

Figure 6-16 Modeled vs. Observed Inbound Flows to MAG Region for SCTG 1 Before Calibration



In total, the supply chain model simulates 184,944 Kilo tons of commodities for the MAG/PAG region which is 99.2% of total observed flows in FAF4.1. Figure 6-17 compares total modeled and observed commodity flows by SCTG type. As it can be seen in the figure, total modeled commodity flows is very close to the observed flows in FAF.

Figure 6-17 Modeled vs. Observed Inbound Flows by SCTG After Calibration

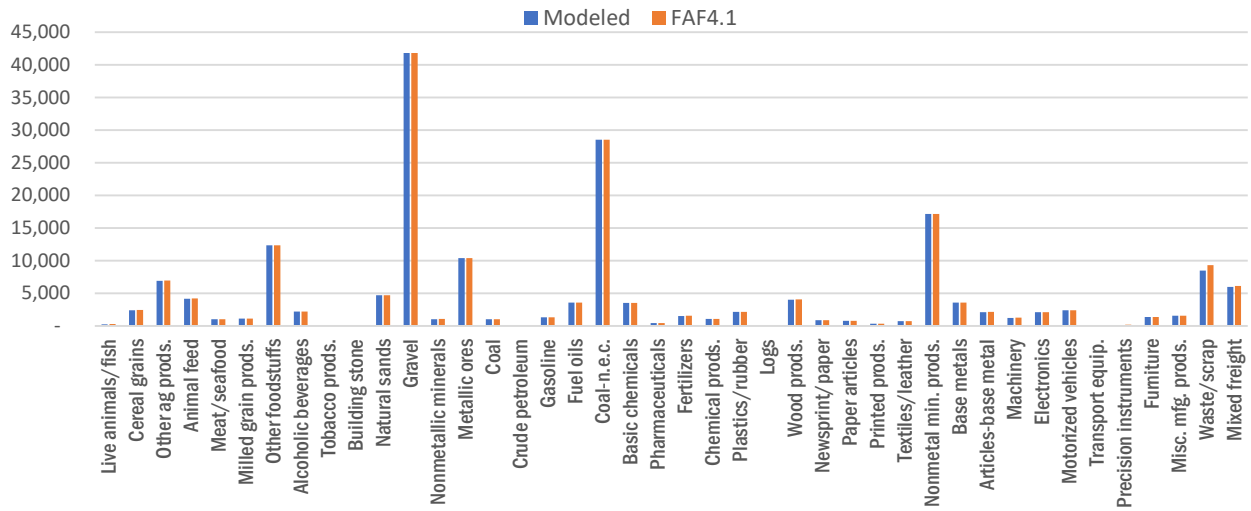


Table 6-26 compares the modeled directional flows with FAF4.1. As the table shows, total directional commodity flows show very close match with observed flows in FAF.

Table 6-26 Modeled vs. Observed Total Directional Flows

(a) Modeled Flows

OD	MAG	PAG	Total
MAG	68,614	9,458	78,072
PAG	4,078	21,795	25,873
Total	**72,692**	**31,253**	**103,945**

Direction	Flow (kton)	% of Total
II (Internal)	103,945	56%
EI (Inbound)	59,916	32%
IE (outbound)	21,083	12%
Total	**184,944**	

(b) FAF Flows

OD	MAG	PAG	Total
MAG	69,270	9,507	78,777
PAG	4,199	21,860	26,059
Total	**73,469**	**31,367**	**104,836**

Direction	Flow (kton)	% of Total
II (Internal)	104,836	56%
EI (Inbound)	60,038	32%
IE (outbound)	21,548	12%
Total	**186,423**	

Table 6-27 shows modeled inbound (External) commodity flows to MAG/PAG region for the top 10 origins by weight. It also compares the modeled flows to the observed inbound flows in FAF data. The same comparison was done for the outbound (external) commodity flows which is presented in Table 6-28. The close match to the FAF data shows the level of precision in supplier selection model and market clearing algorithm at the aggregate level.

Table 6-27 Modeled vs. Observed Inbound Flows to MAG/PAG Region

Modeled			Observed		
Origin FAF Zone	Description	Inbound Kton	Origin FAF Zone	Description	Inbound Kton
350	NM	23,345*	350	NM	23,348*
61	CA Los A	8,262	61	CA Los A	8,269
49	AZ rem	6,861	49	AZ rem	6,870
69	CA rem	1,625	69	CA rem	1,627
300	MT	1,324	300	MT	1,325
64	CA San J	1,238	64	CA San J	1,243
190	IA	1,142	190	IA	1,142
321	NV Las V	896	321	NV Las V	897
310	NE	693	310	NE	695
489	TX rem	648	489	TX rem	648

* This number shows total flow by all modes including pipeline.

Table 6-28 Modeled vs. Observed Outbound Flows from MAG/PAG Region

Modeled			Observed		
Destination FAF Zone	Description	Outbound Kton	Destination FAF Zone	Description	Outbound Kton
49	AZ rem	5,465	49	AZ rem	5,536
61	CA Los A	3,251	61	CA Los A	3,272
69	CA rem	2,255	69	CA rem	2,195
321	NV Las V	1,159	321	NV Las V	1,162
350	NM	933	350	NM	946
64	CA San J	879	64	CA San J	894
171	IL Chica	655	171	IL Chica	660
261	MI Detro	599	261	MI Detro	611
485	TX El Pa	465	485	TX El Pa	472
484	TX Dalla	382	484	TX Dalla	385

The FAF4.1 data was also used for calibration of the mode choice model. The model parameters were tuned to replicate the similar modal split on aggregate level (FAF zone). Table 6-29 compares modeled and observed modal splits by weight.

Table 6-29 Modeled vs. Observed Modal Split

Modeled			Observed		
Alternative	Kton	Modal split	Alternative	Kton	Modal split
Parcel-Air	2,536	1.4%	Parcel-Air	3,215	1.7%
Rail	16,772	9.1%	Rail	14,971	8.0%
Truck	140,792	76.1%	Truck	142,052	76.2%
Pipeline	24,844	13.4%	Pipeline	26,132	14.0%
Total	184,944	100%	Total	186,370	100%

6.4.3 Model Procedures

This section describes the steps of running supply chain model. See Chapter 9 for more detail information.

Step 1. Freight Generation & Data Preparation

This step takes firm synthesizer output (list of firms), classify firms into agent clusters, determine input and output commodities for each agent and calculate production and consumption amounts (annual kilo-ton). The process is described in the following sub-steps.

Step 1.1. Read, Enumerate and Summarize Firm Data

This step reads the external firm (outside MAG/PAG region) list and firm synthesizer output, classifies firms based on location (MAG TAZ), industry class (NAICS 6-digit) and employee size classes (1 to 8), reformats the firms (agents) list to be used in the next steps, and generates a final firm list.

Step 1.2. Read and Process Data for Calculation of Make-Use Rates

This step reads required data for calculation of commodity production and consumption (PC) rates (per employee) and process/reformat the data to prepare for calculation in the next steps. The model estimates production and consumptions rates per employee for each industry class (by NAICS 6-digit). The calculation of PC rates per employee is based on the latest (2007) benchmark industry input-output (IO) accounts provided by Bureau of Economic Analysis (BEA) at national level and employment information for the same year at national scale. A commodity to industry correspondence table is developed by consultant and used for linking industries (6-digit NAICS codes) to the relevant commodities (SCTG code). Since BEA data is based on monetary values, commodity values per pound are calculated from FAF3 data at national level and used for converting monetary values to pounds. The BEA IO accounts uses a different industry classification than the NAICS system. Thus, a crosswalk is developed and used to convert NAICS industry codes to IO industry codes. The process includes following steps:

1) Read NAICS to SCTG crosswalk

2) Read IO to SCTG crosswalk

3) Read SCTG unite values
4) Read NAICS to IO industry crosswalk
5) Merge firm data (output from previous step) with NAICS to IO crosswalk to determine IO industry class for each agent on the firm list
6) Read employment data (CBP) for the same year as BEA IO tables
7) Merge employment data with NAICS to IO crosswalk to determine IO industry code for each NAICS employment record. Sum over IO code to aggregate the employment table.
8) Generate intermediate employment information file
9) Read IO make table, convert table to linear format and write intermediate IO make file.
10) Read IO use table, convert table to linear format and write intermediate IO use file.

Step 1.3. Calculate Commodity Make Rates and Develop Supplier Agents List

This step uses intermediate output file "io_make_2007.csv" from previous step and calculates commodity production rates per employee by industry class. It also reads in a regional make and use rate table, obtained from IMPLAN regional data, to replace the national based rates by regional rates for firms in the MGA/PAG region. Almost all 6-digit NAICS industries are assumed to produce one type of commodity (SCTG) except for some industry classes such as wholesales that produce two or more commodities. For these especial industries a simulation process is used to simulate produced commodity using probability values obtained from 2012 CFS microdata. These industries, produced commodities and their shares are shown in the following table. Since 2012 CFS micro-data provides 4-digit NAICS code for Wholesale trades, the model uses only first four digits of NAICS codes in the firm data to simulate produced commodity.

The procedure includes several steps that are described below:

1) Uses the intermediate **io_make_2007** data file and reformats the data file. As a results the **iomaketable** is generated which shows total annual produced commodities ($million) by each IO industry.
2) The **iomaketable** is merged with the total national employment data, processed in previous step and the intermediate **iosupplier** file is generated.
3) Writes intermediate check file
4) Production rates (**mfactors**) are calculated by dividing total annual produced value by total employee data.
5) The model identifies produced commodity types by each agent in the firm data by merging the crosswalk between NAICS and SCTG codes with firm list and generating **suppliers** data.
6) Run the simulation process to determine type of commodity produced by especial industries (2 or more commodity).

7) Merge the **supplier** firm list with commodity unit value file using SCTG code

8) Merge the **supplier** firm list with **iosupplier** list that includes production rates using common IO industry code.

9) Calculate production amount (kton) for each firm in the list.

10) Replace the national-based rates with regional rates **regional_make_rate** for firms in MAG/PAG region and recalculate production amounts (kton)

11) Clean the list of supplier by removing redundant attributes and generating final **SupplierFirms**.

Step 1.4. Calculate Commodity Use Rates and Develop Buyer Agents List

This step uses intermediate output file "io_use_2007.csv" from previous step and calculates commodity consumption rates per employee by industry class. It also replaces national-based use rates by regional rates for regional firms. One step in this process is to convert IO industry types used by BEA to SCTG codes. Almost all 6-digit IO industries are connected to one type of commodity (SCTG) except for some industry classes that are related to two or more commodities. For these especial industries a similar simulation process is used to simulate consumed commodity using employment information from CBP. These IO industries, related SCTG codes and their shares are presented in the following table.

The procedure includes several steps that are described below:

1) Uses the intermediate **io_use_2007** data file and reformats the data file. As a results the **iousetable** is generated which shows total annual consumed commodities ($million) by each IO industry.

2) The **iousetable** is merged with the IO to SCTG crosswalk, processed in previous step and an intermediate iobuyers file is generated. This file includes list of buyer industries and the consumed SCTG.

3) Writes intermediate check file

4) Run the simulation process to determine type of commodity consumed by especial industries (2 or more commodity).

5) Merge the **iobuyers** data with the total national employment data, processed in previous step for calculating rates per employee.

6) Calculate consumption rates (**ufactors**) by dividing total annual consumption value by total employee data.

7) Merge the firm list (**firmdata**) with **iobuyers** list that includes consumption rates using common IO industry code and generates **buyer** firm list.

8) Adjust Consumed commodities by wholesale trade industry in **buyers** list using the same probabilities from CFS micro data.

9) Calculate consumption amount (kton) for each firm in the list.

10) Replace the national-based rates with regional rates **regional_use_rate** for firms in MAG/PAG region and recalculate consumption amounts (kton)

11) Clean the list of supplier by removing redundant attributes and generating final **BuyerFirms**.

Step 1.5. Adjustment of Supply and Demand & Print Supplier-Buyer Firm Lists

This step uses FAF3.4 2015 commodity production and consumption data to calibrate the firm level production and consumption estimates from previous step. The FAF zone level production and consumption are used as marginal values and a straightforward proportional fitting method is used to adjust firm level values in a way that sum of firm level values matches the marginal FAF values. The adjustment process is described below:

1) Aggregate firm-level productions from **SupplierFirms** data table to FAF zone level.

2) Read 2015 FAF production data.

3) Calculate zone level proportional fitting factors by dividing observed zonal production by zonal estimates from **SupplierFirms**.

4) Apply fitting factor to firm level estimates.

5) Clean and finalize supplier firms list "**SupplierFirms**"

Repeat the same steps for BuyerFirms list.

Step 2. Supplier Selection Model

This step uses supplier and buyer firm data as input and run the supplier selection model to allocate commodities between supplier-buyer pairs.

Step 2.1. Calling JAVA application and running market clearing algorithm & Run the Supplier Selection Model

This step of R code includes only the commands to call SSM java code and run the supplier selection model.

Step 2.2. Read the SSM output and Prepare for Mode-Shipment Size Model

This step takes supplier selection outputs and clean the firm level commodity flow data by removing redundant attributes. It prepares the input data for running the path choice model.

Important Note: Currently, no data is available on employment and business establishments in the main foreign destination countries. Modeling these export flows is not doable with available data. Therefore, a placeholder is used in this step for future import/export flow estimates.

Step 3. Transport, Mode and Path Choice Model

This step uses modeled annual commodity flows from supplier selection model and determine mode of transportation and shipment size for individual firm-level flows. A nested logit model is applied to jointly model mode and shipment size choice.

Step 3.1. Read Required Input Data, Process and Prepare Data for Mode-Shipment Size Model

Required input data for application of mode-shipment size choice model are read in this step. In addition, the commodity flow data is processed and all variables for model application are created. The required variables include:

- Commodity value dummy variable
- Commodity type dummy variable
- Export dummy variable
- Distance
- Transportation cost
- Transportation time

The data preparation process is described below:

1) Read regional commodity unite values ($/lbs) obtained from CFS microdata.
2) Add the unite value variable to the modeled annual firmflowdata and create dummy variable for low, medium and high value commodities based on regional commodity prices. The following table presents defined price levels for classification of commodities.

Table 6-30 Commodity Classification Based on Unit Values

Commodity Category	Value ($/lbs)
Low value	Less than 1.0
Medium value	1 – 3.99
High value	4 or greater than 4

3) Reclassify commodities based on type and create dummy variables based on the following table.

Table 6-31 Commodity Classification Based on Type of Commodity

Commodity Category	SCTG codes included
Manufactured	9, 21, 23, 29, 30, 33, 34, 35, 36, 37, 38, 39, 40, 43
Intermediate	1, 3, 5, 6, 7, 8, 10, 19, 20, 24, 26, 27, 28, 31, 32
Bulk	2, 4, 11, 12, 13, 14, 15, 25
Fuel & Fertilizers	17, 18, 22

4) Generate dummy variable for exporting firms. As noted before, this dummy variable is just a placeholder. Currently no export flow is simulated in the model.
5) Read Great Circle Distance (GCD) matrix and add GCD to each flow record in firmflowdata.
6) Set the minimum GDC for intra-zonal flows to 5 miles.
7) Read in truck routed distance and add the info to each flow record.
8) Read in rail routed distance and add the info to firmflowdata records.
9) Create a dummy variable for rail availability.

All the routed distance information are obtained from the network skims that MAG staff provided.

10) Calculate transportation cost by mode for each OD flow record in firmflowdata. Cost is estimated using the following assumptions.

Table 6-32 Transportation Costs Calculation Assumptions by Mode

Alternative(s)	Distance (Source: Network Skims and Zone system GIS shapefiles)	Cost per Ton-Mile	Minimum Cost ($)
Rail	Routed Distance	$0.0395	200
Truck	Great Circle Distance	$0.375	10
Parcel or Air	Great Circle Distance	$2.49	15

11) Calculate transportation time by mode for each OD flow record in firmflowdata. Transportation time is estimated using the network skims and following assumptions.

Table 6-33 Transportation Time Assumptions by Mode

Geography	Mode	Distance (Source: Network Skims and Zone system GIS shapefiles)	Travel Time (hour)	Speed (MPH)
All Trips	Rail	Routed Distance	12 hours + Distance/Speed	22
Intra-Sun Corridor	Truck	Routed Distance	4 hours + Distance/Speed	20
	Parcel or Air	Routed Distance	1 hour + Distance/Speed	40
External	Truck, 0-650 miles	Routed Distance	16 hours + Distance/Speed	65
	Truck, 651-1,299 miles	Routed Distance	16 hours + Distance/Speed	38
	Truck, 1,300+ miles	Routed Distance	16 hours + Distance/Speed	32
	Parcel or Air	Great Circle Distance	12 hours + Distance/Speed	250

Step 3.2. Determine Mode Choice and Shipment Size for Annual Flows

This step calculates utilities by mode for each OD firm flow record, estimate choice probabilities for each alternative and determine amount of commodities transported by each alternative based on calculated probabilities. The process is performed in as follows.

1) Calculate utilities for each alternative. The model estimation results presented are used for calculation of utilities.
2) Calculate choice probabilities for each alternative.
3) Fix out of range (very low and very high) utilities that result in zero or infinite probabilities.
4) Calculate selection propensity for each alternative
5) Calculate modal-shipment Size splits
6) Calculate total simulated flows
7) Adjust predicted flows for unavailable alternatives
8) Calculate total simulated flows by each alternative
9) Determine Rail yard used for rail shipments based on rail network skim
10) Add external highway station for external shipments

Step 3.3. Process and Print Out Final Results Annual Flows

This step print out output files and produces summary reports.

- Input files: Updated firmflowdata with selected mode and size alternatives from previous step.
- Output files: CommodityFlows_firmlevel.csv contains all simulated records with the selected mode and size alternatives at firm level.

6.5 Additional References:

Pourabdollahi, Z., Karimi, B., Mohammadian, A., and Kawamura, K. A Hybrid Agent-Based Computational Economics and Optimization Approach for Supplier Selection Problem, International Journal of Transportation Science and Technology, 6, pp. 344-355, 2017.

Pourabdollahi, Z., Karimi, B., Mohammadian, A., and Kawamura, K. Shipping Chain Choices in Long Distance Supply Chains: Descriptive Analysis and a Decision Tree Model, Transportation Research Records: Journal of the Transportation Research Board, Washington D.C., 2410, pp. 58-66, 2014.

Pourabdollahi, Z., Karimi, B., Mohammadian, A., and Kawamura, K.An Agent-based Supply Chain and Freight Transportation Model: Case Study for Chicago Metropolitan Area, Contemporary Issues in Supply Chain Management and Logistics, Business Expert Press, 2017.

Pourabdollahi, Z., Karimi, B., Mohammadian, A., and Kawamura, K. An Agent-Based Supply Chain and Freight Transportation Model: Case Study for Chicago Metropolitan Area, 95th Annual Transportation Research Board Meeting, Washington D.C., January 10-14, 2016.

Pourabdollahi, Z., Mohammadian, A., and Kawamura, K. A Behavioral Freight Transportation Modeling System: An Operational and Proposed Framework, 14th Annual International Conference on Electronic Commerce, ACM, August 2012, 196-203.

Pourabdollahi, Z., Mohammadian, A., and Kawamura, K., A. A Behavioral Freight Transportation Modeling System, 4th Transportation Research Board Conference on Innovations in Travel Modeling (ITM), Tampa, Florida, April 29-May 2, 2012.

Pourabdollahi, Z., and Moahhamadian, A. A Conceptual Framework for a Behavioral Freight Transportation Modeling System with Logistics Choices, 4th METRANS National Urban Freight Conference, Long Beach, California, October 11-14, 2011.

Pourabdollahi, Z., and Moahhamadian, A. An Enhanced Freight Activity Micro-simulation Estimator (FAME) Model using Freight Analysis Framework (FAF3.0) dataset, The Mid-Continent Transportation Research Symposium, Ames, Iowa, August 18-19, 2011.

Samimi, A., A. Mohammadian, and K. Kawamura. "A Behavioral Freight Movement Microsimulation Model: Method and Data", Journal of Transportation Letters: The International Journal of Transportation Research. 2(1):53-62, 2010.

Samimi, A., A. Mohammadian, and K. Kawamura. "Behavioral Freight Movement Modeling", in Bhat, C.R. and R.M. Pendyala, Travel Behaviour Research in an Evolving World. Chapter 8, 22 pages, Lulu Publishers, USA, 2012.

Samimi, A., A. Mohammadian, K. Kawamura, and Z. Pourabdollahi, "An Activity-based Freight Mode Choice Microsimulation Model", in Transportation Letters, 6(3), pp. 142-151, 2014.

Samimi, A., K. Kawamura, and A. Mohammadian, "A Behavioral Analysis of Freight Mode Choice Decisions", in Journal of Transportation Planning and Technology, Vol. 34 (8), pp. 857-869, 2011.

Samimi A., Z. Pourabdollahi, A. Mohammadian, and K. Kawamura. An Activity-based Freight Mode Choice Microsimulation Model, in proceedings of: 91th Annual Transportation Research Board Meeting, Washington D.C., January, 2012.

Samimi, A., A. Mohammadian, and K. Kawamura, "Freight Demand Microsimulation in the US", Proceedings of the 12th World Conference on Transport Research (WCTR), Lisbon, Portugal, July 11-15, 2010.

Samimi, A., A. Mohammadian, K. Kawamura, "Behavioral Freight Movement Modeling". Proceedings of the 2009 IEEE International Conference on Service Operations, Logistics and Informatics, July 22-24, Chicago, IL.

Samimi, A., A. Mohammadian, K. Kawamura, "Freight Movement Modeling: Methodology and Data Needs", TRANSLOG 2009 (Transportation and Logistics) 1st Annual International Conference June 17th-18th, 2009, McMaster University, Hamilton, ON, Canada

Samimi, A., A. Mohammadian, K. Kawamura, "A Behavioral National Freight Model: Data and Method", proceedings of the 4th International Transport Economics Conference, June 15-16, 2009, University of Minnesota

Samimi, A., A. Mohammadian, K. Kawamura, "A National-Level Micro-Simulation for Goods Movement in the United States: Data and Method", TRB's Data and Tools for Linking Goods Movement, Air Quality, and Transportation Infrastructure Decisions Workshop, June 2-3, 2009, Irvine, California

7 Tour-Based Truck Model

Consultant Technical Lead –Jason Lemp, Brent Selby and Arun Kuppam (Cambridge Systematics)

MAG Technical Lead –Lavanya Vallabhaneni, Shuyao Hong, Kyunghwi Jeon, Daehyun You, and Haidong Zhu

The objective of truck tour model is to develop truck trip chains by industry sector by truck type. These truck trip chains are then grouped into the major linkages based on land uses the trucks make stops at and the probability of making another stop based on the number of previous stops. The tour-based model generates the number of stops by industry sector, number of stops on a tour, stop purposes, and the location and time of day of stops. This chapter describes the processing of truck GPS data from two different sources, and developing tour-based models by truck type using the processed GPS data.

7.1 ATRI Truck GPS Data

7.1.1 2011 ATRI Data

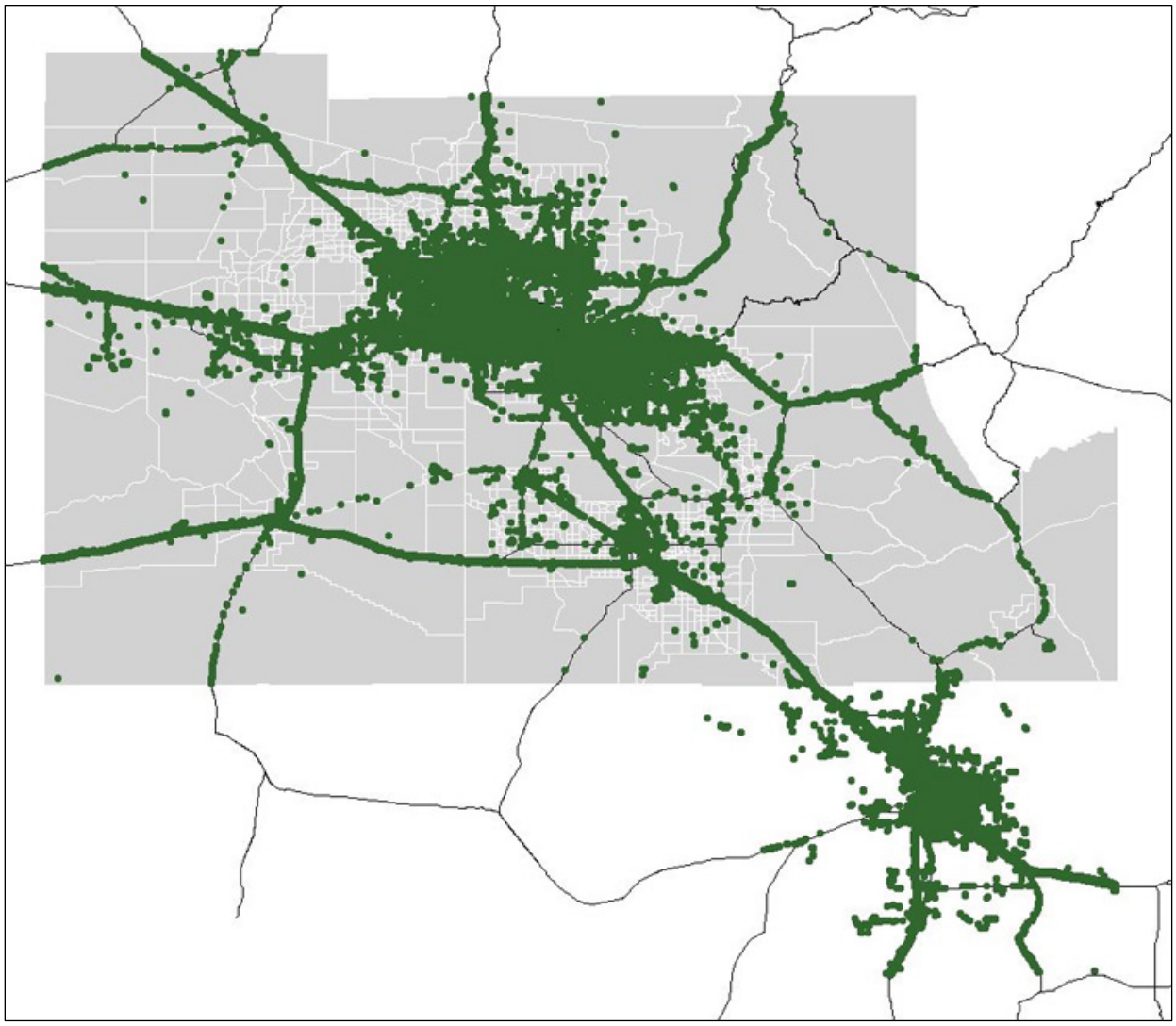
The 2011 ATRI truck GPS data was obtained in 2012 to update MAG's 2010 truck trip-based model and to develop a new tour-based truck model for heavy trucks. The ATRI data was purchased by MAG for the period from April 1, 2011 to April 30, 2011. The data delivery contained 3,429,603 GPS event records from 22,657 unique trucks that indulged in 58,637 tours, as shown in Table 7-1. The locations of these GPS records are shown in Figure 7-1. At these GPS events, the vehicle may be stopped or moving. In principle, only certain stopped records can be grouped into tours or trips, but tours or trips cannot be precisely computed without further processing.

Several criteria were employed to process the GPS event records to determine if the events where the truck is reported as starting (or entering) and stopping (or exiting) define truck trip ends, and/or which will be called a transition event. Between these transition events, the vehicle may be either moving or stopped. The moving records add no information about the trip ends. The stopped records add no additional information about the trip end that already has been determined. Only the transition events are useful in determining the trip ends. Both stopped and moving GPS records were eventually deleted from the GPS database without any impacts on calculating truck trips and tours. Each truck generated about 151 GPS events and the number of tours per truck is about 2.6 (derived from Table 7-1).

Table 7-1 2011 ATRI Truck GPS Data

All Trucks	April 2011
GPS Events	3,429,603
Truck Tours	58,637
Trucks	22,657

Figure 7-1 ATRI GPS Truck Events in April 2011

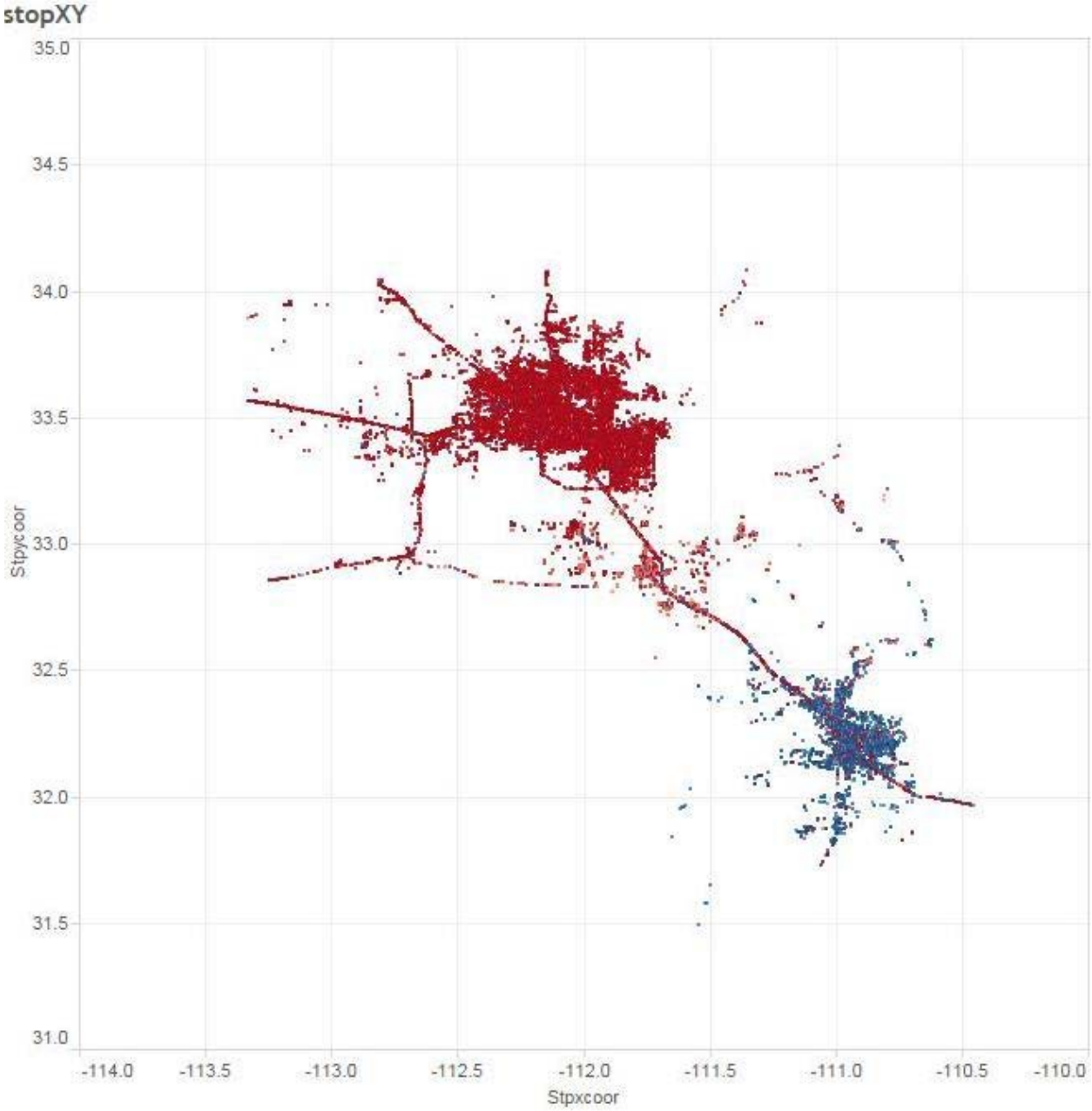


7.1.2 2014 ATRI Data

The 2014 ATRI truck GPS data was obtained in 2015 as part of the SHRP2 C20 Freight Model project. In order to get a good mix of data across different seasons, MAG obtained samples from four different months – January, April, July, and October 2014. The total number of GPS events amounted to about 11.4 million that included about 1.96 million in January (10 days), 3.85 million in April (17 days), 1.92 million in July (10 days), and 3.69 million in October 2014 (17 days). These GPS events are from 39,080 trucks that indulged in 81,090 tours. A combination of ATRI staff and MAG staff processed these data and produced the final ATRI truck trip table for CS' analyses.

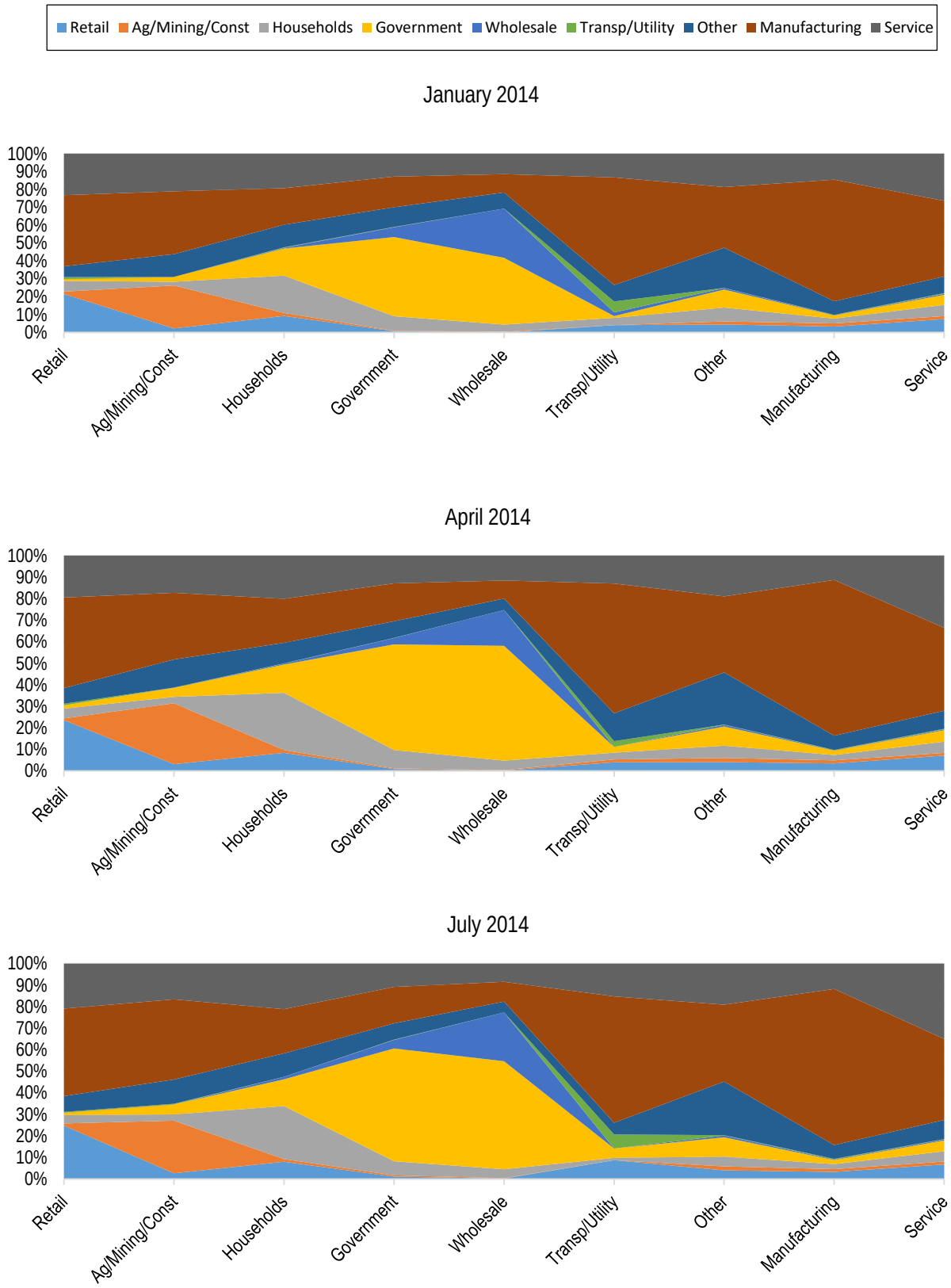
As the SHRP2 C20 project encompasses the MAG-PAG megaregion, MAG obtained ATRI data for the combined MAG and PAG modeling regions. Table 7-2 presents the distribution of truck GPS events across the two regions.

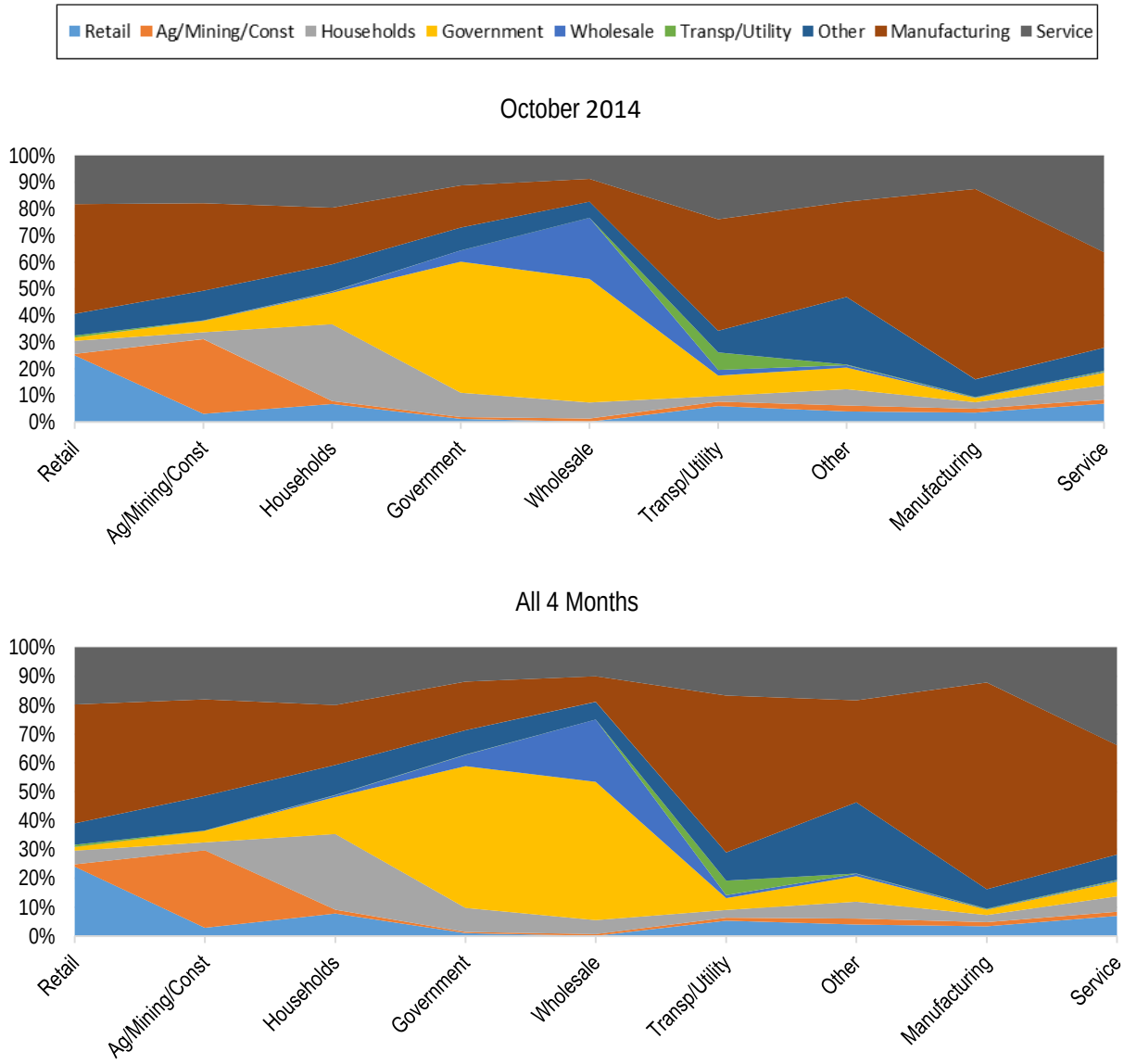
Figure 7-2 ATRI GPS Truck Events in 2014



In order to examine any seasonal differences, the tour formation and stops within each tour were summarized. Figure 7-3 presents the distribution of stops by purpose (y-axis) across different tour types (x-axis) derived from the processed ATRI truck trip table. It is very clear that the distributions are virtually the same across the four different months, and so combining them for data analyses and truck tour modeling should not be a problem.

Figure 7-3 Distribution of Stop Purposes by Tour Purpose




Retail
Ag/Mining/Const
Households
Government
Wholesale
Transp/Utility
Other
Manufacturing
Service
October 2014
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Retail
Ag/Mining/Const
Households
Government
Wholesale
Transp/Utility
Other
Manufacturing
Service
All 4 Months
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Retail
Ag/Mining/Const
Households
Government
Wholesale
Transp/Utility
Other
Manufacturing
Service

7.1.3 Comparison of 2014 ATRI Data versus 2011 ATRI Data

The 2014 data indicated about 291 GPS events from each truck which is twice as large as what the 2011 ATRI data showed (151). It is possible that the GPS technology has evolved quite a bit in three years that is now picking up truck GPS events more frequently. The number of tours per truck is a little over 2.0 that is less than the 2.6 estimate from the 2011 ATRI data. This also may be attributed to the higher GPS frequency in the 2014 ATRI data. Further comparison was conducted as a means to QA/QC the 2014 processed ATRI data. This included trip length frequency distributions (TLFDs) and stops per truck as shown in Figure 7-4 and Figure 7-5.

Figure 7-4 Trip Length Frequency Distributions - 2014 versus 2011 ATRI Data

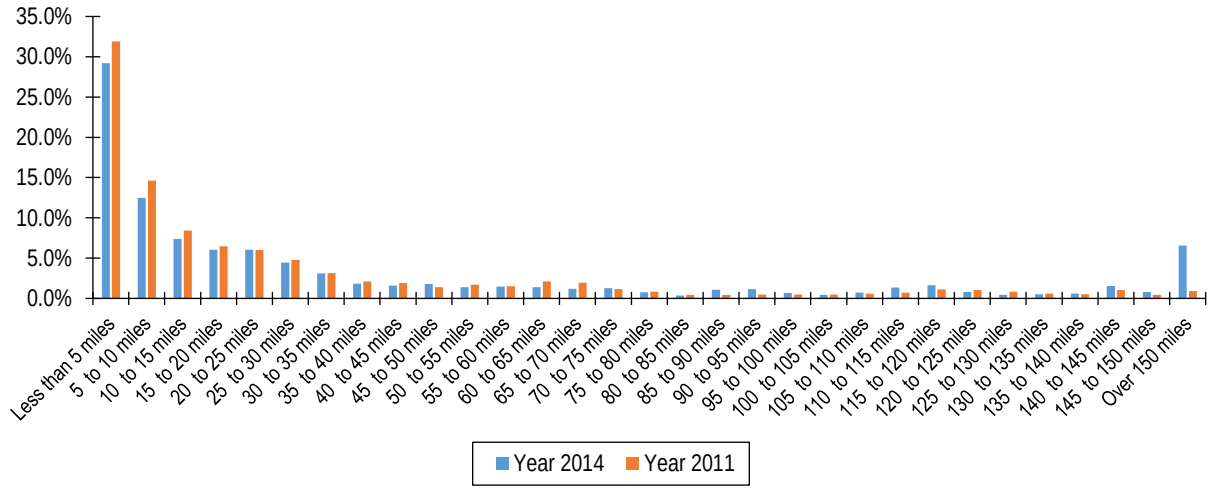
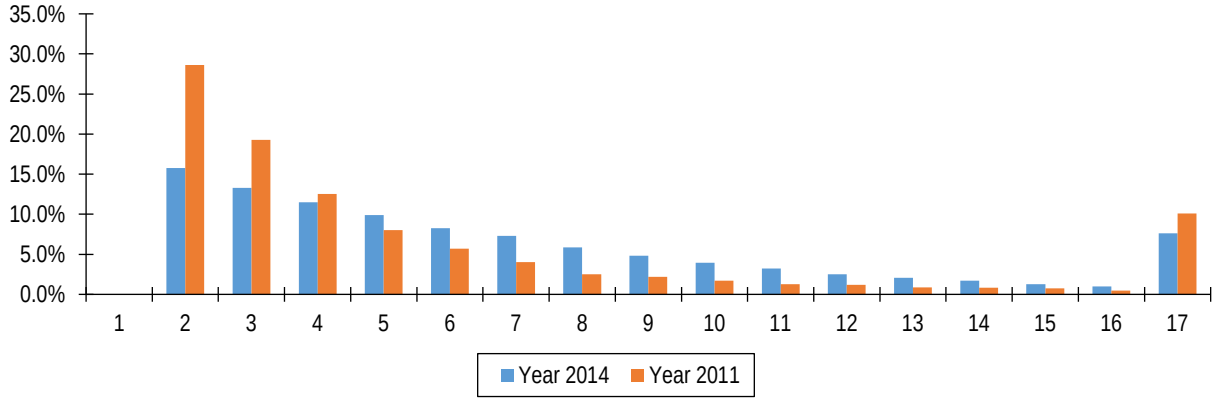


Figure 7-5 Stops per Truck - 2014 versus 2011 ATRI Data



For the most part, the TLFDs are very comparable across the two time points though the 2011 data shows a little higher number of shorter trips that are less than 10 miles in distance. The distribution of stops per truck in Figure 7-5 indicates that the 2014 data produced fewer percentage of truck tours with stops less than four per truck. This finding lines up well with the aggregate estimate of tours per truck across the two time points. However, it is to be noted that the data from 2011 and 2014 are from different samples of trucks; so at a disaggregate level, we are bound to see differences, but at the aggregate level, there are several similar trends.

7.1.4 Analyses with 2014 ATRI Data

The 2014 processed ATRI data was further analyzed before using it for estimating tour-based models. Figure 7-6 depicts the distribution of truck tours by stop duration, where the average stop duration is 42 minutes. This is the time for an average heavy truck to conduct business at a stop during a tour, which seems very reasonable.

Figure 7-6 Distribution of Tours by Stop Duration (2014 ATRI Data)

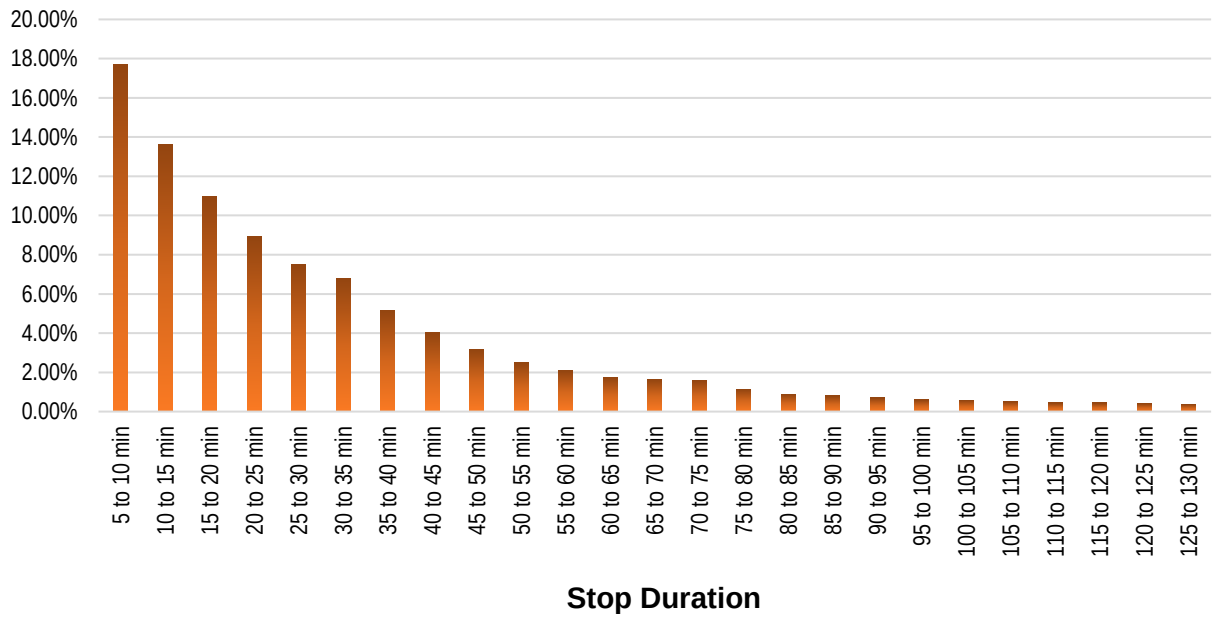
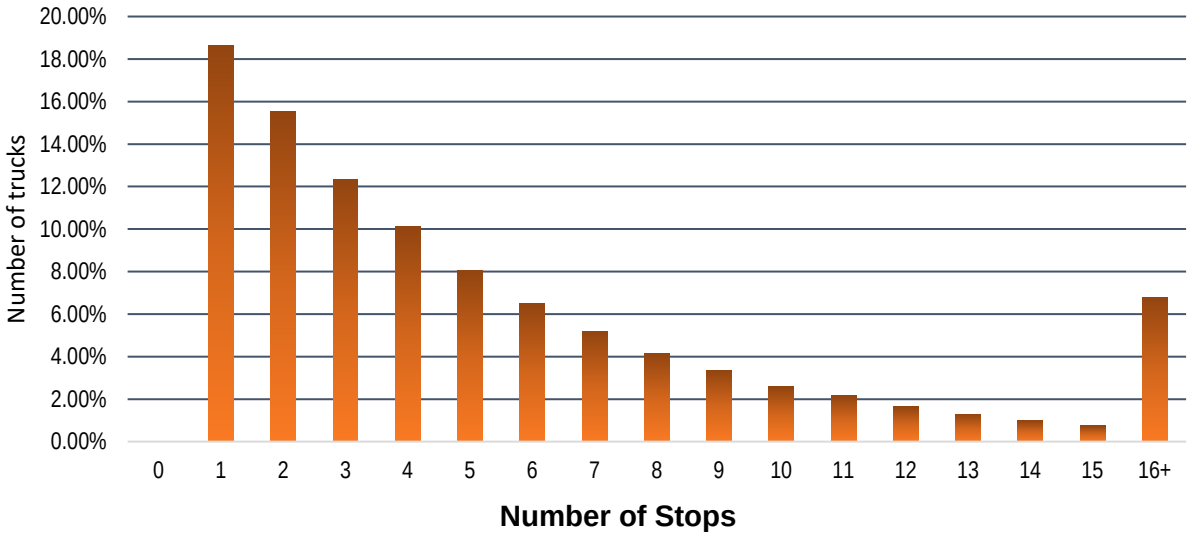


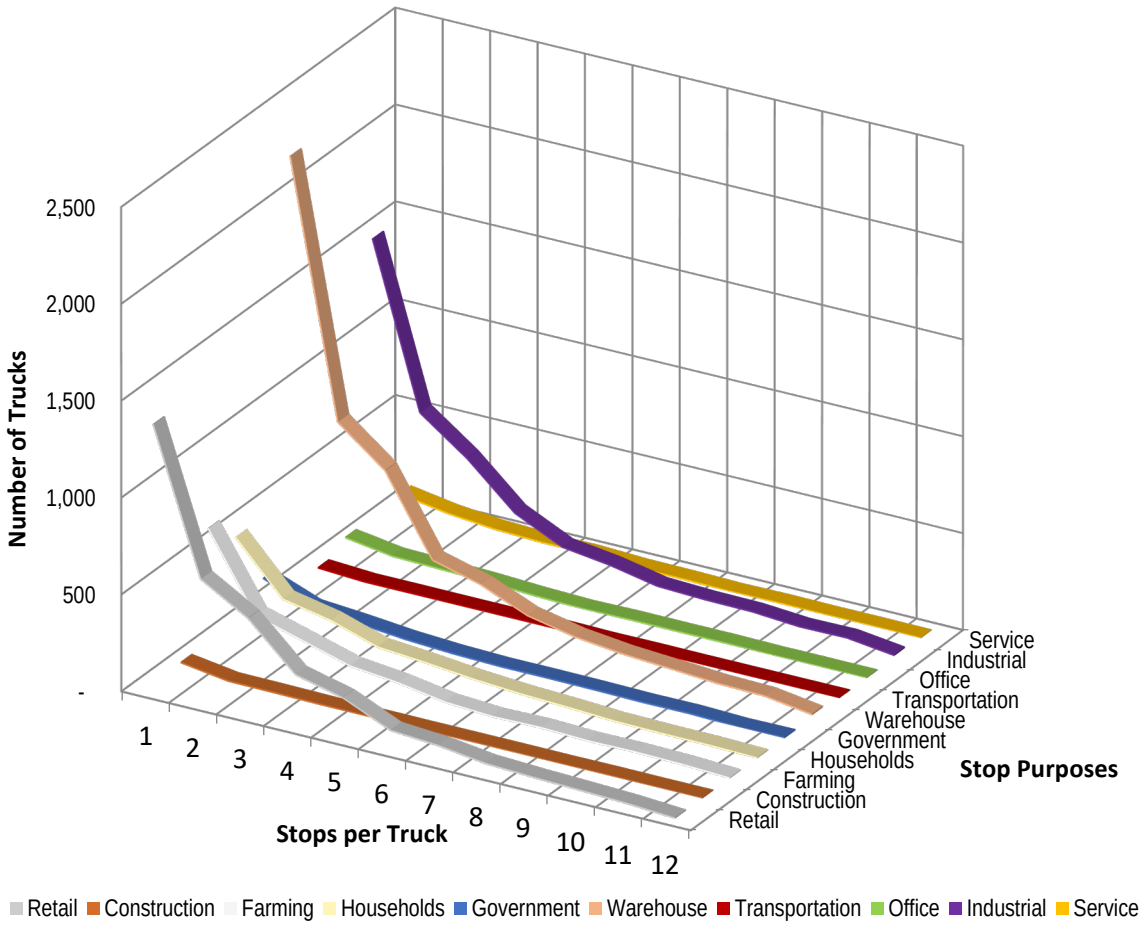
Figure 7-7 below shows the distribution of truck tours by stops per truck. The average is 3.39 stops per truck for all tours combined. However, there is a large variation in this metric across different tour purposes as shown in Figure 7-8.

Figure 7-7 Distribution of Tours by Stops per Truck (2014 ATRI Data)



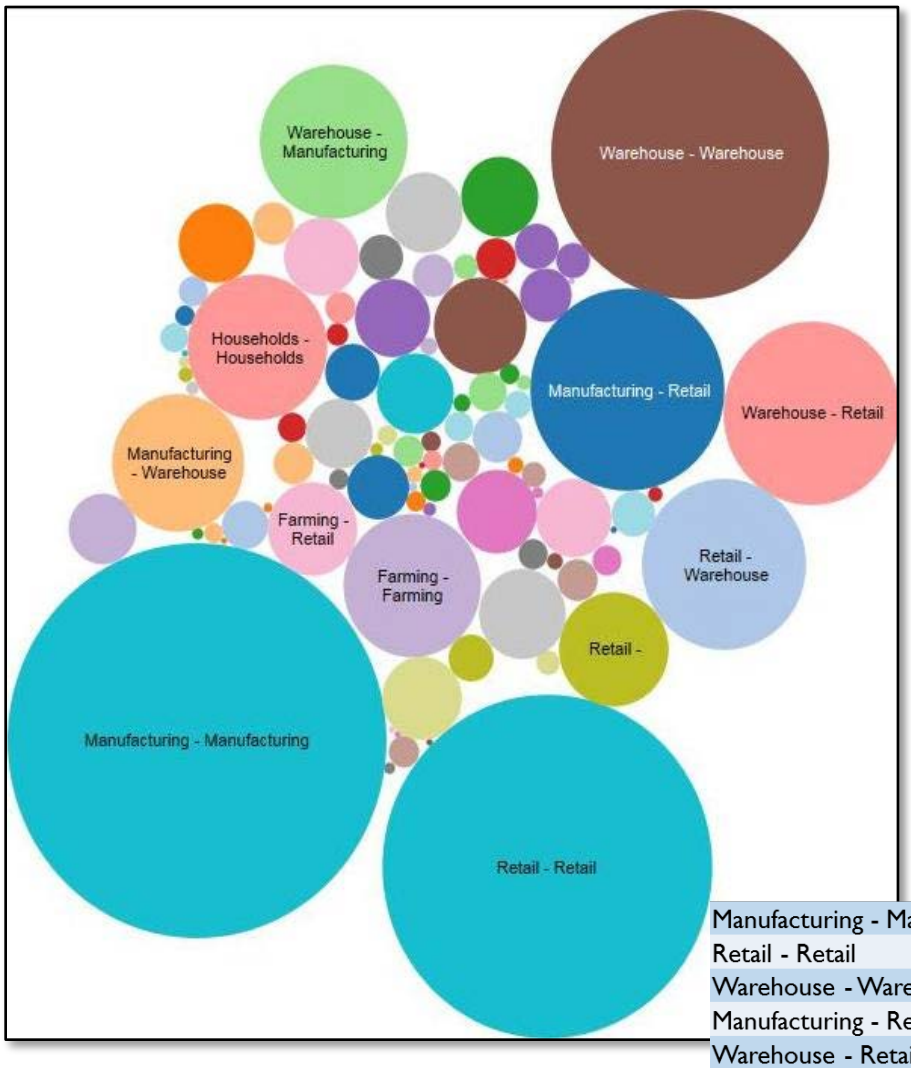
Retail, warehouse and industrial (or manufacturing) tour types are by far generating most number of stops per truck. This is seen by the curves going up the left side of the cube below while the rest of the tour types have pretty flat curves indicating a fewer number of stops per truck.

Figure 7-8 Stops per Truck by Tour Purpose (2014 ATRI Data)



The data also was analyzed to see if it can provide any good data for calibrating the supply chain model. So the logistic supply chains were summarized using the land use-to-land use (LU-LU) information. This is shown in Figure 7-9. The table in this figure shows that the top 11 logistic chains account for 75 percent of all truck trips. This information can be used to cross-check to see how the supply chain model is producing similar information and linking different LUs.

Figure 7-9 Logistic Supply Chains



Manufacturing - Manufacturing	22%
Retail - Retail	17%
Warehouse - Warehouse	12%
Manufacturing - Retail	6%
Warehouse - Retail	5%
Retail - Warehouse	4%
Warehouse - Manufacturing	3%
Households - Households	3%
Farming - Farming	3%
Manufacturing - Warehouse	3%
Retail - Manufacturing	2%
Households - Retail	1%
Farming - Retail	1%
Manufacturing - Farming	1%
Farming - Manufacturing	1%
Manufacturing - Households	1%
Retail - Households	1%
Government - Government	1%
Warehouse - Households	1%
Warehouse - Farming	1%
Households - Warehouse	1%
Retail - Farming	1%

Figure 7-10 presents the LU to LU interchanges that can be used to calibrate the tour-based location choice models.

Figure 7-10 LU to LU – MAG and PAG

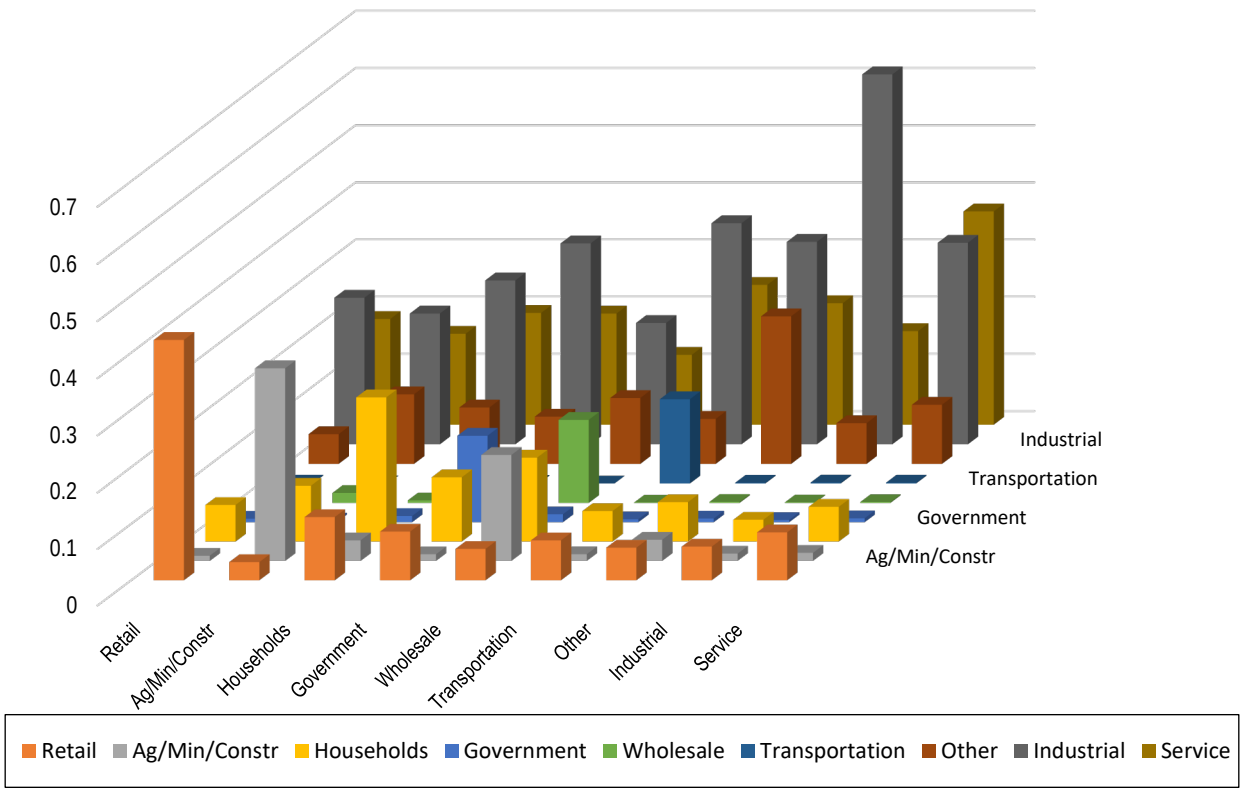


Figure 7-11 shows the distribution of average trip lengths in minutes by tour purpose. The purpose if represented along the x-axis while the travel times are along the y-axis. This figure shows boxplot where the second and third quartile of truck tours are inside the box while the long whiskers above boxes represent the fourth quartile and outliers. The average trip length across all purposes is shown by the 'average' line which is about 25 minutes.

Figure 7-11 Distribution of Truck Tours by Purpose (x-axis) and by Average Travel Time (y-axis)

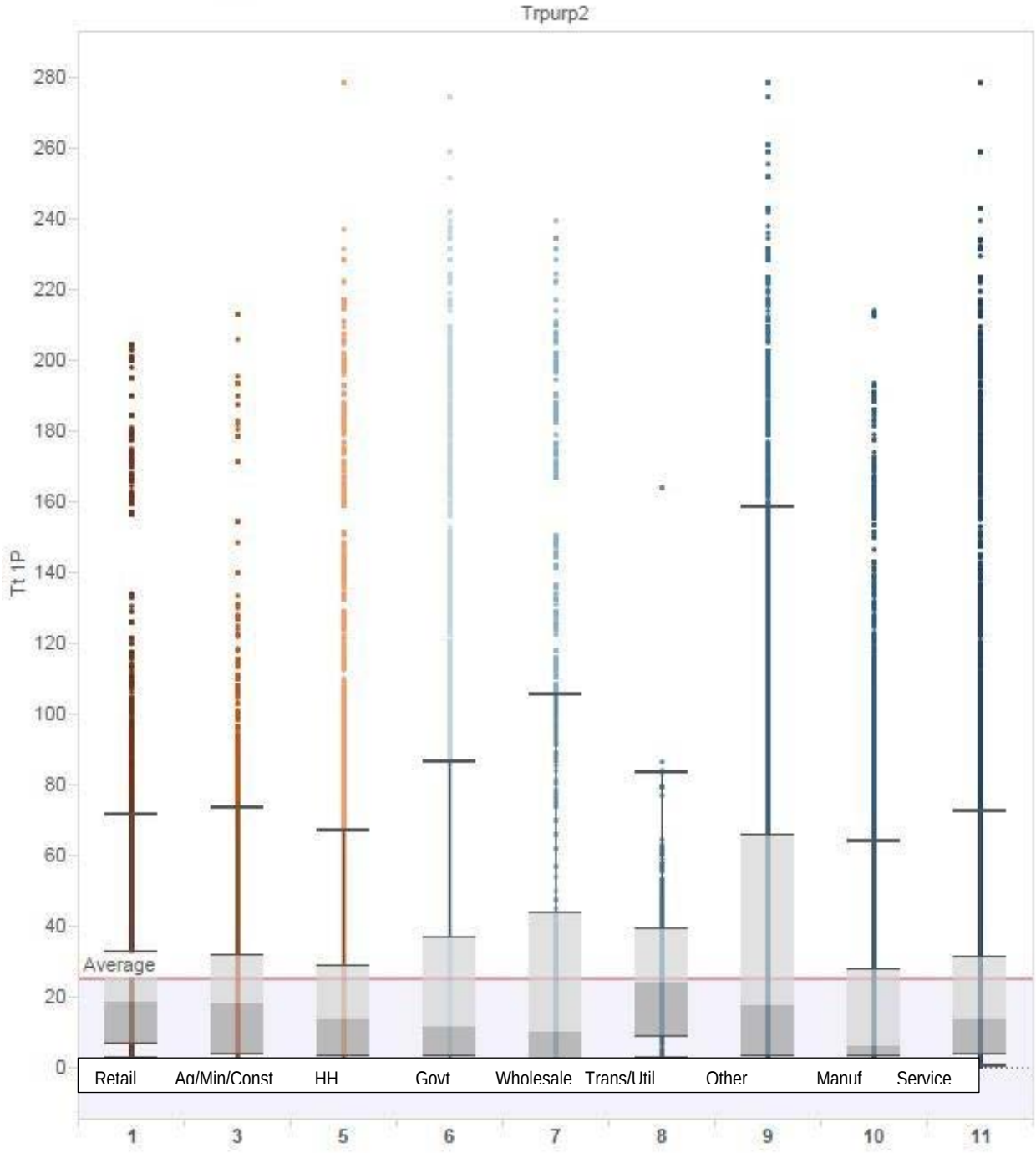
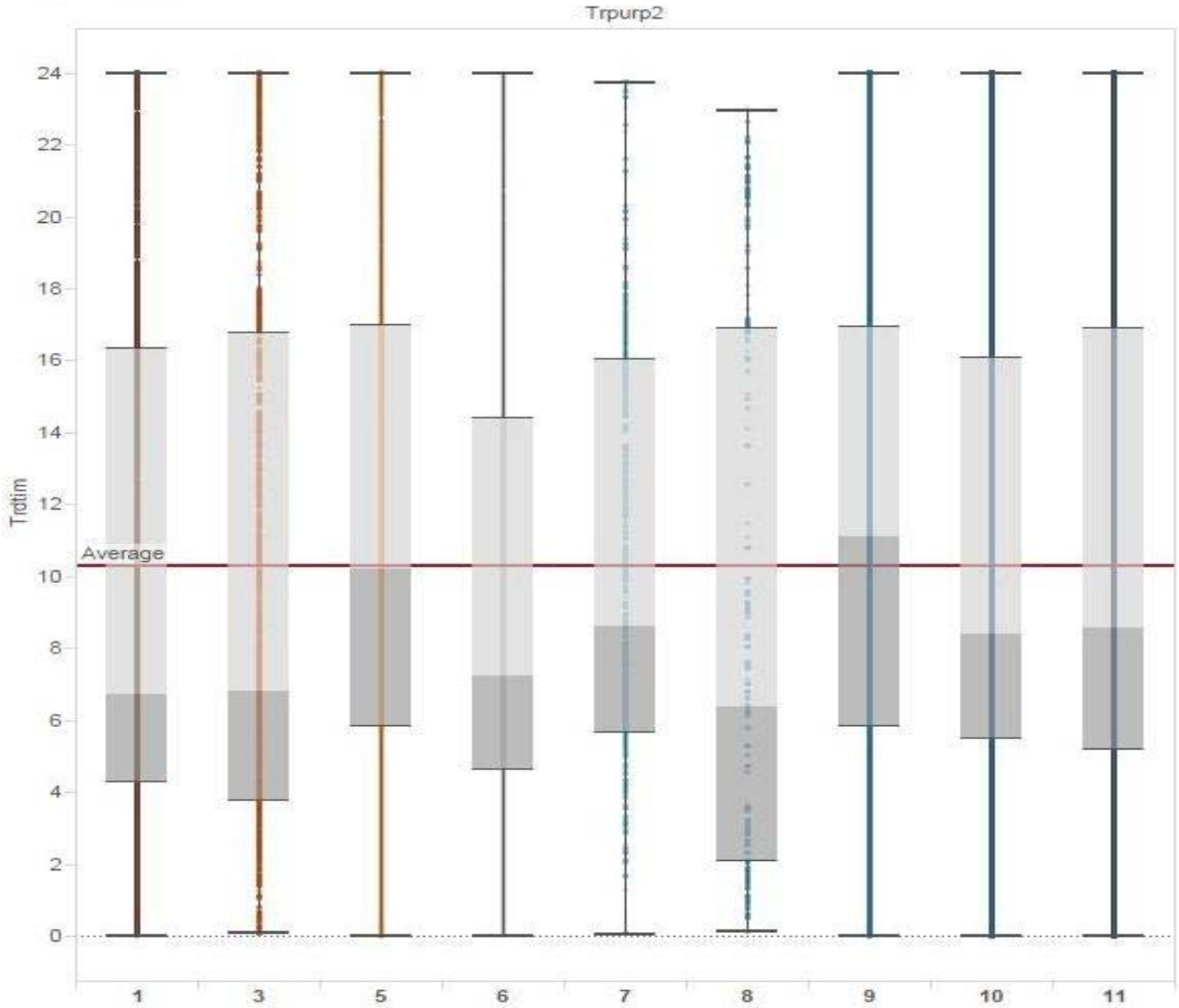


Figure 7-12 shows the time-of-day distribution of truck departures by tour purpose. The purpose is represented along the x-axis while departure time-of-day is along the y-axis. This figure shows a boxplot where the second and third quartile of truck tours are inside the box while the long whiskers above boxes represent the fourth quartile and outliers. The average departure time-of-day is a few minutes past 10 a.m. in the morning.

Figure 7-12 Distribution of Departure Time-of-Day (y-axis) by Purpose (x-axis)



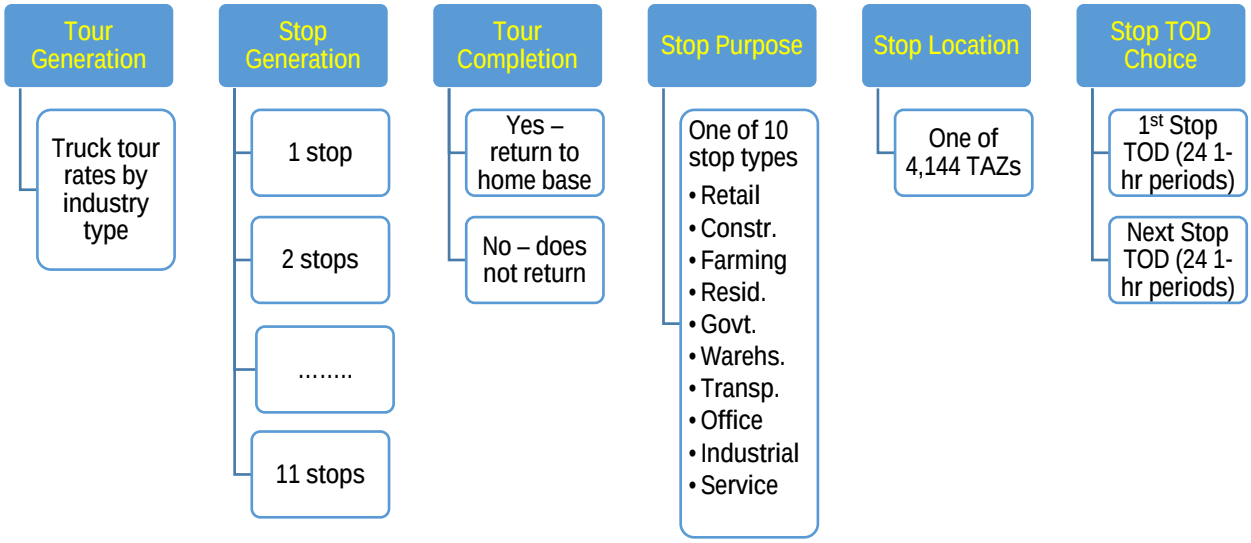
7.1.5 Conclusions

In summary, after the examination of the processed truck trip database and the associated descriptive statistics, we have determined that the criteria and thresholds used in processing the data seem reasonable and corroborated by additional QA/QC measures. Further, a comparison of this data against the previous 2011 data also indicates that the methodology used is technically sound and the data can be used for tour-based model estimation.

7.2 Tour-Based Methodology for Trucks

Truck tours are modeled through a sequence of models as shown in Figure 7-13. These models include predicting tour generation at the zonal level by tour purpose (i.e., starting land use type), the number of stops for each tour, the purpose of those stops, the location of stops, and the time-of-day for stops.

Figure 7-13 Tour-Based Truck Model



The basic concept behind truck tour-based models is consistent with activity-based passenger models. These models focus on the tour characteristics of truck trips after using shipment type information from the supply chain model. This model is very similar in structure and format of the current truck tour-based model developed for MAG using the 2011 ATRI truck GPS data[60]. In this model, a series of choice models are employed in order to determine the type of vehicle that will be used to conduct the business of the tour, the purpose of each stop (goods pickup or delivery, service, return to home), and the location of the next stop.

The tour-based components track the activity of trucks, and since these components will operate at the vehicle level, they will only generate estimates of a single mode[61]. Trucks are associated with establishments, and the truck activity is seen as a function of the type of activity that occurs at that establishment. The tour-based components operate within zones, as do the trip-based truck models, and the activity estimates are aggregated for all of the establishments in a zone. The tour-based model described here generates the number of stops that have to be made in each zone for a particular type of truck (e.g., retail, manufacturing), and then pieces these trips together into tours. The number of stops on a tour, the purpose of stops, the location, and time-of-day of stops are all estimated from the model based on the type of truck making the tour, the activities conducted by the truck, the characteristics of the stops, and the traffic conditions in the network.

7.3 Light Trucks

7.3.1 Model Estimation

The processed StreetLight truck GPS data was the basis for developing the estimation dataset for light trucks. The StreetLight data had a small sample of light trucks that yielded a truck tour database of 754 trucks that engaged in 3,023 tours and 5,989 trips on weekdays. This number includes only tours made

[60] Lemp, J., et. al., Development of a Tour-Based Truck Travel Demand Model, Presented at ITM, 2014.
[61] Fischer, M. J., An Innovative Framework for Modeling Freight Transportation in Los Angeles County, January 2005.

on weekdays. This dataset was narrowed down further for model estimation for a number of reasons. Only tours that made 11 or fewer stops were considered, which was used for consistency with the previous version of the truck tour model. In addition, only tours that started and remained in the MAG region were considered. External tours could not be considered since land use and travel time information was not available. After incorporating these restrictions, 3,003 tours remained, which are summarized by purpose in Table 7-2.

Table 7-2 Summary of Light Truck Trips

Stop Purpose	Trips	Share
Retail (base/reference purpose)	793	26%
Construction	106	4%
Agri/Mining	62	2%
Households	842	28%
Government	461	15%
Warehousing	0	0%
Transportation	22	1%
Office	95	3%
Industrial	142	5%
Service/Other	480	16%
Total	**3,003**	**100%**

One major assumption was that "tour purpose" was defined by the land use in the truck's starting location for the tour. That is, if a tour has a "retail" land use as the starting origin, then the whole tour is classified as a "retail tour". This is the best that could be done because the GPS data does not divulge the industry type of each truck. Upon further examining the data, it was found that nearly 90 percent of the truck tours are incomplete, that is, the trucks do not return to their home base or the starting origin. About 11 percent of truck tours completed the tour and returned to starting origin.

7.3.2 Tour Generation

The tour-generation model estimates the number of tours generated in each zone by truck tour purpose. Truck tour purpose is defined as the starting land use type of the tour. Using a combination of existing light truck trip rates, tour completion percentage and average stops per tour, the tour rates were computed by tour purpose. These rates are multiplied by the appropriate employment variable for each tour purpose to produce number of tours.

7.3.3 Stop Frequency Model

The stop frequency model predicts the number of stops on each truck tour. This is a multinomial logit (MNL) model where the number of stops represent the choice alternatives. The choices were limited to 11 stops for consistency with the previous version of the truck tour model.

The model estimation results are presented in Table 7-3. The key variables that were found to be significant in explaining stop frequency were the starting land use of the tour and zonal land use variables. The zonal variables that influence stop making behavior are employment by type and households at the starting zone of the tour.

Table 7-3 Stop Frequency Model – Light Trucks

Variables		Coefficient	t-statistic
Constants	2 Stops	-1.083	-13.0
	3 Stops	-1.890	-16.9
	4 Stops	-2.797	-19.4
	5 Stops	-2.938	-18.9
	6 Stops	-3.421	-18.4
	7 Stops	-3.960	-17.1
	8 Stops	-4.090	-16.4
	9 Stops	-4.898	-13.7
	10 Stops	-4.421	-14.8
	11 Stops	-5.419	-11.6
Number of Stops – Agri/Mining		-0.141	-0.7
Log (No. of Stops) – Agri/Mining		0.673	1.1
Number of Stops – Households		-0.273	-2.7
Log (No. of Stops) – Households		0.417	1.6
Number of Stops – Transportation		-3.016	-2.0
Log (No. of Stops) – Transportation		5.815	2.3
Number of Stops – Office		-0.100	-0.4
Log (No. of Stops) – Office		-0.075	-0.1
Number of Stops – Manufacturing		-1.444	-3.2
Log (No. of Stops) – Manufacturing		2.642	3.0
Number of Stops – Service		-0.060	-0.5
Log (No. of Stops) – Service		-0.133	-0.4
(Total Emp. Start Zone) / (No. of Stops)		0.00000	-0.1
(Total HHs Start Zone) / (No. of Stops)		-0.00029	-4.2
Number of Observations			**2,516**
Log Likelihood at Convergence			**-3,324.7**
Log Likelihood Constants Only			**-3,360.8**
Rho Squared			**0.0107**

Another key variable was the accessibility index that is expressed as a logarithmic function of travel time between two zones and employment at the destination zone.

$$Accessibility\ Index\ (AI) = \ln\left(1 + Sum_j\left[(\exp(-0.05 * TT_{ij}) / 10000) * EMP_j\right]\right)$$

Where,

TT_{ij} is travel time between i and j; and

EMP_j is employment at the destination zone j.

7.3.4 Tour Completion Model

The tour completion model predicts whether the tour returns to its starting location or ends at another location. This is a binomial logit choice model with two alternatives: tour does not complete or tour completes. These results are shown in Table 7-4.

Table 7-4 Tour Completion Model – Light Trucks

Variable	Coefficient	t-statistic
Constant	-2.069	-12.4
Tour Purpose = Agri/Mining	1.153	2.0
Tour Purpose = Households	0.616	3.8
Tour Purpose = Government	0.613	3.2
Tour Purpose = Transportation	1.537	3.1
Tour Purpose = Office	-0.144	-0.2
Tour Purpose = Manufacturing	0.920	1.3
Tour Purpose = Service	-0.142	-0.5
Number of Stops on Tour	-0.194	-3.4
No. of Stops – Agri/Mining	0.663	3.1
No. of Stops – Office	0.174	0.6
No. of Stops – Manufacturing	-0.422	-1.0
No. of Stops – Service	0.177	1.4
Total Employment in Start Zone	-4.26E-05	-1.0
Total Emp. – Agri/Mining	-4.14E-04	-0.7
Total Emp. – Office	-1.56E-04	-0.6
Total Emp. – Manufacturing	1.91E-05	0.1
Number of Observations		**3,011**
Log Likelihood at Convergence		**-1019.3**
Log Likelihood Constants Only		**-1073.9**
Rho Squared		**0.051**

The tour purpose and the number of stops on the tour make a significant impact on the tour completion probability. The greater the number of stops on the tour, the less the likelihood of a tour being completed. Manufacturing, transportation, and agricultural/mining tours are more prone to completing the tour while office and service trucks are less likely to completing the tour. The land use variables like employment have little influence on the completion of the tour for light trucks.

7.3.5 Stop Purpose Model

The stop purpose model predicts the purpose (i.e., land use type) of each stop that is predicted by the stop frequency model. This is a MNL model that predicts purpose of the stops in sequence, that is, from the first stop to the last stop. The alternatives or choices used in this model are the same land use types as defined in the trip-based truck model. The following alternative choice set is considered:

- Retail;
- Agri/Mining;
- Households;
- Government;
- Transportation;
- Office;
- Manufacturing; and
- Service.

The estimated model is presented in Table 7-5. All the coefficients are segmented by tour purpose. This influences the type of stop purpose significantly. The starting land use of the tour influences the stop purpose of subsequent stops on the tour. Other key explanatory variables that were found to be significant in this model are previous stop purpose, where certain purpose to purpose interchanges are much more prevalent than others, and the number of previous stops by purpose, which includes the total number of stops of each type already simulated for the tour. The accessibility indices that are segmented by tour purpose also were found to be significant in explaining the stop purpose. The zonal land use variables including employment by type and households at the starting zone also influence the stop purpose.

Table 7-5 Stop Purpose Model – Light Trucks

Variable	Alternative	Coefficient	t-statistic
Constants	Agri/Mining	-3.630	-18.3
	Households	2.269	8.7
	Government	-0.221	-1.2
	Transportation	-1.410	-1.7
	Office	-5.204	-6.7
	Manufacturing	-1.748	-8.4
	Service	0.805	2.8
Previous Stop Purpose = Agri/Mining	Agri/Mining	2.795	7.3
	Service	1.727	1.4
Previous Stop Purpose = Households	Manufacturing	-0.751	-3.5
Previous Stop Purpose = Government	Households	-0.286	-2.5
	Government	0.431	3.5
Previous Stop Purpose = Office	Office	1.561	4.2
	Service	0.628	2.9
Previous Stop Purpose = Manufacturing	Households	-0.778	-3.5
	Government	-0.699	-1.6
Previous Stop Purpose = Service	Transportation	0.752	2.0
Tour Purpose = Retail	Households	-0.509	-6.3
	Government	-0.591	-4.7
	Office	-0.215	-1.2
	Manufacturing	-0.679	-4.0
Tour Purpose = Agri/Mining	Agri/Mining	2.589	6.4
	Government	-1.613	-1.6
	Service	-2.643	-1.8
Tour Purpose = Households	Government	-0.623	-4.8
Tour Purpose = Government	Government	0.637	5.0
	Manufacturing	-1.207	-4.0
	Service	0.287	2.4
Tour Purpose = Transportation	Transportation	2.328	3.1
Tour Purpose = Office	Households	-0.754	-3.3
	Government	-0.988	-3.0
	Office	-0.660	-1.5
Tour Purpose = Manufacturing	Government	-0.839	-2.2
	Service	0.470	2.7
Tour Purpose = Service	Service	0.622	5.8
Number of Previous Retail Stops	Agri/Mining	-0.568	-2.6
	Households	-0.384	-4.5
	Government	-0.546	-6.1
	Transportation	-1.320	-1.8
	Office	-0.382	-2.2
	Service	-0.350	-3.3
Number of Previous Agri/Mining Stops	Agri/Mining	0.503	3.1
	Service	-1.852	-1.9
Number of Previous Households Stops	Households	0.454	7.5
	Office	-0.362	-2.0

Variable	Alternative	Coefficient	t-statistic
	Manufacturing	-0.788	-2.7
	Service	0.141	1.6
Number of Previous Government Stops	Households	0.279	3.4
	Government	0.395	5.1
Number of Previous Transportation Stops	Transportation	1.737	3.2
Number of Previous Office Stops	Government	-0.741	-2.2
Number of Previous Manufacturing Stops	Households	-0.669	-2.7
	Government	-0.758	-2.2
	Manufacturing	0.727	3.7
Number of Previous Service Stops	Agri/Mining	-1.641	-2.2
	Service	0.623	6.7
Total Stops on Tour	Agri/Mining	0.216	5.2
	Households	0.121	6.6
	Government	0.045	2.2
	Transportation	-0.492	-2.7
Log (1 + Stop Sequence Number)	Households	-1.620	-7.6
	Office	2.017	4.0
	Service	-0.954	-4.0
(AI to Households) / (Stop Seq. No.)	Households	-0.706	-9.0
(AI to Total Emp.) / (Stop Seq. No.)	Government	-0.116	-2.5
	Transportation	-0.411	-2.2
	Office	0.786	4.0
(AI to Manufacturing Emp.) / (Stop Seq. No.)	Manufacturing	0.279	2.0
(AI to Service Emp.) / (Stop Seq. No.)	Service	-0.529	-6.1
Number of Observations			**5,426**
Log Likelihood at Convergence			**-8,065.9**
Log Likelihood Constants Only			**-9,123.6**
Rho Squared			**0.116**

7.3.6 Stop Location Choice Model

The stop location choice model predicts the location of each stop simulated for the tour, and is similar in design to a destination choice model employed for distributing passenger trips. Every zone in the region is a potential choice for this model. Similar to any other destination choice model, size variables are included in the model. These include employment at the stop location by type.

Two types of accessibility variables are included in the model:

1. Direct zone-to-zone accessibility variables or travel time between: a) previous stop location to current stop location; and b) first stop location to current stop location; AND
2. Aggregate accessibility measures. This is important to describe the accessibility of a stop zone to employment types corresponding to the next stop purpose.

Other variables include zonal area type such as CBD, rural, and suburban. This is defined as a combination of employment and population density. Most of these variables are segmented by the

starting land use of the tour, previous stop purpose, current stop purpose, and number of stops on tour by purpose. These results are shown in Table 7-6.

Table 7-6 Stop Location Choice Model – Light Trucks

Utility Variables	Coefficient	t-statistic
Intrazonal with Previous Stop	-0.165	-0.6
Intrazonal with Tour Start Zone	1.739	6.1
Intrazonal with Previous Stop and Tour Start Zone	-2.525	-5.4
Intrazonal with Previous Stop – Stop Purpose = Manufacturing		N/A
Intrazonal with Tour Start Zone – Stop Purpose = Manufacturing		N/A
Intrazonal with Previous Stop – Prev. Purpose = Manufacturing		N/A
Log(1 + Pk Travel Time)	-2.283	-42.8
Log(1 + Pk Travel Time) – First Stop on Tour, Total Stops on Tour >= 2	0.212	2.9
Log(1 + Pk Travel Time) * Total Stops on Tour	-0.073	-6.3
Log(1 + Pk Return Travel Time) / (Stops Remaining + 1) – Total Stops on Tour >=2	-1.052	-13.3
Log(1 + Pk Travel Time) – Stop Purpose = Agri/Mining	-0.836	-3.7
Log(1 + Pk Travel Time) – Stop Purpose = Households	0.061	0.9
Log(1 + Pk Travel Time) – Stop Purpose = Transportation	0.289	1.1
Log(1 + Pk Travel Time) – Stop Purpose = Manufacturing	0.449	3.1
Size Variables		
Retail Stop, Retail Emp.		N/A
Retail Stop, Service Emp.	-2.999	-15.8
Agri/Mining Stop, Agri/Mining Emp.		N/A
Agri/Mining Stop, Total Emp.	-4.579	-6.3
Agri/Mining Stop, Construction Emp.	-0.790	-1.3
Households Stop, Households		N/A
Government Stop, Total Emp.		N/A
Government Stop, Retail Emp.		N/A
Government Stop, Agri/Mining Emp.	2.694	2.3
Government Stop, Service Emp.	1.291	2.4
Transportation Stop, Total Emp.		N/A
Transportation Stop, Service Emp.		N/A
Transportation Stop, Manufacturing Emp.		N/A
Office Stop, Total Emp.		N/A
Office Stop, Agri/Mining Emp.		N/A
Office Stop, Construction Emp.	0.745	0.5
Office Stop, Manufacturing Emp.		N/A

Utility Variables	Coefficient	t-statistic
Office Stop, Households	-1.511	-1.6
Manufacturing Stop, Manufacturing Emp.		N/A
Manufacturing Stop, Total Emp.	-3.981	-9.0
Manufacturing Stop, Warehouse Emp.	-0.566	-1.0
Service Stop, Total Emp.		N/A
Service Stop, Retail Emp.		N/A
Number of Observations		**5,403**
Log Likelihood at Convergence		**-8,698.3**
Log Likelihood at Zero		**-21,243.7**
Rho Squared		**0.591**

7.3.7 Stop Time-of-Day Choice Model

The stop time-of-day choice model predicts the time period of each stop on a tour. Two separate models were estimated for time-of-day choice. The first is used for the departure time of a tour's first trip, and the second is used for subsequent trips. The reason for defining two separate models is that subsequent trip departure times should depend, in part, on the timing of a tour's previous trips. Thus, duration between trips is an important variable in the subsequent time-of-day choice model. Both models are MNL models, where the alternatives include each one-hour period of the day (24 alternatives in total). In application, the one-hour periods are aggregated back to the four existing time periods used in the regional model – AM peak, midday, PM peak and night.

There are two main reasons for defining the alternatives as one-hour periods rather than the four existing time periods used in MAG's regional model. First, this ensures alternatives are of uniform size, and no special considerations are needed to adjust variables for the size. Second, the more refined time period definitions should allow travel time between stops to have important implications. In addition, it allows availability restrictions to be well defined for the subsequent stop period model. Since truckers involved in interstate commerce are not allowed to work more than 12 consecutive hours by law, availability restrictions are important.

The following variables were found to be significant in the time-of-day choice models shown in Table 7-7 and Table 7-8:

- Starting land use for the tour;
- Previous and current stop purpose;
- Number of stops on tour by purpose;
- Travel distance/time from previous stop to current stop; and
- Previous stop time-of-day (if not first stop).

Table 7-7 Time-of-Day Choice Model – First Trip – Light Trucks

TOD Period	Segment	Variable	Coefficient	t-statistic
AM	All	Constant	1.875	6.7
AM	Tour Purpose – Agri/Mining	Constant	-1.002	-2.2
AM	Tour Purpose – Households	Constant	-0.041	-0.2
AM	Tour Purpose – Government	Constant	1.090	6.2
AM	Tour Purpose – Transportation	Constant	0.126	0.2
AM	Tour Purpose – Office	Constant	-0.788	-2.2
AM	Tour Purpose – Manufacturing	Constant	0.627	1.4
AM	Tour Purpose – Service	Constant	0.442	3.0
AM	Stop Purpose – Households	Constant	0.320	2.7
AM	Stop Purpose – Manufacturing	Constant	0.635	1.7
AM	All	Number of Stops	0.020	0.6
AM	All	Tour Completed (0/1)	-0.211	-0.8
AM	Tour Purpose – Manufacturing	Tour Completed (0/1)	0.374	0.6
AM	All	Log(1 + Pk Travel Time to First	0.012	0.1
AM	All	AM Shift Variable	0.038	0.7
AM	Tour Purpose –Manufacturing	AM Shift Variable	0.239	1.1
MD	All	Constant	2.881	11.4
MD	Tour Purpose – Households	Constant	0.275	1.9
MD	Tour Purpose – Government	Constant	0.921	6.1
MD	Tour Purpose – Transportation	Constant	-0.115	-0.2
MD	Tour Purpose – Manufacturing	Constant	-0.512	-1.0
MD	Stop Purpose – Agri/Mining	Constant	-0.366	-1.4
MD	Stop Purpose – Manufacturing	Constant	0.605	1.8
MD	All	Number of Stops	-0.264	-7.1
MD	Tour Purpose – Manufacturing	Number of Stops	0.749	3.0
MD	All	Tour Completed (0/1)	-0.300	-1.3
MD	All	Log(1 + Pk Travel Time to First	-0.058	-0.7
MD	All	MD Shift Variable	0.007	0.3
PM	All	Constant	3.253	11.0
PM	Tour Purpose – Agri/Mining	Constant	0.953	2.2
PM	Tour Purpose – Households	Constant	-0.027	-0.2
PM	Tour Purpose – Transportation	Constant	-1.151	-1.3
PM	Tour Purpose – Office	Constant	-0.050	-0.2
PM	Stop Purpose – Agri/Mining	Constant	-1.654	-3.3

TOD Period	Segment	Variable	Coefficient	t-statistic
PM	Stop Purpose – Households	Constant	0.363	2.9
PM	Stop Purpose – Manufacturing	Constant	0.342	0.9
PM	All	Number of Stops	-0.807	-10.4
PM	All	Tour Completed (0/1)	0.199	0.8
PM	All	Log(1 + Pk Travel Time to First	0.155	1.7
PM	All	PM Shift Variable	-0.369	-8.7
PM	Tour Purpose – Manufacturing	PM Shift Variable	0.095	0.5
NT	All	NT Shift Variable Early	0.084	2.1
NT	Tour Purpose – Manufacturing	NT Shift Variable Early	0.184	1.4
NT	All	NT Shift Variable Late	0.174	4.9
NT	Tour Purpose – Manufacturing	NT Shift Variable Late	0.044	0.3

Number of Observations	**2,528**
Log Likelihood at Convergence	**-6,949**
Log Likelihood Constants Only	**-7,135**
Rho Squared	**0.026**

Table 7-8 Time-of-Day Choice Model – Subsequent Trips – Light Trucks

Variable	Coefficient	t-stat
Constant – 1-hour Period as Previous Stop	-0.281	-0.1
Constant – First 1-hour Period after Previous Stop	0.557	0.2
Constant – Second 1-hour Period after Previous Stop	0.486	0.2
Constant – Third 1-hour Period after Previous Stop	0.955	0.5
Constant – Fourth 1-hour Period after Previous Stop	1.264	0.7
Constant – Fifth or Sixth 1-hour Period after Previous Stop	1.086	0.7
Constant – Seventh or Eighth 1-hour Period after Previous Stop	0.850	0.7
Departure Time Shift Variable [62]	-0.753	-3.0
Departure Shift – Prev. Stop = Agri/Mining	0.130	1.7
Departure Shift – Prev. Stop = Households	0.023	0.6
Departure Shift – Prev. Stop = Government	0.059	1.3
Departure Shift – Prev. Stop = Office	0.148	1.8
Departure Shift – Prev. Stop = Manufacturing	0.008	0.1
Departure Shift – Prev. Stop = Service	-0.005	-0.1
Departure Shift * Log (1 + Peak Travel Time to Previous Stop)	0.038	2.1
Departure Shift * Total Number of Stops on Tour	-0.051	-5.4
Departure Shift * Total Number of Stops Simulated	-0.019	-1.5
Number of Observations		**2,495**
Log Likelihood at Convergence		**-3637.1**
Log Likelihood Constants Only		**-5754.8**
Rho Squared		**0.368**

7.3.8 Comparison to Survey

All the tour model components were coded and implemented in R. Each component was individually assessed and compared to observed survey data. The reasonability of the explanatory variables were determined by their magnitude, t-statistic, and their relation to the dependent variable. Some of the key findings include:

- Retail, wholesale, construction, and government tours make more stops than other tour purposes;
- Tours with only one stop are most frequently closed tours, and close tour frequency is at its lowest when the number of stops is about seven;
- Stop purpose is often strongly influenced by the tour type or the first land use of the truck origin;

[62] This variable is equal to the number of periods the alternative is after the departure period of the previous stop.

- Travel time has a negative effect on location choice utility, and is more pronounced as the number of stops on tour increase; and
- Previous stop purpose has a marginally significant impact on the time period of the next stop.

The individual model outputs also were compared against the Streetlight data to assess the model performance. These results are depicted in Figure 7-14 through Figure 7-17. These comparisons indicate that the model components are predicting very closely to the observed data for the most part. There are some differences which can be further improved upon with more rigorous calibration and validation of the model.

Figure 7-14 Stop Frequency Comparison - Tour Model Outputs versus Survey Data

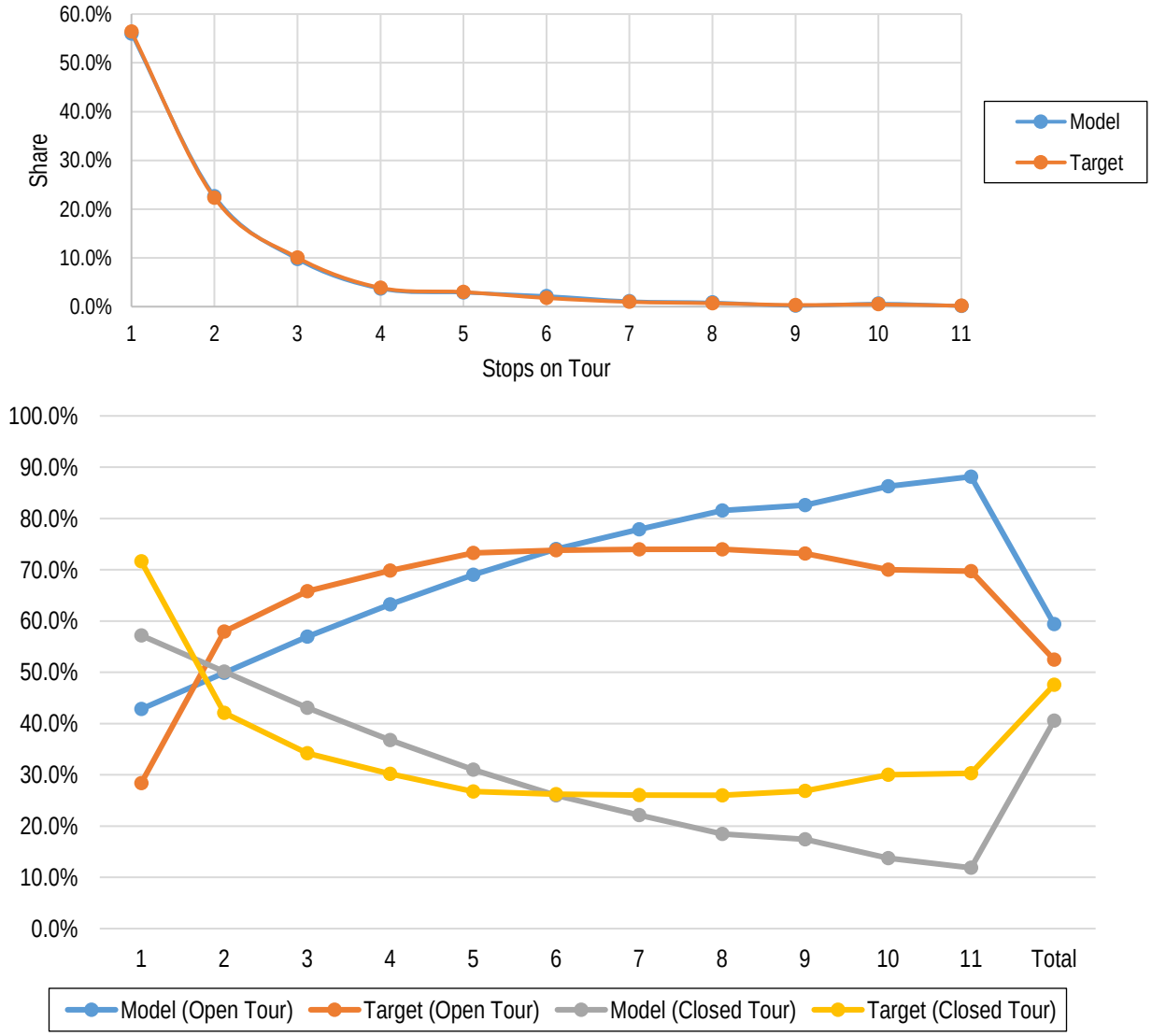


Figure 7-15 Stop Purpose Frequency - Model Outputs versus Survey Data

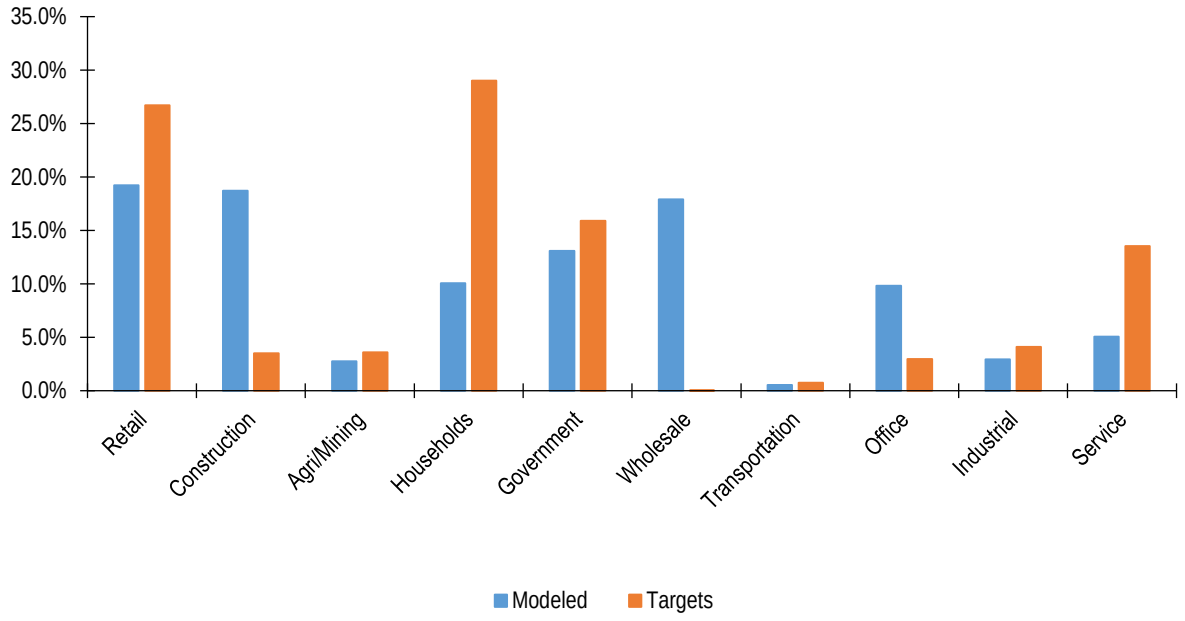


Figure 7-16 Time-of-Day of First Stop - Model Outputs versus Survey Data

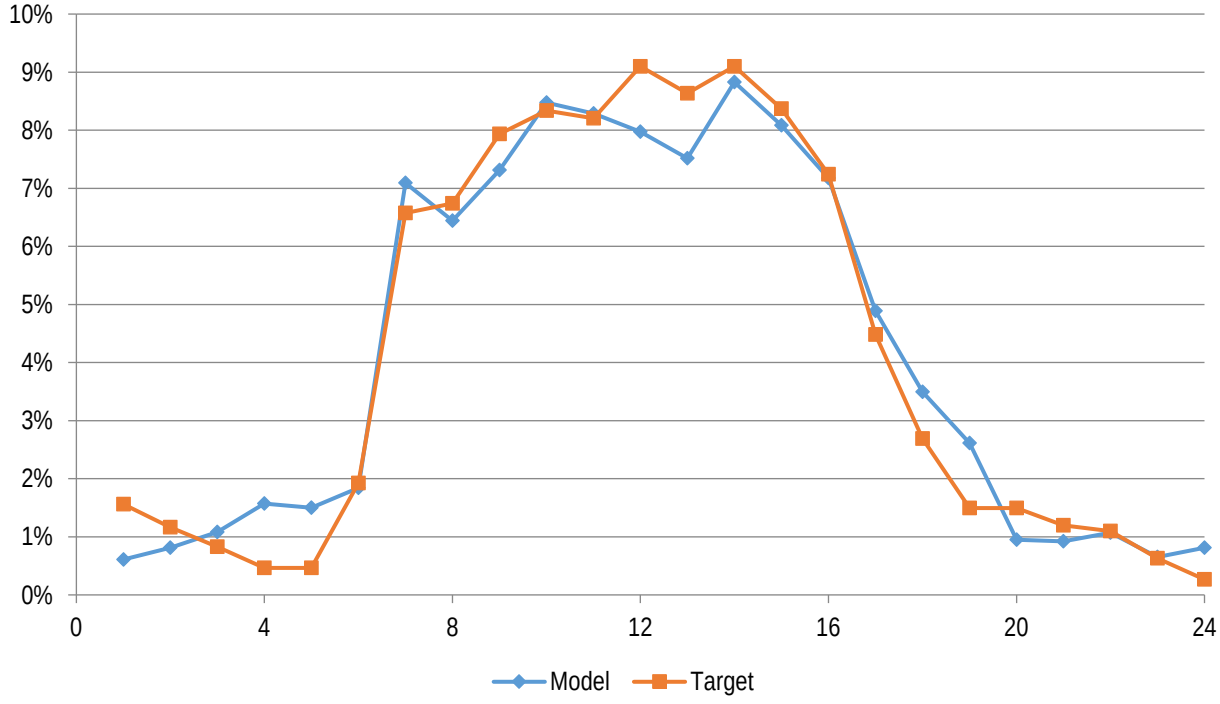
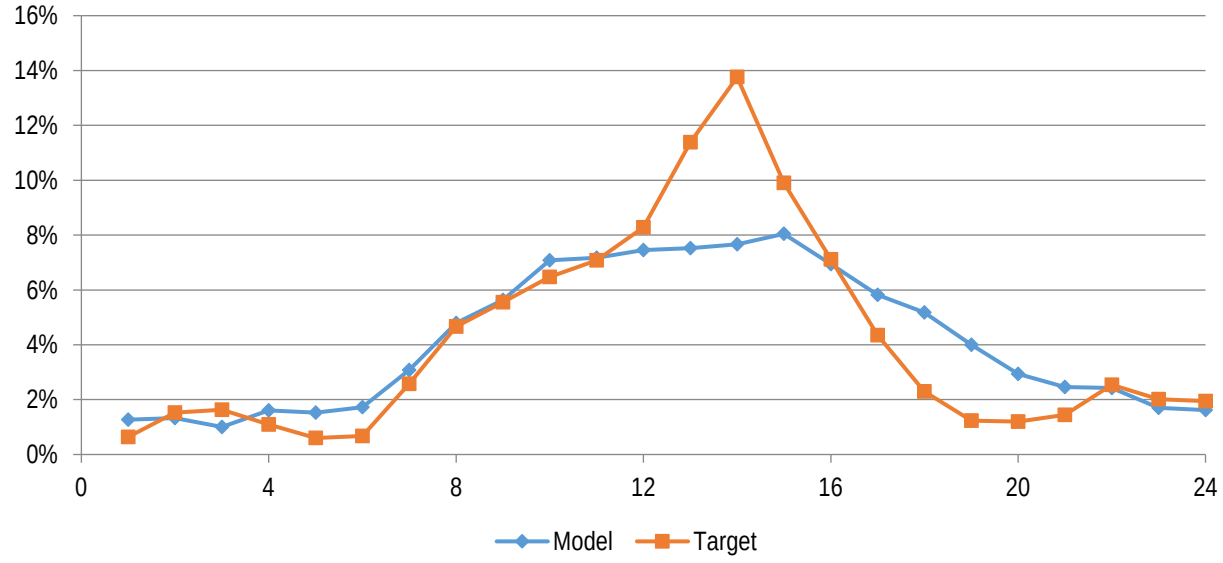


Figure 7-17 Time-of-Day of Subsequent Stops - Model Outputs versus Survey Data



7.4 Medium Trucks

7.4.1 Model Estimation

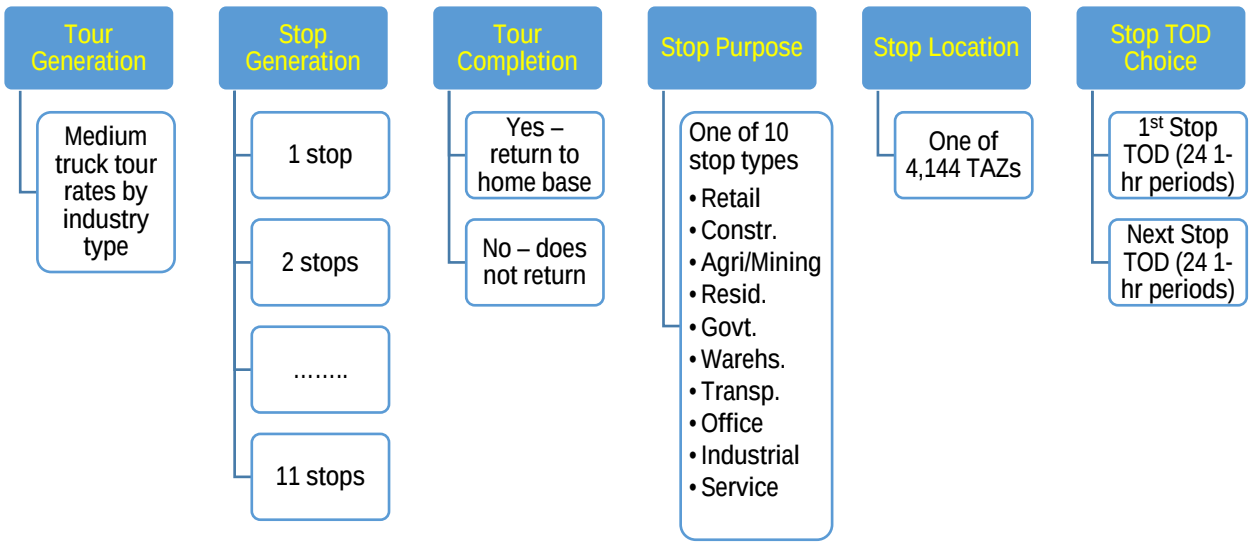
For medium trucks, the StreetLight truck GPS database was used to develop an estimation dataset. This database yielded a truck tour database of 12,206 trucks that engaged in 142,495 tours and 742,904 trips on weekdays. This number includes only tours made on weekdays. This dataset was narrowed down further for model estimation for a number of reasons. Only tours that made 11 or fewer stops were considered, which was used for consistency with the previous version of the truck tour model. In addition, only tours that started and remained in the MAG region were considered. External tours could not be considered since land use and travel time information was not available. After incorporating these restrictions, 128,532 tours remained, which are summarized by purpose in Table 7-9 below.

Table 7-9 Summary of Medium Truck Trips

Stop Purpose	Trips	Share
Retail	21,544	17%
Construction	6,744	5%
Agriculture/Mining	1,717	1%
Households	52,156	41%
Government	10,443	8%
Warehousing	145	0%
Transportation	1,774	1%
Office	5,162	4%
Industrial	12,439	10%
Service / Other	16,408	13%
Total	128,532	100%

One major assumption was that "tour purpose" was defined by the land use in the truck's starting location for the tour. That is, if a tour has a "retail" land use as the starting origin, then the whole tour is classified as a "retail tour". This is the best that could be done because the GPS data does not divulge the industry type of each truck. Upon further examining the data, it was found that nearly 90 percent of the truck tours are incomplete, that is, the trucks do not return to their home base or the starting origin. About 10 percent of truck tours completed the tour and returned to starting origin. Truck tours are modeled through a sequence of models as shown in Figure 7-18. These models include predicting tour generation at the zonal level by tour purpose (i.e., starting land use type), the number of stops for each tour, the purpose of those stops, the location of stops, and the time-of-day for stops.

Figure 7-18 Tour-Based Truck Model



7.4.2 Tour Generation

The tour generation model estimates the number of tours generated in each zone by truck tour purpose. Truck tour purpose is defined as the starting land use type of the tour. Using a combination of existing light truck trip rates, tour completion percentage and average stops per tour, the tour rates were computed by tour purpose. These rates are multiplied by the appropriate employment variable for each tour purpose to produce number of tours.

7.4.3 Stop Frequency Model

The stop frequency model predicts the number of stops on each truck tour. This is a multinomial logit (MNL) model where the number of stops represent the choice alternatives. The choices were limited to 11 stops for consistency with the previous version of the truck tour model.

The model estimation results are presented in Table 7-10. The key variables that were found to be significant in explaining stop frequency were the starting land use of the tour and zonal land use variables. The zonal variables that influence stop making behavior are employment and households at the starting zone of the tour.

Table 7-10 Stop Frequency Model - Medium Trucks

Variables		Coefficient	t-statistic
Constants	2 Stops	-0.573	-41.5
	3 Stops	-0.956	-55.8
	4 Stops	-1.205	-64.1
	5 Stops	-1.354	-69.3
	6 Stops	-1.529	-76.0
	7 Stops	-1.723	-83.0
	8 Stops	-1.956	-89.5
	9 Stops	-2.200	-93.9
	10 Stops	-2.480	-96.4
	11 Stops	-2.809	-97.3
Number of Stops – Agri/Mining		0.006	0.2
Log (Number of Stops) – Agri/Mining		-0.075	-0.8
Number of Stops – Households		0.145	18.1
Log (No. of Stops) – Households		-0.212	-7.1
Number of Stops – Transportation		0.204	7.7
Log (No. of Stops) – Transportation		-0.512	-5.2
Number of Stops – Office		-0.054	-3.0
Log (No. of Stops) – Office		0.206	3.3
Number of Stops – Manufacturing		0.053	4.5
Log (No. of Stops) – Manufacturing		-0.027	-0.6
Number of Stops – Service		0.069	6.2
Log (No. of Stops) – Service		-0.090	-2.3
(Total Emp. Start Zone) / (No. of Stops)		0.00000	0.6
(Total HHs Start Zone) / (No. of Stops)		-0.00016	-14.2
Number of Observations			**110,566**
Log Likelihood at Convergence			**-242,360.8**
Log Likelihood Constants Only			**-243,445.4**
Rho Squared			**0.004**

7.4.4 Tour Completion Model

The tour completion model predicts whether the tour returns to its starting location or ends at another location. This is a binomial logit choice model with two alternatives: tour does not complete or tour completes. These results are shown in Table 7-11.

Table 7-11 Tour Completion Model – Medium Trucks

Variable	Coefficient	t-statistic
Constant	-2.223	-82.3
Tour Purpose = Agri/Mining	1.554	15.6
Tour Purpose = Households	-0.162	-6.1
Tour Purpose = Government	0.590	18.0
Tour Purpose = Transportation	0.325	4.4
Tour Purpose = Office	-0.026	-0.2
Tour Purpose = Manufacturing	0.756	12.5
Tour Purpose = Service	0.105	2.1
Number of Stops on Tour	-0.022	-5.7
No. of Stops – Agri/Mining	-0.042	-2.0
No. of Stops – Office	0.011	0.6
No. of Stops – Manufacturing	-0.044	-4.3
No. of Stops – Service	-0.012	-1.2
Total Employment in Start Zone	2.59E-05	3.9
Total Emp. – Agri/Mining	-3.39E-04	-5.8
Total Emp. – Office	-2.84E-05	-1.2
Total Emp. – Manufacturing	-8.02E-05	-4.8
Number of Observations		**132,017**
Log Likelihood at Convergence		**-42464.9**
Log Likelihood Constants Only		**-43156.1**
Rho Squared		**0.016**

The tour purpose and the number of stops on the tour make a significant impact on the tour completion probability. The greater the number of stops on the tour, the less the likelihood of a tour being completed. Manufacturing, agricultural/mining, and government tours are more prone to completing the tour, though the effect dissipates even faster with higher numbers of stops. The land use variables like employment also influence the completion of the tour as they do the stop making behavior.

7.4.5 Stop Purpose Model

The stop purpose model predicts the purpose (i.e., land use type) of each stop that is predicted by the stop frequency model. This is a MNL model that predicts purpose of the stops in sequence, that is, from the first stop to the last stop. The alternatives or choices used in this model are the same land use types as defined in the trip-based truck model. The following alternative choice set is considered:

- Retail;
- Agriculture/Mining;

- Households;
- Government;
- Transportation;
- Office;
- Manufacturing; and
- Service.

The estimated model is presented in Table 7-12. All the coefficients are segmented by tour purpose. This influences the type of stop purpose significantly. The starting land use of the tour influences the stop purpose of subsequent stops on the tour. Other key explanatory variables that were found to be significant in this model are previous stop purpose, where certain purpose to purpose interchanges are much more prevalent than others, and the number of previous stops by purpose, which includes the total number of stops of each type already simulated for the tour. The accessibility indices that are segmented by tour purpose also were found to be significant in explaining the stop purpose. The zonal land use variables, including employment by type and households at the starting zone also influence the stop purpose.

The accessibility index that is expressed as a logarithmic function of travel time between two zones and employment at the destination zone.

$$\textit{Accessibility Index (AI)} = \ln\left(1 + \mathrm{Sum}_j\left[\left(\exp\left(-0.05 * TT_{ij}\right) / 10000\right) * EMP_j\right]\right)$$

Where;

TT_{ij} is travel time between i and j; and

EMP_j is employment at the destination zone j.

Table 7-12 Stop Purpose Model - Medium Trucks

Variable	Alternative	Coefficient	t-statistic
Constants	Agri/Mining	-3.777	-53.3
	Households	1.483	46.4
	Government	-1.062	-23.1
	Transportation	-3.752	-36.6
	Office	-3.523	-52.3
	Manufacturing	-1.298	-43.0
	Service	-0.214	-5.8
Previous Stop Purpose = Retail	Agri/Mining	0.080	2.4
	Transportation	0.037	1.0
	Office	0.183	9.7
	Manufacturing	0.028	1.3
	Service	0.051	4.2
Previous Stop Purpose = Agri/Mining	Agri/Mining	1.172	26.7
	Households	-0.117	-3.0
	Government	0.525	11.6
	Office	-0.251	-2.7
	Manufacturing	0.198	3.2
	Service	0.061	1.0
Previous Stop Purpose = Households	Households	-0.058	-6.5
	Government	0.026	1.6
	Transportation	-0.277	-7.3
	Manufacturing	-0.059	-2.6
Previous Stop Purpose = Government	Households	-0.159	-9.5
	Government	0.152	7.3
	Manufacturing	-0.220	-6.3
	Service	-0.040	-1.9
Previous Stop Purpose = Transportation	Households	-0.533	-15.9
	Government	-0.257	-4.7
	Transportation	0.760	15.3
	Service	0.101	2.6
Previous Stop Purpose = Office	Agri/Mining	-0.774	-9.2
	Households	-0.193	-9.6
	Government	-0.214	-6.1
	Office	0.232	8.5
	Service	0.057	2.4
Previous Stop Purpose = Manufacturing	Agri/Mining	0.067	1.3
	Households	-0.364	-20.2
	Government	-0.172	-5.9
	Office	-0.045	-1.4
	Manufacturing	0.073	3.0
	Service	0.048	2.6

Variable	Alternative	Coefficient	t-statistic
Previous Stop Purpose = Service	Agri/Mining	-0.400	-7.6
	Households	-0.213	-16.8
	Government	-0.038	-1.8
	Transportation	0.014	0.3
	Office	0.114	4.9
	Manufacturing	-0.077	-3.3
	Service	0.059	4.3
Tour Purpose = Retail	Agri/Mining	0.046	0.8
	Households	-0.231	-15.1
	Government	-0.269	-10.5
	Office	-0.214	-7.1
	Manufacturing	-0.181	-7.4
	Service	-0.228	-11.6
Tour Purpose = Agri/Mining	Agri/Mining	2.379	35.7
	Households	-0.292	-6.6
	Government	0.384	7.3
	Office	-0.329	-3.3
	Manufacturing	-0.200	-3.0
	Service	-0.532	-7.9
Tour Purpose = Households	Agri/Mining	-0.029	-0.5
	Households	0.090	6.5
	Government	-0.094	-4.0
	Office	0.091	3.4
	Manufacturing	-0.361	-15.7
	Service	0.058	3.3
Tour Purpose = Government	Agri/Mining	1.636	28.4
	Households	0.244	12.2
	Government	1.089	39.7
	Office	-0.257	-6.1
	Manufacturing	-0.306	-8.2
	Service	0.159	6.1
Tour Purpose = Transportation	Transportation	1.676	43.4
Tour Purpose = Office	Agri/Mining	0.035	0.4
	Households	-0.038	-1.8
	Government	-0.083	-2.3
	Office	0.348	9.5
	Manufacturing	-0.752	-18.0
	Service	-0.063	-2.3
Tour Purpose = Manufacturing	Agri/Mining	0.159	2.5
	Households	-0.356	-20.1
	Government	-0.229	-7.8
	Transportation	0.016	0.4
	Office	-0.234	-7.0

Variable	Alternative	Coefficient	t-statistic
	Manufacturing	0.257	10.3
	Service	-0.250	-11.4
Tour Purpose = Service	Agri/Mining	0.288	4.6
	Households	-0.132	-8.2
	Government	-0.126	-4.6
	Transportation	0.394	11.0
	Office	-0.199	-6.3
	Manufacturing	-0.224	-8.5
	Service	0.340	17.6
Number of Previous Retail Stops	Agri/Mining	-0.446	-28.6
	Households	-0.248	-56.2
	Government	-0.333	-47.5
	Transportation	-0.215	-15.6
	Office	-0.321	-36.7
	Manufacturing	-0.261	-45.0
	Service	-0.188	-35.1
Number of Previous Agri/Mining Stops	Agri/Mining	0.858	53.6
	Households	0.014	1.0
	Government	0.047	2.7
	Office	-0.170	-5.9
	Manufacturing	-0.054	-2.4
	Service	-0.415	-17.8
Number of Previous Households Stops	Households	0.418	108.9
	Office	-0.150	-17.9
	Manufacturing	-0.255	-32.4
	Service	0.054	10.7
Number of Previous Government Stops	Households	0.068	8.9
	Government	0.611	71.0
	Office	-0.052	-3.8
	Manufacturing	-0.026	-1.9
	Service	0.094	9.6
Number of Previous Transportation Stops	Transportation	0.929	53.2
	Manufacturing	0.157	9.3
	Service	-0.020	-1.2
Number of Previous Office Stops	Households	-0.136	-14.7
	Government	-0.115	-7.6
	Transportation	-0.166	-6.5
	Office	0.577	50.4
	Service	0.071	6.7
Number of Previous Manufacturing Stops	Agri/Mining	0.002	0.1
	Households	-0.164	-19.9
	Government	-0.133	-10.3
	Office	-0.127	-8.6

Variable	Alternative	Coefficient	t-statistic
	Manufacturing	0.588	73.0
Number of Previous Service Stops	Agri/Mining	-0.447	-18.5
	Households	-0.012	-2.2
	Government	-0.060	-6.7
	Transportation	-0.116	-6.1
	Office	-0.025	-2.5
	Service	0.422	73.7
Total Stops on Tour	Agri/Mining	0.038	9.5
	Households	0.049	39.5
	Government	0.009	4.3
	Transportation	0.038	8.5
	Office	0.017	7.0
	Manufacturing	0.004	1.9
	Service	0.041	27.4
Log (1 + Stop Sequence Number)	Agri/Mining	0.274	8.3
	Households	-0.950	-46.7
	Government	-0.103	-4.1
	Transportation	0.066	1.1
	Office	0.949	22.4
	Service	-0.607	-26.2
(AI to Agri/Mining Emp.) / (Stop Seq. No.)	Agri/Mining	0.854	3.6
(AI to Households) / (Stop Seq. No.)	Households	-0.623	-68.2
(AI to Total Emp.) / (Stop Seq. No.)	Government	-0.177	-15.2
(AI to Total Emp.) / (Stop Seq. No.)	Transportation	-0.007	-0.2
(AI to Total Emp.) / (Stop Seq. No.)	Office	0.353	21.5
(AI to Manufacturing Emp.) / (Stop Seq. No.)	Manufacturing	-0.068	-5.1
(AI to Service Emp.) / (Stop Seq. No.)	Service	-0.309	-30.3
Number of Observations	**640,953**		
Log Likelihood at Convergence	**-940170.9**		
Log Likelihood Constants Only	**-1039658.7**		
Rho Squared	**0.096**		

7.4.6 Stop Location Choice Model

The stop location choice model predicts the location of each stop simulated for the tour, and is similar in design to a destination choice model employed for distributing passenger trips. Every zone in the region is a potential choice for this model. Similar to any other destination choice model, size variables are included in the model. These include employment at the stop location by type.

Two types of accessibility variables are included in the model:

1. Direct zone-to-zone accessibility variables or travel time between: a) previous stop location to current stop location; and b) first stop location to current stop location.

2. Aggregate accessibility measures. This is important to describe the accessibility of a stop zone to employment types corresponding to the next stop purpose.

Other variables include zonal area type such as CBD, rural, and suburban. This is defined as a combination of employment and population density. Most of these variables are segmented by the starting land use of the tour, previous stop purpose, current stop purpose, and number of stops on tour by purpose. These results are shown in Table 7-13.

Table 7-13 Stop Location Choice Model - Medium Trucks

Utility Variables	Coefficient	t-statistic
Intrazonal with Previous Stop	0.841	17.8
Intrazonal with Tour Start Zone	5.084	101.8
Intrazonal with Previous Stop and Tour Start Zone	-1.932	-14.7
Intrazonal with Previous Stop – Stop Purpose = Manufacturing	2.027	8.7
Intrazonal with Tour Start Zone – Stop Purpose = Manufacturing	0.332	2.1
Intrazonal with Previous Stop – Prev. Purpose = Manufacturing	-0.804	-3.8
Log(1 + Pk Travel Time)	-2.000	-207.0
Log(1 + Pk Travel Time) – First Stop on Tour, Total Stops on Tour >= 2	0.431	45.4
Log(1 + Pk Travel Time) * Total Stops on Tour	-0.089	-78.6
Log(1 + Pk Return Travel Time) / (Stops Remaining + 1) – Total Stops on Tour >=2	-1.048	-123.2
Log(1 + Pk Travel Time) – Stop Purpose = Agri/Mining	-1.024	-27.9
Log(1 + Pk Travel Time) – Stop Purpose = Households	0.172	20.7
Log(1 + Pk Travel Time) – Stop Purpose = Transportation	0.461	13.8
Log(1 + Pk Travel Time) – Stop Purpose = Manufacturing	0.435	24.6
Size Variables		
Retail Stop, Retail Emp.		N/A
Retail Stop, Service Emp.	-2.561	-151.2
Agri/Mining Stop, Agri/Mining Emp.		N/A
Agri/Mining Stop, Total Emp.	-4.343	-57.5
Agri/Mining Stop, Construction Emp.	-2.221	-27.0
Households Stop, Households		N/A
Government Stop, Total Emp.		N/A
Government Stop, Retail Emp.		N/A
Government Stop, Agri/Mining Emp.	1.252	6.1
Government Stop, Service Emp.		N/A
Transportation Stop, Total Emp.		N/A
Transportation Stop, Service Emp.		N/A
Transportation Stop, Manufacturing Emp.		N/A
Office Stop, Total Emp.		N/A
Office Stop, Agri/Mining Emp.	2.350	23.3
Office Stop, Construction Emp.		N/A
Office Stop, Manufacturing Emp.		N/A
Office Stop, Households	-2.549	-32.2
Manufacturing Stop, Manufacturing Emp.		N/A
Manufacturing Stop, Total Emp.	-3.346	-92.6
Manufacturing Stop, Warehouse Emp.	-0.445	-7.1
Service Stop, Total Emp.		N/A
Service Stop, Retail Emp.		N/A
Number of Observations		**679,772**
Log Likelihood at Convergence		**-576,939.0**
Log Likelihood at Zero		**-1,630,022.1**
Rho Squared		**0.646**

7.4.7 Stop Time-of-Day Choice Model

The stop time-of-day choice model predicts the time period of each stop on a tour. Two separate models were estimated for time-of-day choice. The first is used for the departure time of a tour's first trip, and the second is used for subsequent trips. The reason for defining two separate models is that subsequent trip departure times should depend, in part, on the timing of a tour's previous trips. Thus, duration between trips is an important variable in the subsequent time-of-day choice model. Both models are MNL models, where the alternatives include each one-hour period of the day (24 alternatives in total). In application, the one-hour periods are aggregated back to the four existing time periods used in the regional model – AM peak, midday, PM peak and night.

There are two main reasons for defining the alternatives as one-hour periods rather than the four existing time periods used in MAG's regional model. First, this ensures alternatives are of uniform size, and no special considerations are needed to adjust variables for the size. Second, the more refined time period definitions should allow travel time between stops to have important implications. In addition, it allows availability restrictions to be well defined for the subsequent stop period model. Since truckers involved in interstate commerce are not allowed to work more than 12 consecutive hours by law, availability restrictions are important.

The following variables were found to be significant in the time-of-day choice models shown in Table 7-14 and Table 7-15:

- Starting land use for the tour;
- Previous and current stop purpose;
- Number of stops on tour by purpose;
- Travel distance/time from previous stop to current stop; and
- Previous stop time-of-day (if not first stop).

Table 7-14 Time-of-Day Choice Model - First Trip - Medium Trucks

TOD	Segment	Variable	Coefficient	t-statistic
AM	All	Constant	3.407	85.9
AM	Tour Purpose – Agri/Mining	Constant	-0.065	-1.2
AM	Tour Purpose – Households	Constant	0.232	10.9
AM	Tour Purpose – Government	Constant	0.088	2.8
AM	Tour Purpose – Transportation	Constant	-0.570	-9.5
AM	Tour Purpose – Office	Constant	-0.165	-4.7
AM	Tour Purpose – Manufacturing	Constant	-0.113	-2.5
AM	Tour Purpose – Service	Constant	-0.007	-0.3
AM	Stop Purpose – Households	Constant	-0.145	-8.9
AM	Stop Purpose – Manufacturing	Constant	-0.034	-1.1
AM	All	Number of Stops	0.101	39.5
AM	All	Tour Completed (0/1)	-0.155	-4.9
AM	Tour Purpose – Manufacturing	Tour Completed (0/1)	-0.172	-2.6
AM	All	Log(1 + Pk Travel Time to First	0.074	7.7
AM	All	AM Shift Variable	-0.001	-0.2
AM	Tour Purpose – Manufacturing	AM Shift Variable	-0.027	-1.4
MD	All	Constant	3.830	97.5
MD	Tour Purpose – Households	Constant	-0.244	-11.8
MD	Tour Purpose – Government	Constant	0.394	14.2
MD	Tour Purpose – Transportation	Constant	-0.450	-7.2
MD	Tour Purpose – Manufacturing	Constant	0.201	4.4
MD	Stop Purpose – Agri/Mining	Constant	-0.100	-2.1
MD	Stop Purpose – Manufacturing	Constant	-0.209	-6.4
MD	All	Number of Stops	-0.049	-17.0
MD	Tour Purpose – Manufacturing	Number of Stops	-0.052	-8.1
MD	All	Tour Completed (0/1)	0.257	8.6
MD	All	Log(1 + Pk Travel Time to First	0.052	5.2
MD	All	MD Shift Variable	-0.197	-52.2
PM	All	Constant	4.231	94.2
PM	Tour Purpose – Agri/Mining	Constant	-0.028	-0.4
PM	Tour Purpose – Households	Constant	-0.306	-12.1
PM	Tour Purpose – Transportation	Constant	-0.279	-3.4
PM	Tour Purpose – Office	Constant	0.139	3.3
PM	Stop Purpose – Agri/Mining	Constant	-0.004	-0.1
PM	Stop Purpose – Households	Constant	0.220	9.6
PM	Stop Purpose – Manufacturing	Constant	-0.236	-5.6
PM	All	Number of Stops	-0.360	-71.4
PM	All	Tour Completed (0/1)	0.041	1.1
PM	All	Log(1 + Pk Travel Time to First	0.005	0.4
PM	All	PM Shift Variable	-0.219	-28.5
PM	Tour Purpose – Manufacturing	PM Shift Variable	0.014	0.6
NT	All	NT Shift Variable Early	0.620	108.2
NT	Tour Purpose – Manufacturing	NT Shift Variable Early	0.090	9.1
NT	All	NT Shift Variable Late	0.412	62.4
NT	Tour Purpose – Manufacturing	NT Shift Variable Late	0.010	0.7
Number of Observations				**121,676**
Log Likelihood at Convergence				**-316938.5**
Log Likelihood Constants Only				**-326948.6**
Rho Squared				**0.031**

Table 7-15 Time-of-Day Choice Model - Subsequent Trips - Medium Trucks

Variable	Coefficient	t-statistic
Constant – Same 1-hour Period as Previous Stop	-2.182	-7.5
Constant – First 1-hour Period after Previous Stop	-0.714	-2.7
Constant – Second 1-hour Period after Previous Stop	-0.668	-2.9
Constant – Third 1-hour Period after Previous Stop	-0.409	-2.0
Constant – Fourth 1-hr Period after Previous Stop	-0.108	-0.6
Constant – Fifth or Sixth 1-hr Period after Previous Stop	-0.411	-2.8
Constant – Seventh or Eighth 1-hr Period after Previous Stop	-0.279	-2.6
Departure Time Shift Variable [63]	-0.821	-27.6
Departure Shift – Prev. Stop = Agri/Mining	0.004	0.3
Departure Shift – Prev. Stop = Households	0.025	6.3
Departure Shift – Prev. Stop = Government	0.052	8.2
Departure Shift – Prev. Stop = Office	0.075	9.6
Departure Shift – Prev. Stop = Manufacturing	0.139	22.0
Departure Shift – Prev. Stop = Service	0.013	2.6
Departure Shift * Log (1 + Peak Travel Time to Previous Stop)	0.057	31.6
Departure Shift * Total Number of Stops on Tour	-0.064	-101.3
Departure Shift * Total Number of Stops Simulated	-0.002	-2.6
Number of Observations		**368,072**
Log Likelihood at Convergence		**-466015.3**
Log Likelihood Constants Only		**-751610.2**
Rho Squared		**0.380**

7.4.8 Comparison to Survey

The medium truck model components were also coded and implemented in R, and each component was individually assessed and compared to observed survey data. The reasonability of the explanatory variables were determined by their magnitude, t-statistic, and their relation to the dependent variable. Some of the key findings include:

- Construction and agriculture/mining tours make more stops per tour than other tour purposes;
- Tours with only one stop are most frequently closed tours, and close tour frequency falls with almost every additional stop, reaching its lowest at 11 stops;
- Stop purpose is often strongly influenced by the tour type or the first land use of the truck origin;

[63] This variable is equal to the number of periods the alternative is after the departure period of the previous stop.

- Travel time has a negative effect on location choice utility, and is more pronounced as the number of stops on tour increase; and
- Previous stop purpose has a significant impact on the time period of the next stop.

The individual model outputs also were compared against the Streetlight data to assess the model performance. These results are depicted in Figure 7-19 through Figure 7-22. These comparisons indicate that the model components are predicting very closely to the observed data.

Figure 7-19 Stop Frequency Comparison - Tour Model Outputs versus Survey Data – Medium Trucks

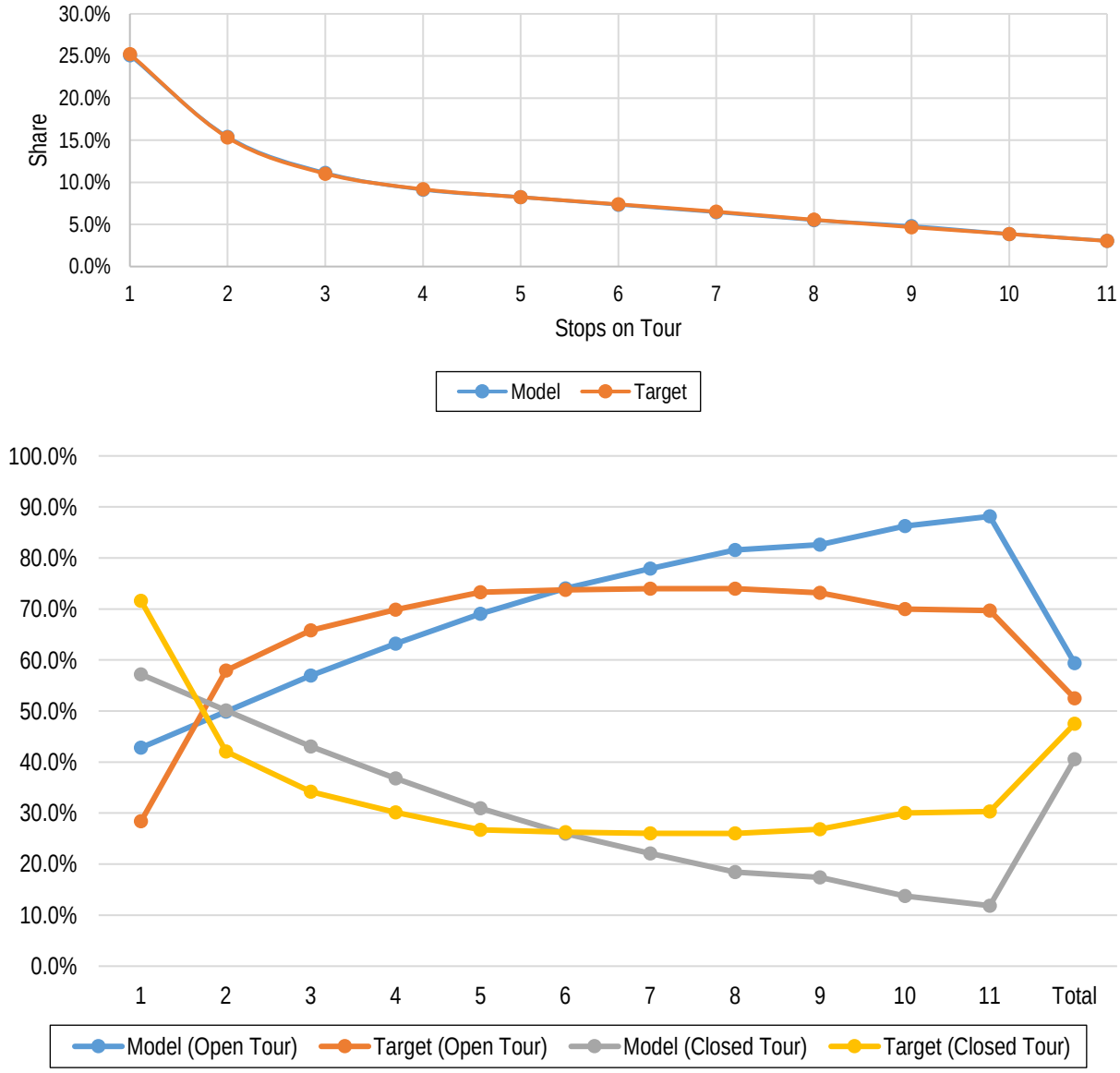


Figure 7-20 Stop Purpose Frequency – Model Outputs versus Survey Data – Medium Trucks

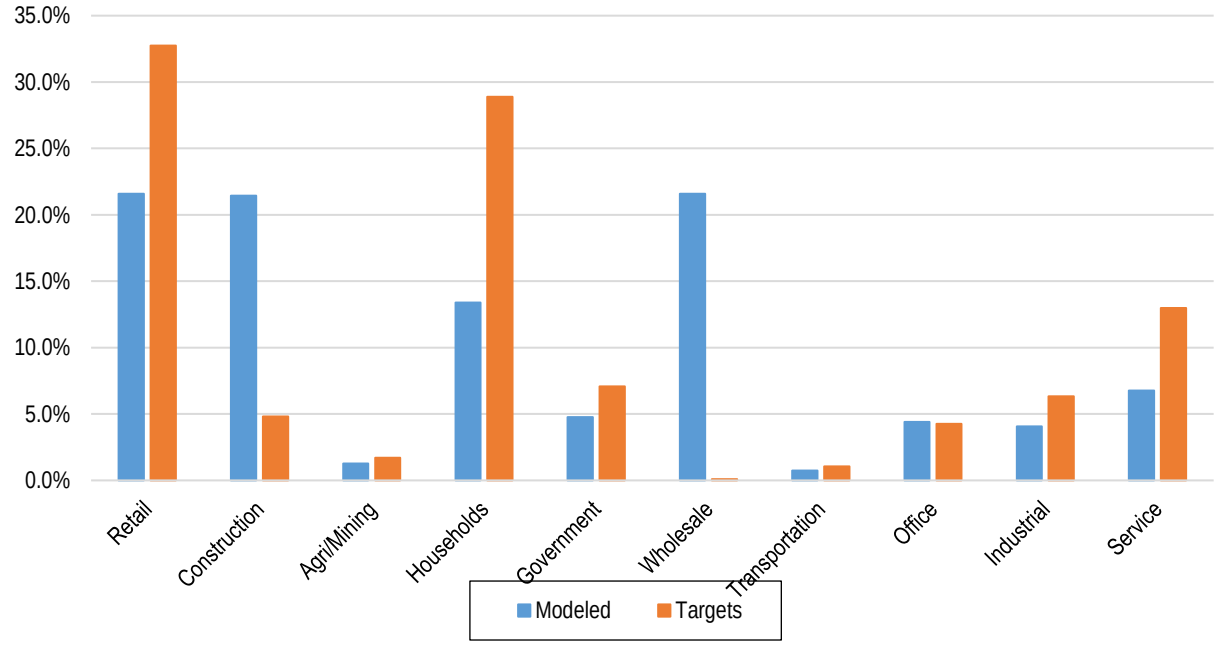


Figure 7-21 Time-of-Day of First Stop - Model Outputs versus Survey Data – Medium Trucks

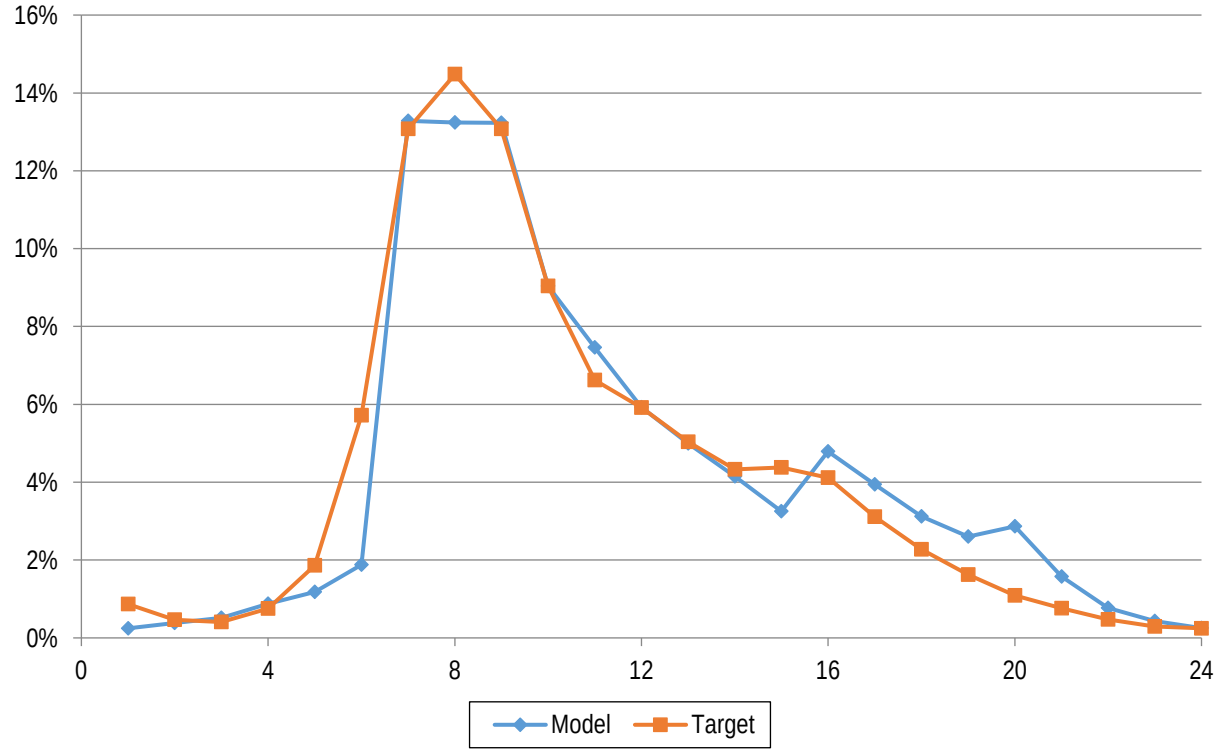
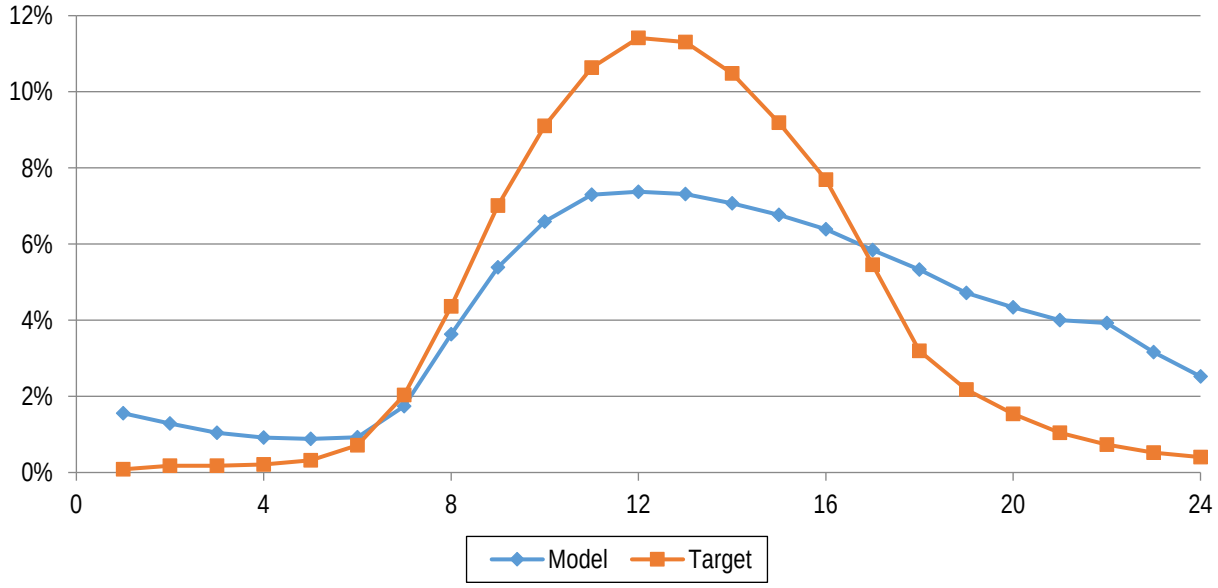


Figure 7-22 Time-of-Day of Subsequent Stops - Model Outputs versus Survey Data – Medium Trucks



7.5 Heavy Trucks

7.5.1 Model Estimation

The ATRI truck GPS database was used to develop heavy truck tour-based models. The processed ATRI GPS data yielded a truck tour database that included data from 39,080 trucks that engaged in 81,090 tours and 408,679 trips. This dataset was narrowed down further for model estimation for a number of reasons. Only tours made on weekdays were used. Only tours that made 11 or fewer stops were considered, which was used for consistency with the previous version of the truck tour model. In addition, only tours that started and remained in the MAG region were considered. External tours could not be considered since land use and travel time information was not available. PAG tours were not used because land use information was not consistent between MAG and PAG regions, making it impossible to code tour and stop purposes in consistent ways across the regions. After incorporating these restrictions, 63,355 tours remained, which are summarized by purpose in Table 7-16.

Table 7-16 Summary of Medium Truck Trips

Stop Purpose	Trips	Share
Retail (base/reference purpose)	3,030	4.8%
Construction	0	0.0%
Farming	1,338	2.1%
Mining	0	0.0%
Households	2,820	4.5%
Government	200	0.3%
Warehousing	0	0.0%
Transportation	179	0.3%
Other	6,465	10.2%
Manufacturing	35,062	55.3%
Service	14,261	22.5%
Total	63,355	100.0%

One major assumption was that "tour purpose" was defined by the land use in the truck's starting location for the tour. That is, if a tour has a "retail" land use as the starting origin, then the whole tour is classified as a "retail tour". This is the best that could be done because the GPS data does not divulge the industry type of each truck. Upon further examining the data, it was found that about half of the truck tours are incomplete, that is, the trucks do not return to their home base or the starting origin. About 52 percent of truck tours completed the tour and returned to starting origin. Heavy truck tours are modeled through a sequence of models that predict tour generation at the zonal level by tour purpose (i.e., starting land use type), the number of stops for each tour, the purpose of those stops, the location of stops, and the time-of-day for stops.

One notable difference between the updated model and the previous version of the model is that the construction, warehousing, and office activity purposes no longer appear as distinct purposes in the updated model. This was done because it was decided to use the land uses of zones as the sole determinant of activity purpose. Moreover, the sample sizes were very low for these tour purposes, which led us to not use these specific purposes distinctly in the model. Warehousing stops are now grouped with manufacturing stops, office stops are now included with the other stop category, and construction stops are included with farming stop category.

7.5.2 Tour Generation

The tour generation model estimates the number of tours generated in each zone by truck tour purpose. Truck tour purpose is defined as the starting land use type of the tour. Using a combination of existing heavy truck trip rates, tour completion percentage and average stops per tour, the tour rates were computed by tour purpose. These rates are multiplied by the appropriate employment variable for each tour purpose to produce number of tours.

7.5.3 Stop Frequency Model

The stop frequency model predicts the number of stops on each truck tour. This is a multinomial logit (MNL) model where the number of stops represents the choice alternatives. The choices were limited to 11 stops for consistency with the previous version of the truck tour model.

The model estimation results are presented in Table 7-17. The key variables that were found to be significant in explaining stop frequency were the starting land use of the tour and zonal land use variables. The zonal variables that influence stop making behavior are employment by type and households at the starting zone of the tour.

Table 7-17 Stop Frequency Model - Heavy Trucks

Variables		Coefficient	t-statistic
Constants	2 Stops	-0.511	-14.2
	3 Stops	-0.747	-15.3
	4 Stops	-0.934	-17.3
	5 Stops	-1.077	-19.2
	6 Stops	-1.174	-20.9
	7 Stops	-1.197	-21.6
	8 Stops	-1.174	-21.3
	9 Stops	-1.261	-22.6
	10 Stops	-1.288	-22.3
	11 Stops	-1.413	-22.9
Number of Stops – Farming		0.076	2.2
Log (No. of Stops) – Farming		-0.850	-6.7
Number of Stops – Households		0.077	2.9
Log (No. of Stops) – Households		-0.887	-8.8
Number of Stops – Transportation		0.222	2.9
Log (No. of Stops) – Transportation		-0.956	-3.2
Number of Stops – Other		-0.116	-5.2
Log (No. of Stops) – Other		-0.134	-1.6
Number of Stops – Manufacturing		0.076	4.1
Log (No. of Stops) – Manufacturing		-0.851	-11.6
Number of Stops – Service		0.038	1.9
Log (No. of Stops) – Service		-0.739	-9.6
(Total Emp. Start Zone) / (No. of Stops)		-0.00004	-4.2
(Total HHs Start Zone) / (No. of Stops)		-0.00015	-7.5
Number of Observations	**63,355**		
Log Likelihood at Convergence	**-125534.6**		
Log Likelihood Constants Only	**-126092.0**		
Rho Squared	**0.004**		

Another key variable was the accessibility index that is expressed as a logarithmic function of travel time between two zones and employment at the destination zone.

$$Accessibility\ Index\ (AI) = \ln\left(1 + Sum_j\left[\left(\exp\left(-0.05 * TT_{ij}\right) / 10000\right) * EMP_j\right]\right)$$

Where,

TT_{ij} is travel time between i and j; and

EMP_j is employment at the destination zone j.

7.5.4 Tour Completion Model

The tour completion model predicts whether the tour returns to its starting location or ends at another location. This is a binomial logit choice model with two alternatives: tour does not complete or tour completes. These results are shown in Table 7-18.

Table 7-18 Tour Completion Model - Heavy Trucks

Variable	Coefficient	t-statistic
Constant	-0.103	-1.4
Tour Purpose = Farming	-1.513	-9.3
Tour Purpose = Households	1.079	15.7
Tour Purpose = Government	-1.021	-3.5
Tour Purpose = Transportation	-2.621	-5.1
Tour Purpose = Other	-1.424	-14.9
Tour Purpose = Manufacturing	0.724	9.0
Tour Purpose = Service	0.775	10.3
Number of Stops on Tour	-0.451	-26.1
No. of Stops – Farming	0.111	2.6
No. of Stops – Other	0.208	9.3
No. of Stops – Manufacturing	0.253	14.3
No. of Stops – Service	0.073	3.7
Total Employment in Start Zone	0.00004	3.4
Total Emp. – Farming	0.00108	14.3
Total Emp. – Other	0.00063	25.1
Total Emp. – Manufacturing	0.00011	6.7
Number of Observations		**63,355**
Log Likelihood at Convergence		**-36895.4**
Log Likelihood Constants Only		**-43836.9**
Rho Squared		**0.158**

The tour purpose and the number of stops on the tour make a significant impact on the tour completion probability. The greater the number of stops on the tour, the less the likelihood of a tour being completed. Manufacturing tours are more prone to completing the tour while farming and service trucks are less likely to completing the tour. The land use variables like employment and accessibility indices do influence the completion of the tour as they do the stop making behavior.

7.5.5 Stop Purpose Model

The stop purpose model predicts the purpose (i.e., land use type) of each stop that is predicted by the stop frequency model. This is a MNL model that predicts purpose of the stops in sequence, that is, from the first stop to the last stop. The alternatives or choices used in this model are the same land use types as defined in the trip-based truck model. The following alternative choice set is considered:

- Retail;
- Farming;
- Households;
- Government;
- Transportation;
- Other;
- Manufacturing; and
- Service.

The estimated model is presented in Table 7-19. All the coefficients are segmented by tour purpose. This influences the type of stop purpose significantly. The starting land use of the tour influences the stop purpose of subsequent stops on the tour. Other key explanatory variables that were found to be significant in this model are previous stop purpose, where certain purpose to purpose interchanges are much more prevalent than others, and the number of previous stops by purpose, which includes the total number of stops of each type already simulated for the tour. The accessibility indices that are segmented by tour purpose also were found to be significant in explaining the stop purpose. The zonal land use variables, including employment by type and households at the starting zone also influence the stop purpose.

Table 7-19 Stop Purpose Model - Heavy Trucks

Variable	Alternative	Coefficient	t-statistic
Constants	Farming	-2.647	-22.5
	Households	-4.089	-32.1
	Government	-8.935	-26.3
	Transportation	-5.200	-14.4
	Other	-1.631	-20.3
	Manufacturing	1.868	52.9
	Service	0.051	1.0
Previous Stop Purpose = Retail	Farming	-1.279	-14.0
	Transportation	-2.409	-10.4
	Other	-1.600	-37.7
	Manufacturing	-1.872	-67.8
	Service	-0.819	-28.5
Previous Stop Purpose = Farming	Farming	2.872	25.7
	Households	2.154	19.9
	Government	1.703	6.3
	Other	1.134	11.2

Variable	Alternative	Coefficient	t-statistic
	Manufacturing	0.668	7.3
	Service	1.070	11.3
Previous Stop Purpose = Households	Households	1.983	51.3
	Government	1.496	12.7
	Transportation	-1.123	-3.9
	Manufacturing	-1.087	-34.1
Previous Stop Purpose = Government	Households	1.468	15.0
	Government	3.235	23.2
	Manufacturing	-0.676	-8.7
	Service	0.264	3.5
Previous Stop Purpose = Transportation	Households	0.850	3.3
	Government	1.579	2.7
	Transportation	4.155	25.8
	Service	0.296	2.6
Previous Stop Purpose = Other	Farming	0.909	12.0
	Households	1.624	36.6
	Government	1.888	15.7
	Other	1.520	45.2
	Service	0.625	20.9
Previous Stop Purpose = Manufacturing	Farming	0.820	12.5
	Households	0.805	21.8
	Government	1.195	11.5
	Other	0.115	3.3
	Manufacturing	0.524	20.3
	Service	0.301	10.5
Previous Stop Purpose = Service	Farming	0.172	2.5
	Households	0.932	27.0
	Government	1.296	13.1
	Transportation	-0.836	-6.2
	Other	-0.153	-4.4
	Manufacturing	-0.657	-24.7
	Service	0.526	18.9
Tour Purpose = Retail	Farming	0.324	2.8
	Households	1.019	12.8
	Government	0.781	3.2
	Other	0.613	10.4
	Manufacturing	0.152	4.1
	Service	0.329	8.1
Tour Purpose = Farming	Farming	1.774	14.0
	Households	1.034	7.7
	Government	1.051	3.1
	Other	0.998	9.1
	Manufacturing	0.729	7.7
	Service	0.556	5.6
Tour Purpose = Households	Farming	0.526	4.5
	Households	1.542	18.3
	Government	1.526	6.3
	Other	0.873	13.4

Variable	Alternative	Coefficient	t-statistic
	Manufacturing	0.566	11.5
	Service	0.666	13.1
Tour Purpose = Government	Farming	1.518	8.2
	Households	1.831	12.8
	Government	2.669	10.1
	Other	1.640	12.5
	Manufacturing	0.744	6.6
	Service	0.978	8.6
Tour Purpose = Transportation	Transportation	-0.611	-2.0
Tour Purpose = Other	Farming	1.153	12.9
	Households	1.380	17.9
	Government	1.621	7.2
	Other	1.303	25.6
	Manufacturing	0.702	18.9
	Service	0.780	19.0
Tour Purpose = Manufacturing	Farming	0.635	7.4
	Households	0.863	11.3
	Government	1.121	5.0
	Transportation	0.467	3.2
	Other	0.857	17.6
	Manufacturing	0.656	20.9
	Service	0.620	16.8
Tour Purpose = Service	Farming	1.072	12.0
	Households	1.254	16.4
	Government	1.494	6.6
	Transportation	0.512	2.9
	Other	1.057	20.9
	Manufacturing	0.657	20.1
	Service	0.740	19.8
Number of Previous Retail Stops	Farming	-0.522	-15.7
	Households	-0.483	-25.6
	Government	-0.532	-11.8
	Transportation	-0.188	-3.4
	Other	-0.470	-28.4
	Manufacturing	-0.202	-24.6
	Service	-0.278	-32.2
Number of Previous Farming Stops	Farming	1.022	21.7
	Households	0.280	5.4
	Government	0.282	3.0
	Other	0.486	10.4
	Manufacturing	0.377	8.7
	Service	0.455	10.3
Number of Previous Households Stops	Households	0.263	13.6
	Other	0.095	5.2
	Manufacturing	0.060	4.2
	Service	0.063	4.6
Number of Previous Government Stops	Households	-0.203	-5.6
	Government	0.176	3.5

Variable	Alternative	Coefficient	t-statistic
	Other	-0.297	-9.1
	Manufacturing	0.097	4.7
	Service	-0.120	-5.1
Number of Previous Transportation Stops	Transportation	0.831	6.7
	Manufacturing	0.245	4.6
	Service	0.294	5.2
Number of Previous Other Stops	Households	-0.191	-10.2
	Government	-0.231	-5.7
	Transportation	-0.196	-3.0
	Other	0.083	6.6
	Service	-0.057	-6.9
Number of Previous Manufacturing Stops	Farming	-0.094	-5.9
	Households	-0.192	-11.9
	Government	-0.279	-8.4
	Other	-0.031	-2.8
	Manufacturing	0.278	55.3
Number of Previous Service Stops	Farming	-0.209	-10.1
	Households	-0.159	-10.1
	Government	-0.285	-8.0
	Transportation	-0.174	-3.2
	Other	-0.127	-10.5
	Service	0.166	27.3
Total Stops on Tour	Farming	-0.107	-18.6
	Households	-0.087	-20.9
	Government	0.066	5.7
	Transportation	-0.073	-4.2
	Other	-0.107	-29.3
	Manufacturing	-0.119	-44.7
	Service	-0.091	-31.6
Log (1 + Stop Sequence Number)	Farming	0.748	12.7
	Households	1.687	18.0
	Government	2.294	11.7
	Transportation	1.267	6.2
	Other	1.075	17.8
	Service	0.495	16.3
(AI to Farming Emp.) / (Stop Seq. No.)	Farming	3.304	6.9
(AI to Households) / (Stop Seq. No.)	Households	0.617	16.8
(AI to Total Emp.) / (Stop Seq. No.)	Government	0.854	10.7
(AI to Total Emp.) / (Stop Seq. No.)	Transportation	0.525	4.8
(AI to Total Emp.) / (Stop Seq. No.)	Other	0.354	15.7
(AI to Manufacturing Emp.) / (Stop Seq. No.)	Manufacturing	0.188	9.2
(AI to Service Emp.) / (Stop Seq. No.)	Service	0.251	14.1
Number of Observations			**276,532**
Log Likelihood at Convergence			**-330681.3**
Log Likelihood Constants Only			**-383907.3**
Rho Squared			**0.139**

7.5.6 Stop Location Choice Model

The stop location choice model predicts the location of each stop simulated for the tour, and is similar in design to a destination choice model employed for distributing passenger trips. Every zone in the region is a potential choice for this model. Similar to any other destination choice model, size variables are included in the model. These include employment at the stop location by type.

Two types of accessibility variables are included in the model:

1. Direct zone-to-zone accessibility variables or travel time between: a) previous stop location to current stop location; and b) first stop location to current stop location; and
2. Aggregate accessibility measures. This is important to describe the accessibility of a stop zone to employment types corresponding to the next stop purpose.

Other variables include zonal area type such as CBD, rural, and suburban. This is defined as a combination of employment and population density. Most of these variables are segmented by the starting land use of the tour, previous stop purpose, current stop purpose, and number of stops on tour by purpose. These results are shown in Table 7-20.

Table 7-20 Stop Location Choice Model - Heavy Trucks

Utility Variables	Coefficient	t-statistic
Intrazonal with Previous Stop	1.733	50.6
Intrazonal with Tour Start Zone	4.788	108.6
Intrazonal with Previous Stop and Tour Start Zone	-4.062	-73.6
Intrazonal with Previous Stop – Stop Purpose = Manufacturing	1.329	19.4
Intrazonal with Tour Start Zone – Stop Purpose = Manufacturing	0.275	5.6
Intrazonal with Previous Stop – Prev. Purpose = Manufacturing	-0.325	-5.0
Log(1 + Pk Travel Time)	-2.023	-132.1
Log(1 + Pk Travel Time) – First Stop on Tour, Total Stops on Tour >= 2	0.120	8.6
Log(1 + Pk Travel Time) * Total Stops on Tour	-0.046	-30.8
Log(1 + Pk Return Travel Time) / (Stops Remaining + 1) – Total Stops on	-0.666	-63.5
Log(1 + Pk Travel Time) – Stop Purpose = Farming	0.341	13.9
Log(1 + Pk Travel Time) – Stop Purpose = Households	-0.635	-30.6
Log(1 + Pk Travel Time) – Stop Purpose = Transportation	0.401	4.1
Log(1 + Pk Travel Time) – Stop Purpose = Manufacturing	0.438	42.0

Size Variables	Coefficient	t-statistic
Retail Stop, Retail Emp.	0.000	N/A
Retail Stop, Service Emp.	-4.910	-41.1
Farming Stop, Farming Emp.	0.000	N/A
Farming Stop, Total Emp.	-3.825	-63.7
Farming Stop, Construction Emp.	-4.548	-9.3
Households Stop, Households	0.000	N/A
Government Stop, Total Emp.	0.000	N/A
Government Stop, Retail Emp.	0.880	0.6
Government Stop, Farming Emp.	3.655	3.8
Government Stop, Service Emp.	2.767	3.8
Transportation Stop, Total Emp.	0.000	N/A
Transportation Stop, Service Emp.	0.079	0.2
Transportation Stop, Manufacturing Emp.	1.631	3.8
Other Stop, Total Emp.	0.000	N/A
Other Stop, Farming Emp.	2.730	32.5
Other Stop, Construction Emp.	0.669	6.5
Other Stop, Manufacturing Emp.	0.016	0.1
Other Stop, Households	-1.456	-24.9
Manufacturing Stop, Manufacturing Emp.	0.000	N/A
Manufacturing Stop, Total Emp.	-7.484	-31.0
Manufacturing Stop, Warehouse Emp.	1.377	44.0
Service Stop, Total Emp.	0.000	N/A
Service Stop, Retail Emp.	1.662	63.7
Number of Observations		**305,577**
Log Likelihood at Convergence		**-426586.4**
Log Likelihood at Zero		**-1201475.5**
Rho Squared		**0.645**

7.5.7 Stop Time-of-Day Choice Model

The stop time-of-day choice model predicts the time period of each stop on a tour. Two separate models were estimated for time-of-day choice. The first is used for the departure time of a tour's first trip, and the second is used for subsequent trips. The reason for defining two separate models is that subsequent trip departure times should depend, in part, on the timing of a tour's previous trips. Thus, duration between trips is an important variable in the subsequent time-of-day choice model. Both models are MNL models, where the alternatives include each one-hour period of the day (24 alternatives in total). In application, the one-hour periods are aggregated back to the four existing time periods used in the regional model – AM peak, midday, PM peak and night.

There are two main reasons for defining the alternatives as one-hour periods rather than the four existing time periods used in MAG's regional model. First, this ensures alternatives are of uniform size, and no special considerations are needed to adjust variables for the size. Second, the more refined time period definitions should allow travel time between stops to have important implications. In addition, it allows availability restrictions to be well defined for the subsequent stop period model. Since truckers involved in interstate commerce are not allowed to work more than 12 consecutive hours by law, availability restrictions are important.

The following variables were found to be significant in the time-of-day choice models shown in Table 7-21 and Table 7-22:

- Starting land use for the tour;
- Previous and current stop purpose;
- Number of stops on tour by purpose;
- Travel distance/time from previous stop to current stop; and
- Previous stop time-of-day (if not first stop).

Table 7-21 Time-of-Day Choice Model – First Trip – Heavy Trucks

TOD Period	Segment	Variable	Coefficient	t-statistic
AM	All	Constant	0.725	10.4
AM	Tour Purpose – Farming	Constant	0.282	3.1
AM	Tour Purpose – Households	Constant	0.247	3.2
AM	Tour Purpose – Government	Constant	0.875	6.0
AM	Tour Purpose – Transportation	Constant	-0.527	-2.4
AM	Tour Purpose – Other	Constant	0.306	4.9
AM	Tour Purpose – Manufacturing	Constant	0.769	13.0
AM	Tour Purpose – Service	Constant	0.182	3.4
AM	Stop Purpose – Households	Constant	0.306	5.4
AM	Stop Purpose – Manufacturing	Constant	0.379	13.5
AM	All	Number of Stops	0.017	5.4
AM	All	Tour Completed (0/1)	-0.320	-6.5
AM	Tour Purpose – Manufacturing	Tour Completed (0/1)	-0.327	-6.0
AM	All	Log(1 + Pk Travel Time to	-0.021	-1.6
AM	All	AM Shift Variable	-0.188	-10.2
AM	Tour Purpose – Manufacturing	AM Shift Variable	0.089	3.7
MD	All	Constant	1.805	32.4
MD	Tour Purpose – Households	Constant	0.085	1.3
MD	Tour Purpose – Government	Constant	0.231	1.3
MD	Tour Purpose – Transportation	Constant	-1.209	-4.3
MD	Tour Purpose – Manufacturing	Constant	0.267	6.6
MD	Stop Purpose – Farming	Constant	-0.205	-2.4
MD	Stop Purpose – Manufacturing	Constant	0.188	6.9
MD	All	Number of Stops	-0.186	-34.7
MD	Tour Purpose – Manufacturing	Number of Stops	0.085	13.4
MD	All	Tour Completed (0/1)	-0.481	-17.2
MD	All	Log(1 + Pk Travel Time to	-0.128	-9.8
MD	All	MD Shift Variable	-0.086	-13.2
PM	All	Constant	2.486	53.8
PM	Tour Purpose – Farming	Constant	0.319	4.6
PM	Tour Purpose – Households	Constant	0.703	12.9
PM	Tour Purpose – Transportation	Constant	-0.375	-2.0
PM	Tour Purpose – Other	Constant	0.845	25.7
PM	Stop Purpose – Farming	Constant	-0.412	-5.3
PM	Stop Purpose – Households	Constant	-0.310	-5.9
PM	Stop Purpose – Manufacturing	Constant	0.082	3.4
PM	All	Number of Stops	-0.247	-79.4
PM	All	Tour Completed (0/1)	0.166	6.8
PM	All	Log(1 + Pk Travel Time to	-0.483	-39.2
PM	All	PM Shift Variable	0.597	73.8
PM	Tour Purpose – Manufacturing	PM Shift Variable	0.050	5.0
NT	All	NT Shift Variable Early	0.194	33.8
NT	Tour Purpose – Manufacturing	NT Shift Variable Early	0.094	12.9
NT	All	NT Shift Variable Late	-0.069	-8.7
NT	Tour Purpose – Manufacturing	NT Shift Variable Late	0.036	3.4
Number of Observations				**72,117**
Log Likelihood at Convergence				**-198053.0**
Log Likelihood Constants Only				**-206654.3**
Rho Squared				**0.042**

Table 7-22 Time-of-Day Choice Model – Subsequent Trips – Heavy Trucks

Variable	Coefficient	t-statistic
Constant – Same 1-hr Period as Previous Stop	-0.166	-0.7
Constant – First 1-hr Period after Previous Stop	1.345	6.2
Constant – Second 1-hr Period after Previous Stop	1.233	6.3
Constant – Third 1-hr Period after Previous Stop	1.100	6.3
Constant – Fourth 1-hr Period after Previous Stop	1.011	6.6
Constant – Fifth or Sixth 1-hr Period after Previous Stop	0.352	2.8
Constant – Seventh or Eighth 1-hr Period after Previous Stop	-0.202	-2.1
Departure Time Shift Variable [64]	-0.791	-32.7
Departure Shift – Prev. Stop = Farming	0.040	2.8
Departure Shift – Prev. Stop = Households	0.050	4.8
Departure Shift – Prev. Stop = Government	-0.069	-2.3
Departure Shift – Prev. Stop = Other	0.110	13.5
Departure Shift – Prev. Stop = Manufacturing	0.056	10.0
Departure Shift – Prev. Stop = Service	0.044	7.1
Departure Shift * Log (1 + Peak Travel Time to Previous Stop)	0.175	79.7
Departure Shift * Total Number of Stops on Tour	-0.049	-64.7
Departure Shift * Total Number of Stops Simulated	-0.067	-66.8
Number of Observations		**240,598**
Log Likelihood at Convergence		**-302932.1**
Log Likelihood Constants Only		**-583524.0**
Rho Squared		**0.481**

7.5.8 Comparison to Survey

All the tour model components were coded and implemented in R, and calibrated to match observed data. The reasonability of the explanatory variables were determined by their magnitude, t-statistic, and their relation to the dependent variable. Some of the key findings include:

- Retail, warehouse, and manufacturing tours make more stops than other tour purposes;
- Tours with only one stop are most frequently closed tours, and close tour frequency is at its lowest when the number of stops is about six;
- Stop purpose is often strongly influenced by the tour type or the first land use of the truck origin;

[64] This variable is equal to the number of periods the alternative is after the departure period of the previous stop.

- Travel time has a negative effect on location choice utility, and is more pronounced as the number of stops on tour increase; and
- Previous stop purpose has a significant impact on the time period of the next stop.

The individual model outputs also were compared against the truck GPS survey data to assess the model performance. These results are depicted in Figure 7-23 through Figure 7-28.

Figure 7-23 Stop Frequency Comparison – Tour Model Outputs versus Survey Data – Heavy Trucks

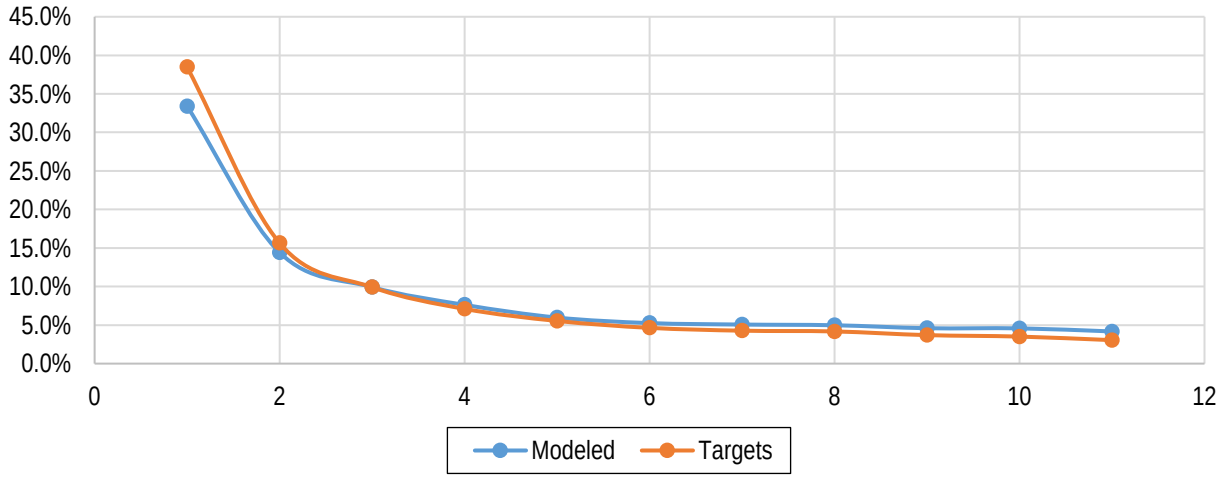


Figure 7-24 Tour Completion versus Stop Frequency - Model Outputs versus Survey Data - Heavy Trucks

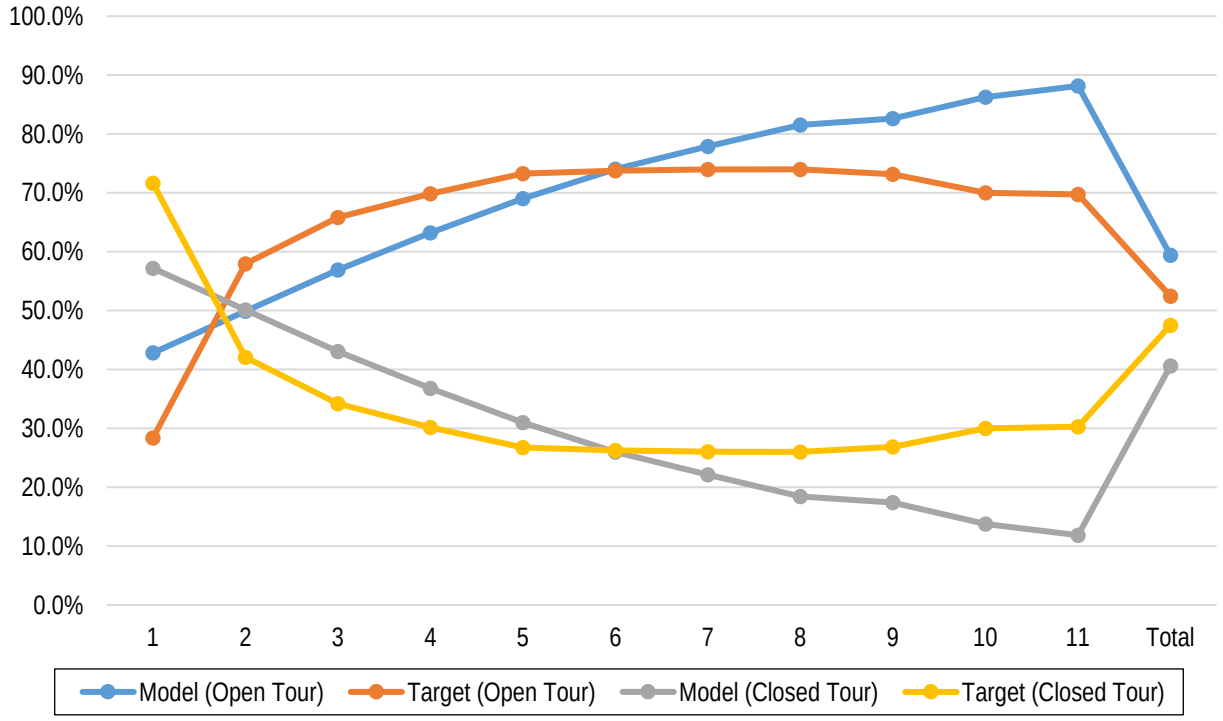


Figure 7-25 Stop Purpose Frequency - Model Outputs versus Survey Data - Heavy Trucks

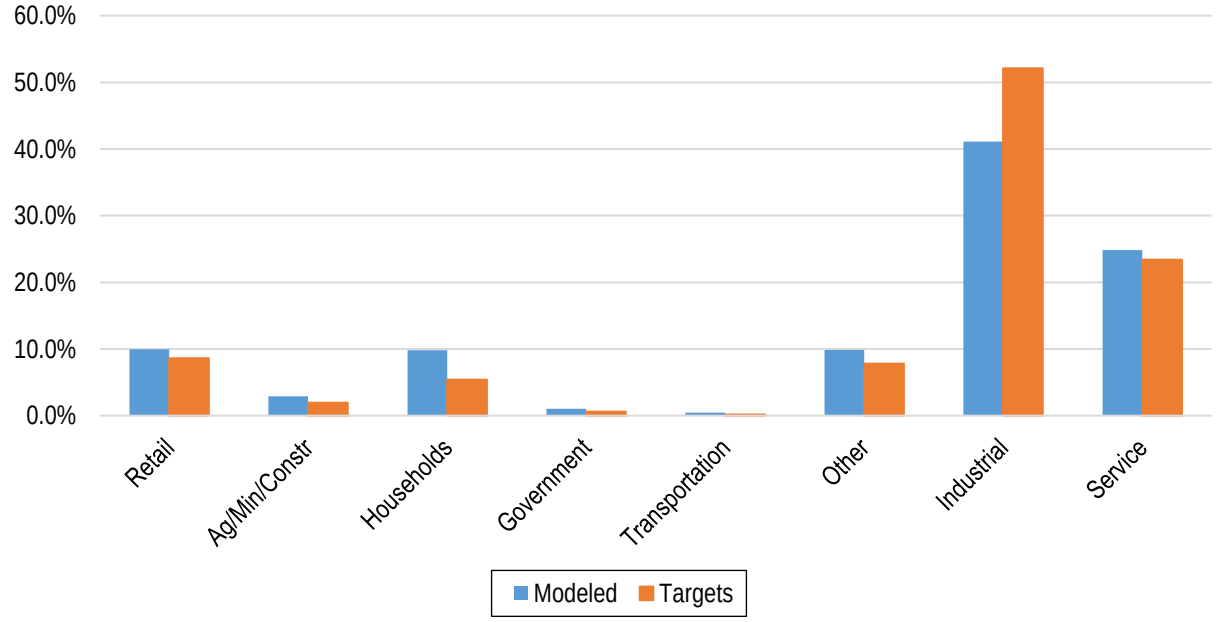


Figure 7-26 Time-of-Day of First Stop - Model Outputs versus Survey Data – Heavy Trucks

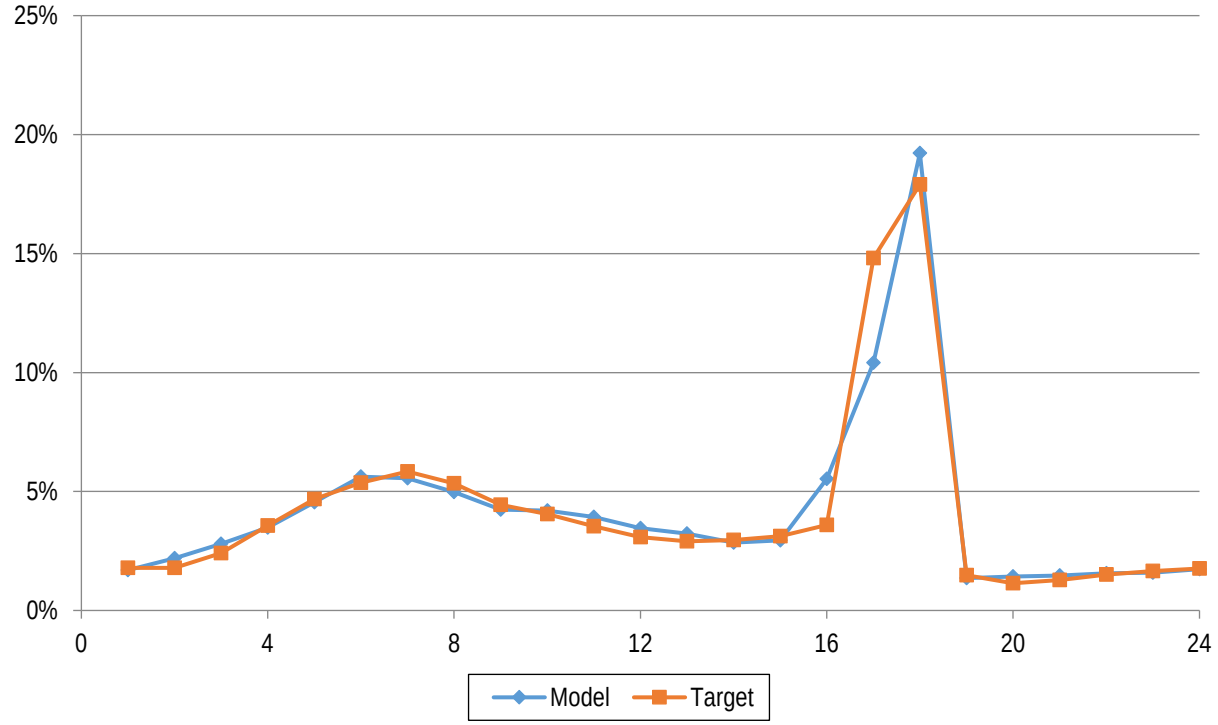


Figure 7-27 Time-of-Day of Subsequent Stops - Model Outputs versus Survey Data – Heavy Trucks

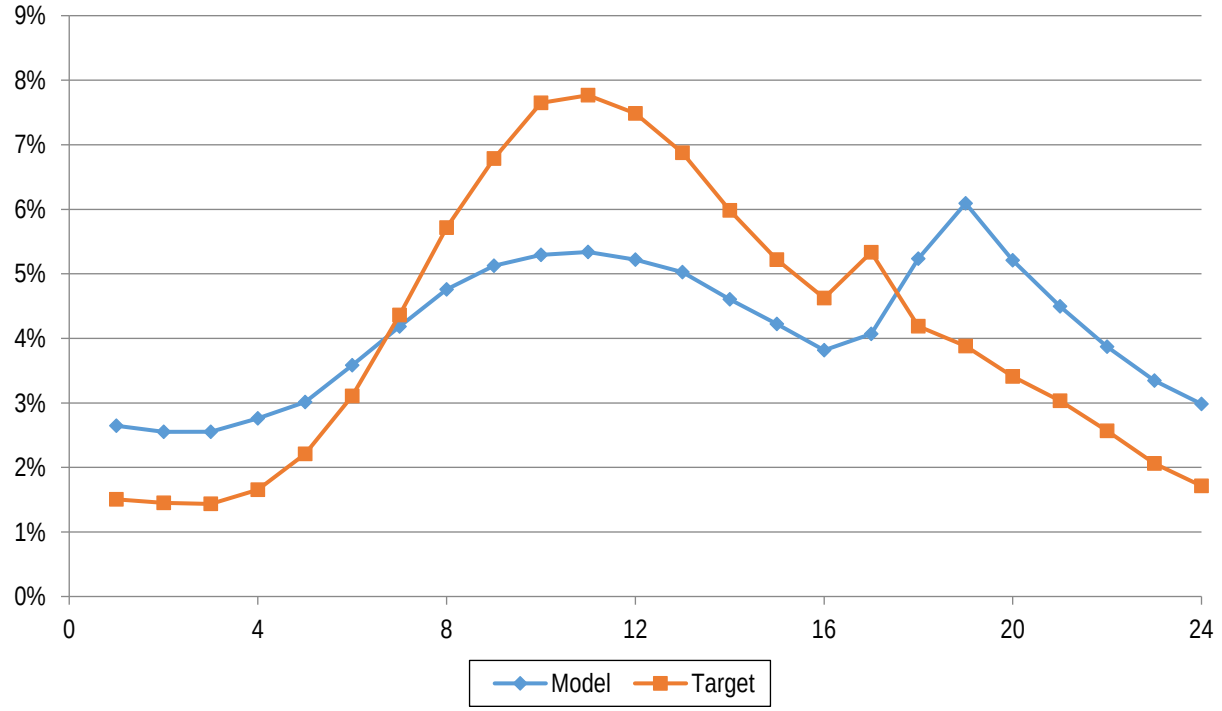
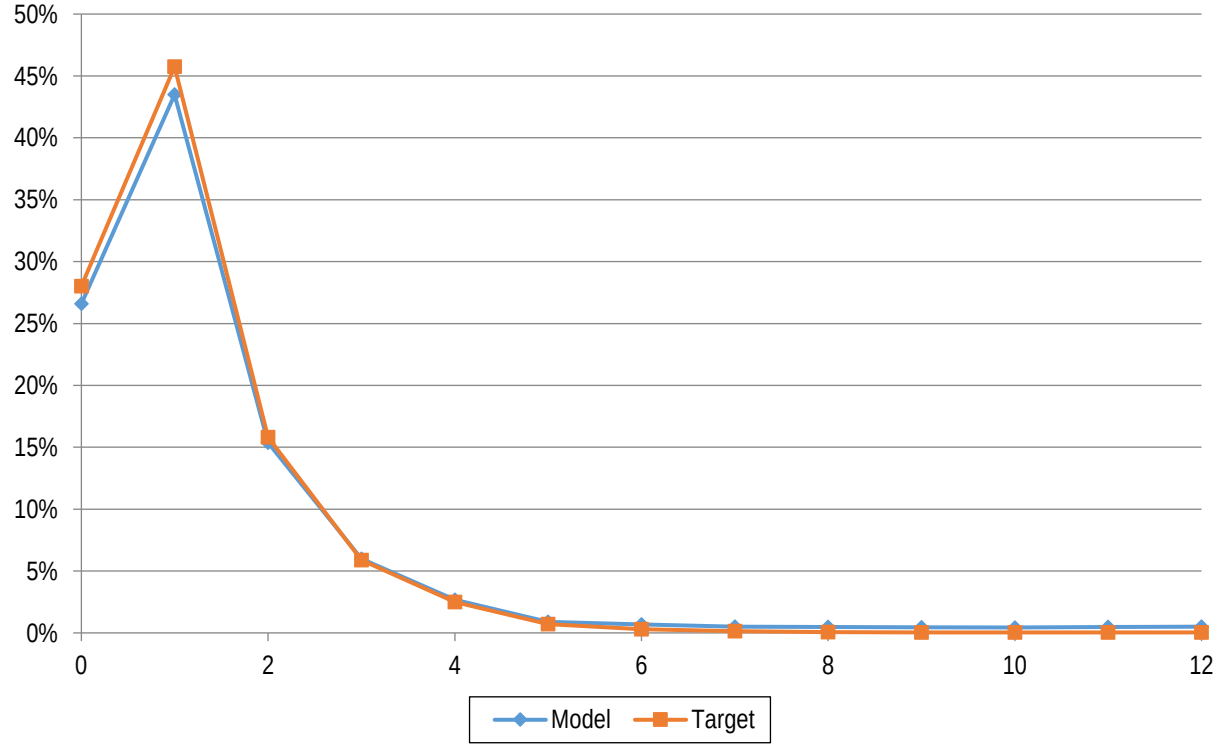


Figure 7-28 Duration of Subsequent Stops (in hours) - Model Outputs versus Survey Data – Heavy Trucks



8 Integration of Supply-Chain and Tour-Based Models

Consultant Technical Lead – Arun Kuppam (Cambridge Systematics) and Zahra Pourabdollahi (RS&H)

MAG Technical Lead – Kyunghwi Jeon, Haidong Zhu, and Daehyun You

The supply chain model outputs annual commodity flows in tons by commodity group that will need to be integrated with the rest of the model system. From a highway assignment standpoint, only the truck flows from the supply chain model are used while rail, water and other modes of freight are not assigned to any networks. The trucks flows are converted to daily truck tours and trips and integrated with the highway assignment model as described in this chapter.

8.1 Integration Methodology

The Supply Chain model uses the Freight Analysis Framework (FAF) database, and outputs the supply chains of cargo that are used to transport freight to, from and within the combine MAG/PAG regions. These supply chains include cargo that is moved by truck. Additionally, a Truck Tour model was created for the combined MAG/PAG region. This section discusses how the outputs from those models are used to create truck trip tables that can be assigned by the MAG/PAG model.

The MAG/PAG assignment model requires truck trip tables as weekday trips by Time-of-Day period, between origins and destinations that are Traffic Analysis Zones (TAZ) or external stations.

The Truck Tour model produces tours of truck that travel from an origin of the tour to the destination of the tour with intermediate stops during the tour. This tour output is in the right time period (i.e., average weekday) and the legs of the tour, i.e., between each stop of the tour, can be unchained into a trip table, by making each leg of the tour into an individual trip between the stops. Additionally the tour model outputs the TOD time period in which each trip begins as the time at which a tour leg left one stop for the next stop.

The output of Supply Chain model is of annual cargo trips by SCTG2 commodity. Its time period must be converted from annual to average weekday. The flow unit must be converted from tons by SCTG to trucks. Some legs of the supply chain will be by rail, or another nontruck mode, and need not be integrated with the Truck Tour model. Only those supply chain legs which use trucks must be further processed. Additionally, the tours from the Supply Chain model already will be included in the tour model estimation data, i.e., ATRI for heavy trucks and StreetLight for medium and light trucks. The model outputs need to be adjusted to ensure that the tours from the Supply Chain model and those from the Truck Tour model are not double counted. Also, to the extent that trips from the Supply Chain model may limit the tours that can be created in the Truck Tour model, these also have to be considered.

To make this conversion from cargo to truck tours and to properly account for any limitations that the Supply Chain model places on truck model formation, it is necessary to understand some basics about cargo and truck tours.

- Cargo Tour: By definition all cargo tours are open. The buyer of the cargo will not be in the same location as the as the supplier of the cargo. If the buyer was in the same location as the supplier, no tour would have been necessary.

- Truck Tour: By definition all truck tours are closed, but only if the time period is long enough and the geography being considered is large enough.

All trucks in tours will eventually return to their starting locations. However a truck tour might span all of North American and might endure for several weeks before it does return to its starting location. Thus, while in theory, all truck tours are closed, in practice, from the perspective of a single region on single day, a piece of the truck tour may appear to be open.

Cargo in tours by definition do not include any movements of trucks without cargo. However a truck tour will include legs when the truck is carrying cargo, as well as legs when it is empty. A supply chain only considers the movement of cargo that is classified as freight. The additional transport of this cargo within a region may not be considered to be freight, especially freight as defined by the Freight Analysis Framework (FAF) that was an input to the Supply Chain model.

In order to covert supply chain outputs to truck tours and to integrate those truck tours with the Truck Tour model output, four types of truck tours are proposed. Those tours are:

Type 1 (Freight Cargo): These are truck tours that are included within the Supply Chain model outputs. They are formed at the Transport Logistics Nodes (TLN) with the region that are an intermediate stop along a supply chain (For example, a TLN can be an intermodal rail terminal). This includes the tours from that TLN that deliver the cargo by trucks to the MAG/PAG buyers in order to complete the supply chain, or that pick up the cargo by truck from MAG/PAG suppliers and deliver it to a TLN in order to begin the supply chain. These would be classified as a "Transportation/Warehousing" tour purpose in the Truck Tour model. However trips to "Transportation/Warehousing" to deliver cargo that is necessary for the operations of those industries, not to transship cargo, are no different than the delivery of cargo to any industry.

Type 2 (Some Non-freight Cargo): In the Supply Chain model, buyers of cargo include wholesalers, a NAICS 42 industry. However that wholesaler only "buys" the cargo to have stock that can then be delivered locally, in tours that are considered to be non-freight truck tours. The local non-freight tours that are classified as "Wholesale" purpose in the Truck Tour model are practically limited by the cargo that is delivered to that wholesaler where the supply chain destination in the MAG/PAG region also has an industry code of NAICS 42. Those deliveries to the NAICS 42 wholesalers can be as either as Type 1, or Type 3 truck tours, or even as direct shipments by other nontruck (e.g., rail) modes.

Type 3 (Freight Cargo): These are the truck tours that are outputs from the Supply Chain model where the supply chain selected is by "direct truck." The supplier can be any be any NAICS industry, and the buyer can be any NAICS industry. A NAICS 42 buyer, imposes a limitation on the non-freight truck tours that can be formed from that location as discussed in Type 2, but the cargo in these supply chains still has to be delivered to that NAICS 42 wholesaler.

Type 4 (All Other Non-freight): All other non-freight truck tours, including the transport of cargo that is not considered as a Type 3 tour, as well as all tours that only involve the provision of services, the transport of construction equipment or materials, etc. In the Truck Tour model this includes all types of tour purposes except those classified as "Wholesale" or "Transportation/Warehouse."

A crosswalk between the truck tours in the Supply Chain model and the tours in the Truck Tour model is shown in Table 8-1. The tour types in the truck tour model are based on the LU activity where the

tour begins. Those Truck Tour purposes that are known not to be in the Supply Chain model (e.g., Households) are shown as Type 4.

Table 8-1 Tour Type and Tour Purposes in Supply Chain and Truck Tour Models

Supply Chain Model	Truck Tour Model
Tour Type	Tour Purposes (Stop Purposes)
Type 3	Retail
Type 3	Ag/Min/Construction
Type 4 (not from SC)	Households
Type 4 (not from SC)	Government
Type 1	Transportation/Warehousing
Type 3	Manufacturing
Type 2 (Limited by SC)	Wholesale
Type 3	Service
Type 4 (not from SC)	Other

An investigation of the attributes of the supply chain outputs was undertaken. It was determined that the location or presence of a TLN in a supply chain was not made. The Supply chain model only outputs the principal modal family of the supply chain, not the locations, if any, of the TLN that might be used by that modal family of supply chains. However the supply chain model outputs do include shipment size ranges. In order to have some supply chains classified as Type 1 Truck Tours through TLNs, it was assumed that all small shipment sizes by definition will pass through a TLN. For Supply Chain truck tours that are Internal-External, the tours will be formed by the suppliers are the origin of the supply chains in the MAG/PAG region, and the TLN must be assumed to be at the external station where the Supply Chain exits the region. For Supply Chain truck tours that are External-Internal, the tours will be formed by the buyers that are the destination of the supply chains in the MAG/PAG region, and the TLN must be assumed to be at the external station where the Supply Chain enters the region. For Supply Chain truck tours that are Internal-Internal, the tours will be formed by the buyers that are the destination of the supply chains in the MAG/PAG region, and the TLN must be assumed to be at the origin supplier of the Supply Chain in the MAG/PAG region.

8.1.1 Supply Chain Model Output Records

In order to process the supply chain records from cargo into tours, it is necessary to use the attributes that also are available for those supply chains output records. The attributes of the Supply Chain Output Model that were found to be useful in this conversion are:

NAICS Industry

The Supply Chain model outputs will indicate the North American Industrial Classification System 2 digit Code (NAISC2) industry code of the MAG/PAG origin or destinations as shown in Table 8-2. Organs or Destinations outside of the MAG/PAG regions are shown only as a "Composite Industry." The actual industry outside of the region provides no information that could be useful to MAG/PAG.

Table 8-2 NAICS2 Industry Codes

NAICS2 Code	Name
11	Agricultural
21	Mining
23	Construction
31	Consumer manufacturing
32	Nondurable Manufacturing
33	Nondurable Manufacturing
42	Wholesale
44	Specialty retail
45	General Retail
48	Transportation
49	Warehousing and Couriers
51	Information
52	Finance and Insurance
53	Real estate
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services

The destination NAICS2 code is used to determine if a truck Tour is Type 1 or 3, or is Type 2.

SCTG2 Code

The Supply Chain model output indicates the Standard Classification of Transported Goods two-digit code (SCTG2). This information, along with truck size, can be used to convert the tons by truck to trucks. The conversion of tons per tuck by Truck typo and SCTG commodity was developed from the Arizona records in the 2002 U.S. Vehicle Inventory and Usage Survey database. Those tons per truck are shown in Table 8-3.

Table 8-3 Tons per Truck by SCTG and Truck Size

SCTG		Tons per Truck	
Code	Name	Medium	Heavy
01	Live Animals and Fish	3.09	19.84
02	Cereal Grains	9.47	22.88
03	Other Agricultural Products	4.73	20.48
04	Animal Feed	7.33	9.57
05	Meat, Fish, and Seafood	3.63	19.30
06	Milled Grain Products	1.89	17.80
07	Other Foodstuffs	4.38	14.73
08	Alcoholic Beverages	5.55	18.26
09	Tobacco Products	6.00	18.49
10	Building Stone	8.98	19.67
11	Natural Sands	13.21	22.69
12	Gravel and Crushed Stone	14.82	22.75
13	Nonmetallic Minerals, n.e.c.	13.46	23.67
14	Metallic Ores	8.24	23.35
15	Coal		
16	Crude Petroleum		
17	Gasoline	5.39	23.62
18	Fuel Oils	6.37	23.80
19	Petroleum Products, n.e.c.	4.25	21.43
20	Basic Chemicals	4.05	19.63
21	Pharmaceutical Products	3.60	18.65
22	Fertilizers	5.10	5.65
23	Chemical Products, n.e.c.	3.20	19.07
24	Plastics and Rubber	2.51	8.07
25	Logs	7.76	23.25
26	Wood Products	3.65	20.64
27	Newsprint	3.49	8.64
28	Paper Articles	3.45	11.59
29	Printed Products	2.77	18.52
30	Textiles and Leather	2.20	17.62
31	Nonmetallic Mineral Products	10.36	22.45
32	Base Metal	2.67	19.80
33	Articles of Base Metal	2.35	18.98
34	Machinery	2.59	18.63
35	Electrical Equipment	2.23	8.63
36	Motorized and Other Vehicles	2.79	17.45
37	Transportation Equip, n.e.c.	7.34	21.76
38	Precision Instruments	2.57	18.93
39	Furniture	1.92	17.01
40	Misc. Manufactured Products	2.69	11.36
41	Waste and Scrap	7.64	21.02
42	Mail	3.70	16.81
43	Mixed Freight	3.00	16.89

Shipment Size

The Supply Chain model output will classify the shipment size of that supply chain as Small (< 150 lbs.), Medium (150 to 1,499 pounds.), Large (1,500 to 34,999 pounds.), and Very Large (>35,000 pounds). As noted previously, in the absence of information on whether the a supply chain uses a TLN, ALL supply chains with a small shipment size are considered as Tour Type 1, while all other shipments sizes are considered to be Tour Type 3.

Tons

The Supply Chain model output will indicate the number of annual tons shipped in that supply chain. This information will be used to convert to the weekday trucks required by the truck table to be assigned to the MAG/PAG highway network.

External MAG/PAG Origin Station

The supply chain model output reports the origin of the supply chain as the MAG/PAG TAZ for supply chains that being in the region and the FAF region for those supply chains that begin outside of the region. While the MAG/PAG TAZ can be used to develop a MAG/PAG truck trip table, the FAF External region is not used in the assignment model. Using a crosswalk between FAF External Regions and MAG/PAG External stations that was developed by MAG staff, the supply chain model output reports a field that converted the origin, if necessary, to the MGA/PAG External Station.

External MAG/PAG Destination Station

The supply chain model output reports the destination of the supply chain as the MAG/PAG TAZ for supply chains that end in the region and the FAF region for those supply chains that end outside of the region. As described above, for the Supply Chain destinations that are outside of the region, this is used to compute the associated MAG/PAG external station.

8.1.2 Vehicle Choice

As shown in Table 8-3, the tons per truck by SCTG2 commodity varies by the size of the truck. Before converting from annual tons in the supply chain to daily trucks it is necessary to know the size of the truck, as medium or heavy, which will be used in the conversion from tons to trucks. The vehicle choice by Truck Tour Type is:

Tour Type 1 (Freight) in the delivery from, or pickup to, TLNs, is assumed to be only by heavy trucks.

Tour Type 2 (non-freight cargo delivery) only applies to the use of medium trucks by wholesalers to deliver goods to their customers. These medium truck tours already are presumably included as weekday trucks in the truck tour model. There also is no need to convert from tons to trucks because the tour flow unit already is in trucks. What is needed is to scale the truck tours that begin in "Wholesalers" in the medium truck tour model. While these tours already are being forecast by the medium truck tour model, it is necessary to check that the number of these truck tours is reasonable. The total annual tonnage received by wholesalers in a TAZ, which can be obtained from processing the supply chain model output with NAICS_42 at the destination, can be converted to annual medium trucks by using Table 8-3. This number should be equal to the number of medium truck tours with a 'wholesale purpose that begin in that TAZ, converted from weekday tours to annual tours by multiplying by 295. If the numbers are not equal, then the medium truck tour model outputs are scaled by the ratio of the supply chain outputs.

Tour Type 3 (freight) does not need a vehicle choice model. It is assumed that direct truck shipments of supply chain are by heavy (combination tractor trailer) trucks.

Tour Type 4 (non-freight) does need a vehicle choice model, but that model already is implicitly being used in the medium and heavy truck tour models. Since these tours are, by definition, not related to the supply chain model outputs, the vehicle choice does not change and need not be computed.

8.1.3 Conversion from Annual Tons to Daily trucks

The documentation of the process to convert from annual supply chain tons to daily truck trips will be presented in order of increasing complexity.

Tour Type 4 (non-freight) are by definition unrelated to the supply chain model tonnages. They already are expressed as weekday tours of medium and heavy trucks. While these tours do have to be unlinked from tours, with many legs, to trips for each individual leg, no conversion is required from tons to trucks. The TAZs of each stop on the tour become the origins and destinations of truck trips. The time of departure from a stop on the tour, determines the TOD period in which that trips is reported.

Tour Type 2 (non-freight cargo delivery) also do not need to be converted from tons to trucks. These are the "wholesaling" purpose tours that are estimated by the medium truck model. While their total may be scaled as described above, once scale, the tour are unlinked into truck trips by TOD exactly as described for Tour Type 4 above.

Tour Type 3 (freight) are reported in annual tons by heavy trucks and do need to be converted to daily TOD trucks. However by definition a direct truck supply chain moves, between the specified origin and the specified destination, in a single leg. For this truck Tour Type, the tour of the cargo is the same as the tour of the truck. Table 8-3 can be applied to the tonnage of that supply chain to convert from annual tons to annual trucks. The annual trucks are converted to weekday trucks by dividing by 265, the equivalent weekdays per year. The supply chain model provides no information by which the TOD period can be determined. It is assumed that the TOD distribution of Type 4 tours for heavy trucks, can be applied to the daily trucks.

Only for **Tour Type 1 (Freight)**, which is synthetically assumed to pass through a TLN, will there be a need to compute truck tours that are different than cargo tours. Tours are developed using the outputs from the supply chain model. These tours are estimated using the same model components (stops per tour, next stops, etc.) for heavy trucks with a transportation purpose as for the Tour Type 4. Those tours were estimated for heavy trucks using ATRI data and for medium trucks using StreetLight data. They are converted from annual tons to daily trucks as part of a two-step process. The tons are first converted from annual to daily using the 295 weekday equivalents. The daily tons at each stop, a supply chain output of the tour is summed. That is then converted to a truck equivalent for the tour using Table 8-4.

8.1.4 Integration Outputs

The results of the converting the truck supply chains tons into truck trips and the unlinking of the outputs of the light, medium and heavy truck tours model into weekday truck trips is shown in Table 8-4.

Table 8-4 Integration Model Outputs

		Light		Medium		Heavy	
	Tour Type	Tours	Trips	Tours	Trips	Tours	Trips
	1	-	-	-	-	812	2,842 [65]
Only Wholesale Purpose	2	-	-	9,572	40,749 [66]	-	-
	3	-	-	17,088	17,088	22,665	22,665
Excluding Wholesale Purpose	4	7,386	15,355	110, 575 [66]	469,726 [66]	25,076	100,725

While the number of Type 1 tours is small, this is consistent with the findings of the Wisconsin SHRP2 C20 project. That project also found that while the number of annual supply chain shipments is large, when converted to daily truck tours and trips, their number was very small compared to all other truck movements. In that project all truck cargo supply chains, not just those shown as Tour Type 3 in Table 8-4, were directly converted from tons to truck trips without converting any of them to tours.

8.2 Vehicle Type Choice Model

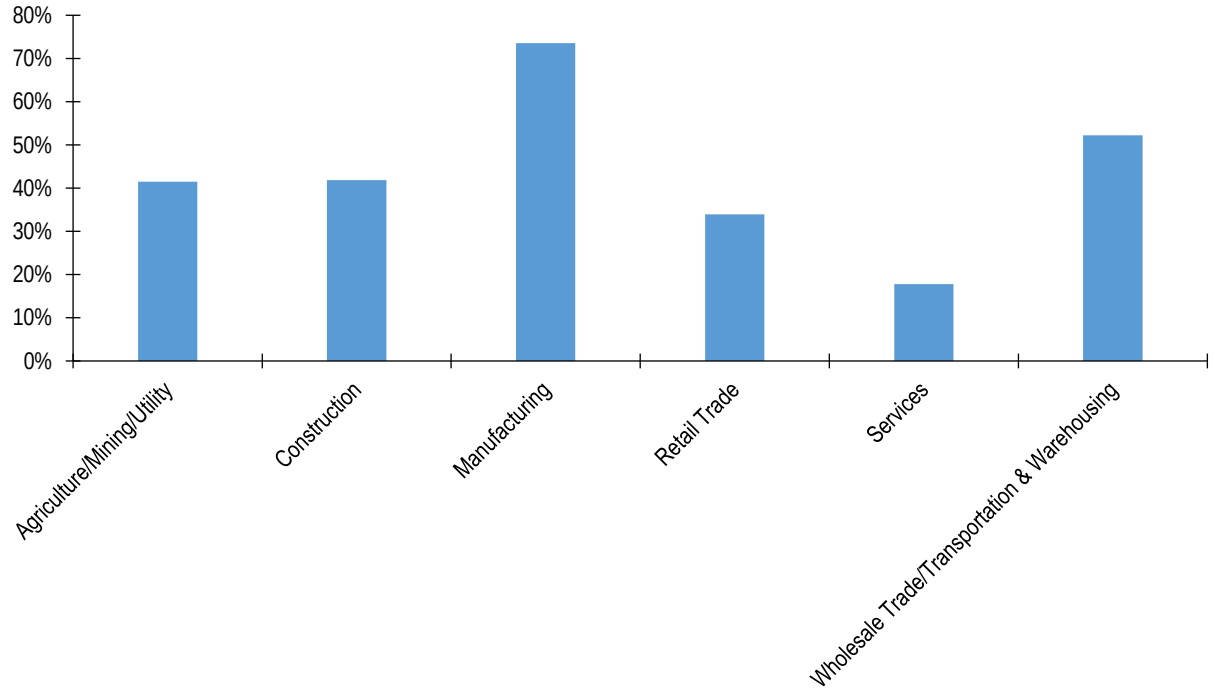
8.2.1 Model Estimation

To predict the usage of heavy versus medium weight trucks by establishments in the MAG region, we employed a logit model. The data came from the 2016 commercial establishment survey. This survey collected data at 416 establishments in the MAG region. From these establishments 229 had heavy or medium trucks based or making stops there. Respondents were asked about all arrivals, but only about the departures of those trucks based on location. The latter are analogous to the start of a truck tour. These data were expanded to produce a weighted dataset of truck behavior which is used in the estimation described here. As a result, many more arrivals were reported than departures.

The data were rearranged to get a weighted list of trucks of each type with the corresponding establishment's characteristics associated with them. The percentage of trucks in the survey which were heavy trucks is shown below.

[65] Estimated, the trips are unlinked as part of all HT tours.
[66] Estimated and scaled. The 9,572 are included in a total of 120,147 tours and 511, 475 trips and they have been allocated accordingly.

Figure 8-1 Percentage Heavy Trucks by Industry



Many binary logit models were specified and tested. Here are the results of the best performing model:

Table 8-5 Model Estimation Results

Variable	Coefficient	t-statistic
Intercept	-1.05	-0.8
Log Employment	0.428	4.9
Construction Establishment	-0.806	-0.6
Manufacturing Establishment	0.79	0.6
Retail Establishment	-0.653	-0.5
Service Establishment	-0.727	-0.6
Wholesale, Transportation, etc.	0.372	0.3
Establishment	–	–
Observations	**430**	
Final log (L)	**-298.1**	
D.O.F.	**7**	
Rho²(0)	**0.188**	
Rho²(c)	**0.097**	

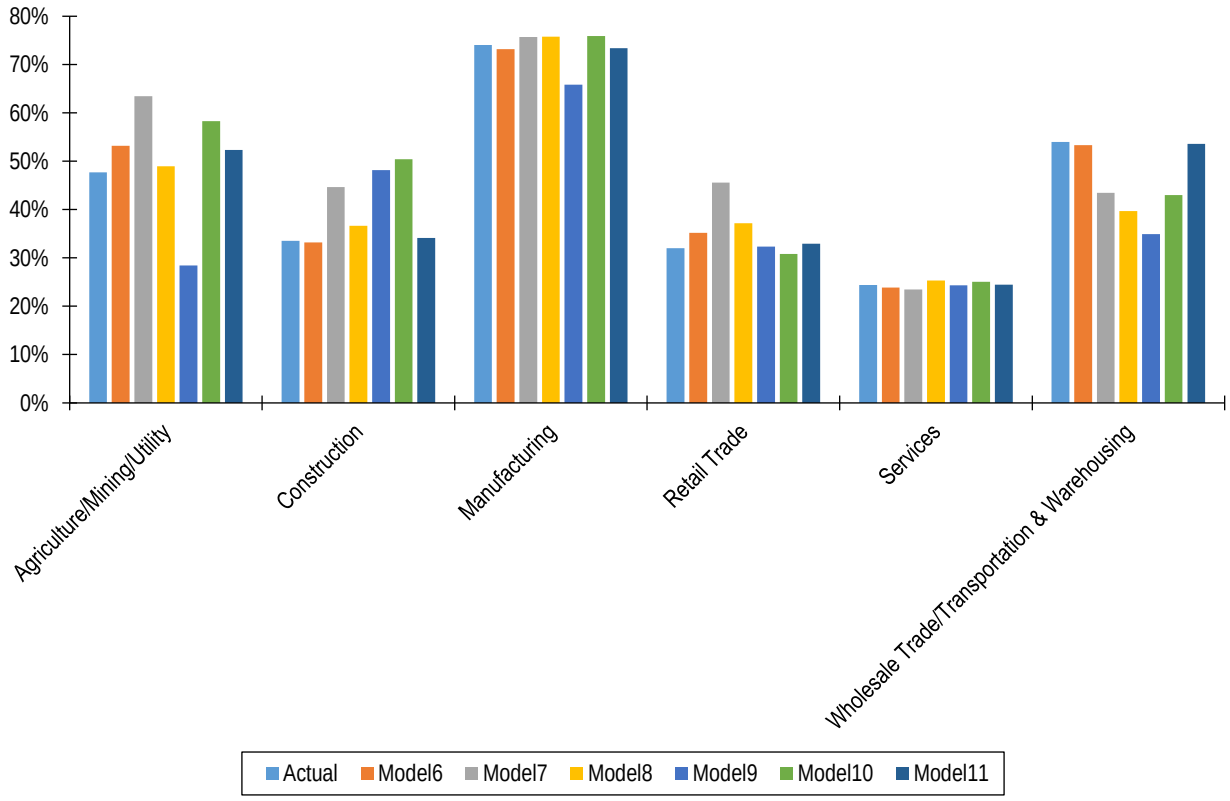
8.2.2 Model Validation

A simulation also was constructed to use this to predict percentages heavy (versus medium) trucks to establishments. The results showed that this regression would reproduce the industry average and display

some of the effect of employment size on percentage. Note the data do not follow a simply increasing trend with higher employment.

The graph below shows a comparison of a few models to the actual percentages by industry. Model 6 is the specification shown in the table above.

Figure 8-2 Shares of Heavy Trucks by Total Industry



8.3 ODME of Processed Truck GPS Data

8.3.1 Methodology

The ODME process estimates the OD trip table based on the observed OD travel pattern provided in the seed ATRI table and the highway counts coded in the network. The ODME is an iterative process that develops an OD table by adjusting the seed OD trips to match the assigned flow with the coded counts. If there is no count along the highway path between an OD pair then the output OD trip value for that OD pair remains unadjusted. The term "unadjusted" means that the OD trip remains same as that of seed. The adjusted and unadjusted OD pairs were identified from the ODME output trip table. The trip values in unadjusted cells were adjusted using the adjustment factor as calculated in equation (1) and equation (2).

Adjustment factor = (Total OD trip value in adjusted OD pairs of ODME output table)

÷

(Total OD trip value in adjusted OD pairs of seed table) (1)

Adjusted OD trip = Unadjusted OD trip * Adjustment factor (2)

8.3.2 Results

An origin-destination matrix estimation (ODME) process was implemented to develop origin-destination (OD) truck trip tables by truck type for the combined MAG-PAG region. The StreetLight truck GPS data was used for light and medium trucks while the ATRI truck GPS data was used for heavy trucks as observed seed OD tables in the ODME process. This process expanded the StreetLight and ATRI truck trip tables to the entire megaregion with the help of light, medium and heavy truck counts. This resulted in expanded light, medium, and heavy trucks.

The seed origin-destination table and the highway counts served as the inputs to the ODME process. The seed origin-destination table represents a sample of observed OD travel pattern, which is derived from the processed StreetLight and ATRI GPS survey data.

Inputs to ODME

The ODME process requires two types of inputs: the seed OD table and the highway counts. The seed OD table represents the observed OD travel pattern, which was derived from the processed Streetlight data to develop the seed OD trip table at the combined MAG-PAG region's TAZ level. The Streetlight data includes OD trips between 316,205 OD pairs for medium trucks and between 4,528 OD pairs for commercial light trucks.

According to the time stamp of survey data, it was found that the data was collected for 30 days. The Streetlight data captures 742,904 OD medium truck trips and 6,005 OD commercial light truck trips within the study region for 30 day time period. The monthly trips were converted to daily level using a factor of 30.

The medium and commercial light truck count data was used as another input to the ODME process. The count data was provided by MAG and PAG for their respective regional boundaries, geocoded onto the highway network, as shown in Figure 8-3 and Figure 8-4. For medium trucks, the counts were available on 564 links on the network, both freeway and non-freeway. For commercial light trucks, the counts were available on 302 links, both freeway and non-freeway.

Figure 8-3 Medium Truck Count Locations

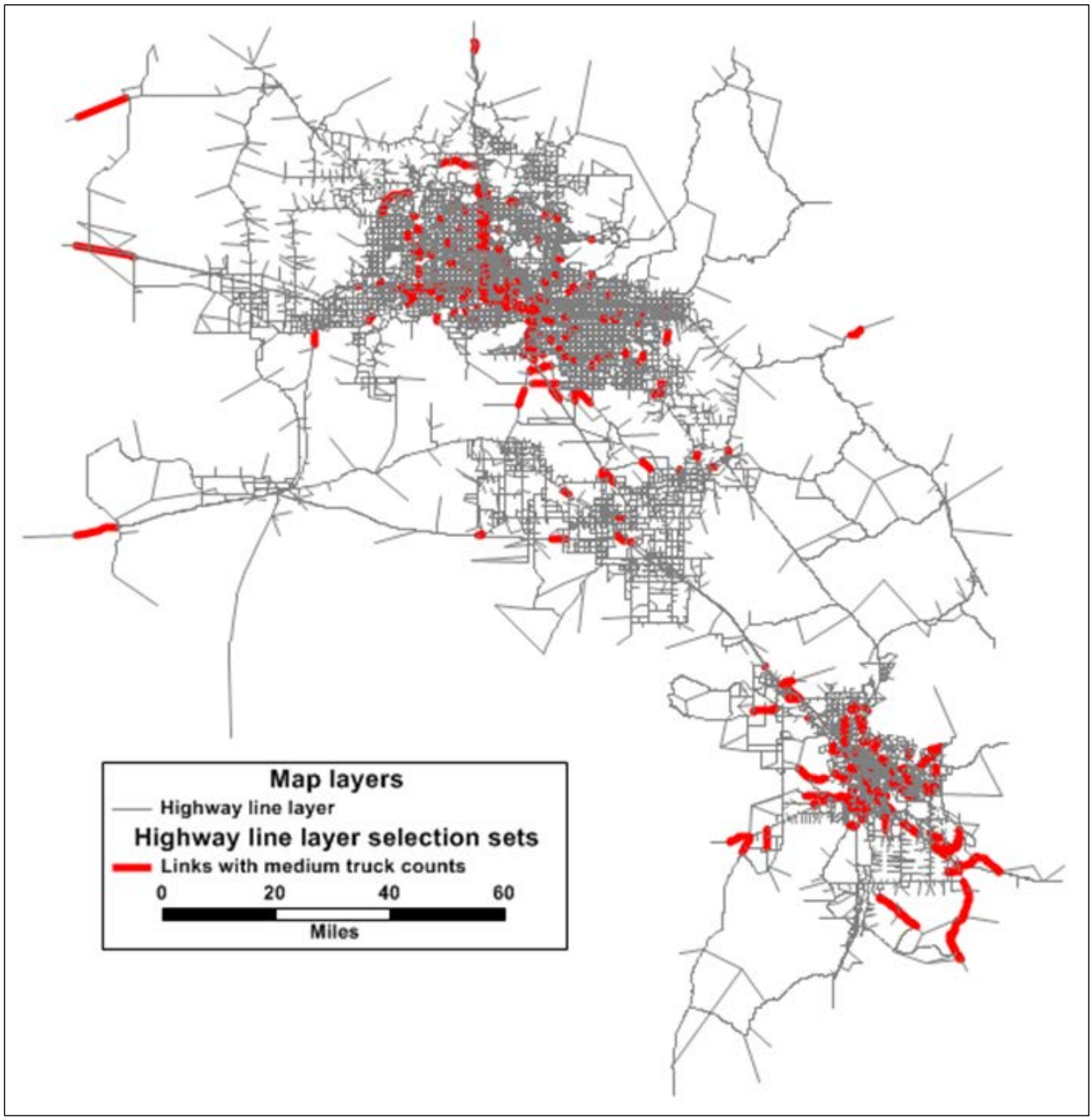
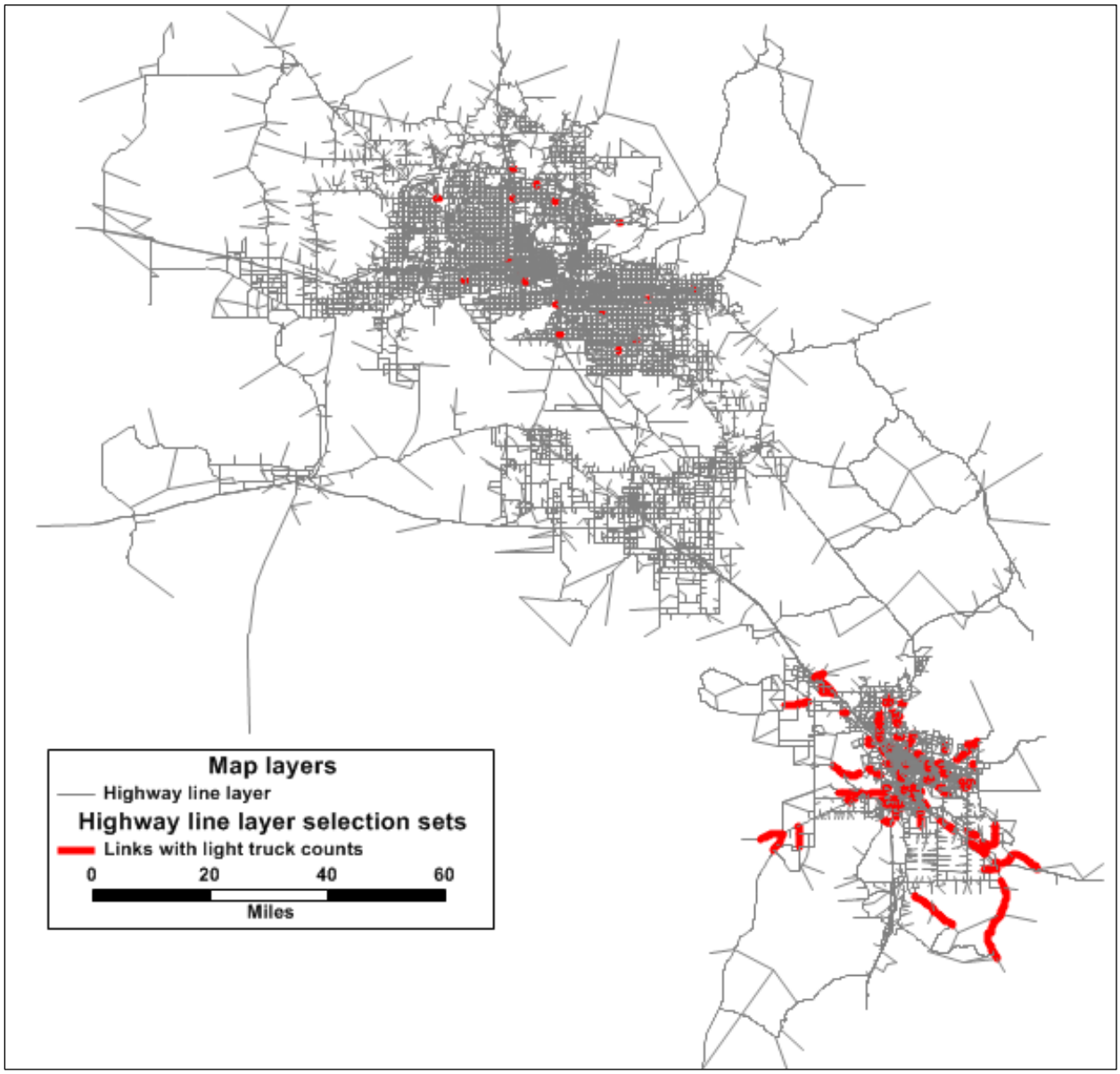
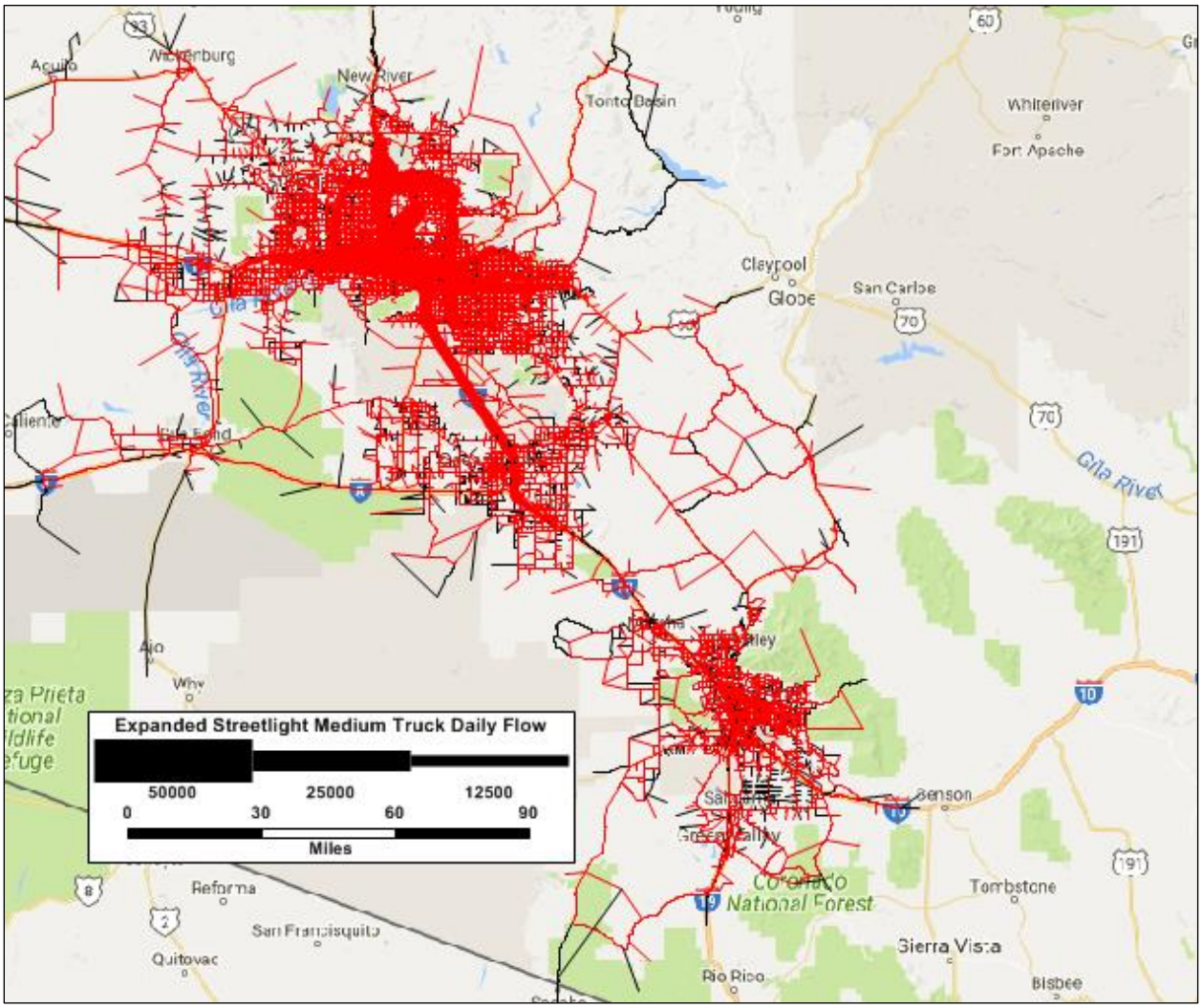


Figure 8-4 Commercial Light Truck Count Locations



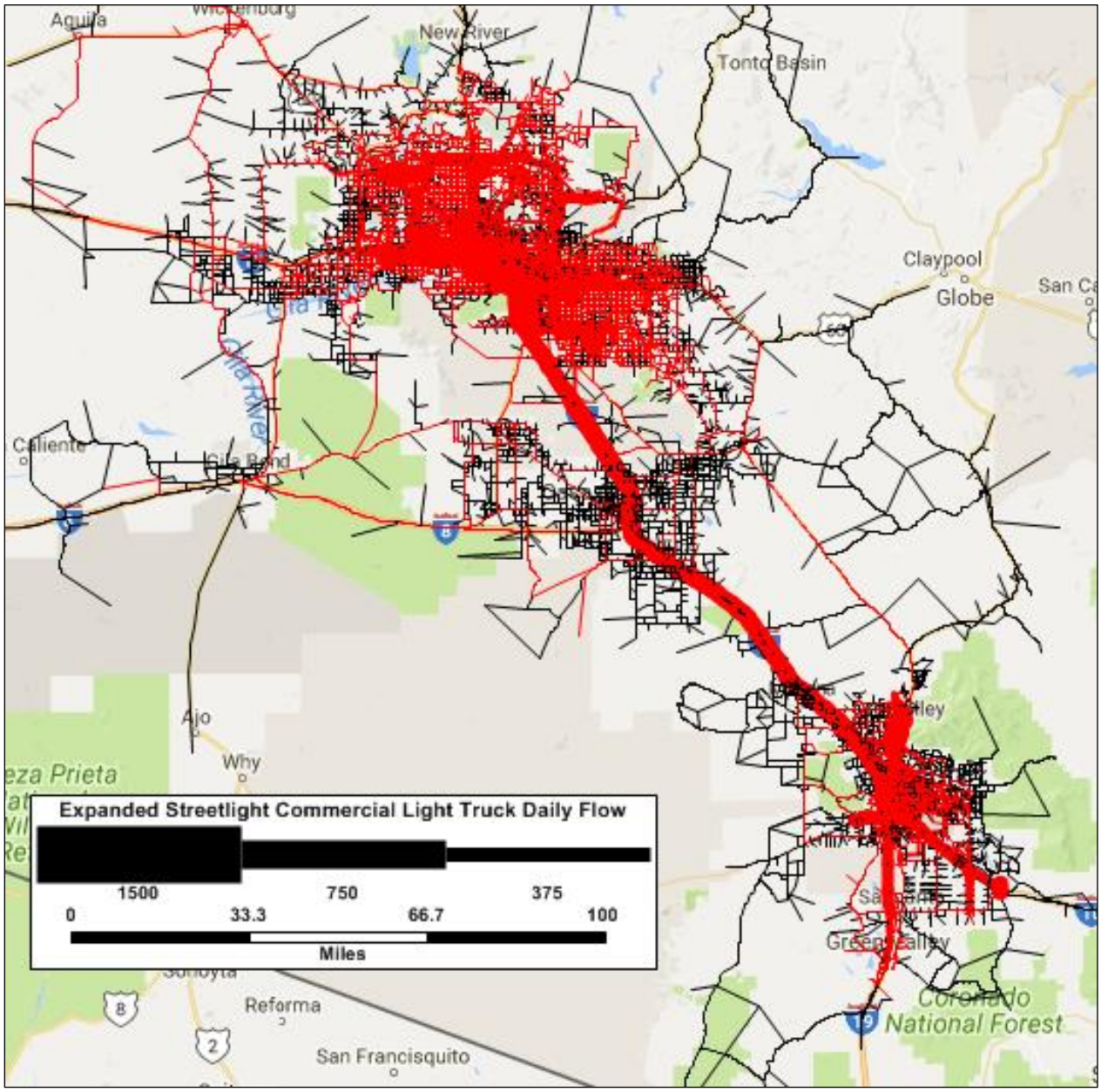
The final adjusted ODME output is then assigned on the highway network and compared with counts to test the performance of the ODME process. Figure 8-5 and Figure 8-6 present the results from the assignment process for medium trucks and commercial light trucks, respectively.

Figure 8-5 Assignment of ODME – Medium Trucks



The expansion of the Streetlight medium truck trip table results in 523,431 daily medium trucks.

Figure 8-6 Assignment of ODME – Commercial Light Trucks



The expansion of the Streetlight commercial light truck trip table results in 15,402 daily commercial light trucks.

Figure 8-7 shows a scatter plot between the medium truck counts and the assigned truck flows from the ODME process. Similarly, Figure 8-8 shows a comparison between the commercial light truck counts and the assigned truck flows from the ODME process. The robust R-squares of both figures show strong correlations between the expanded trip tables and the truck counts for both truck types. Moreover, both charts have slope values close to 1, reaffirming the strong correlation between the expanded trip tables and the truck counts.

Figure 8-7 Comparison between Counts and Assigned ODME Flow of Medium Trucks

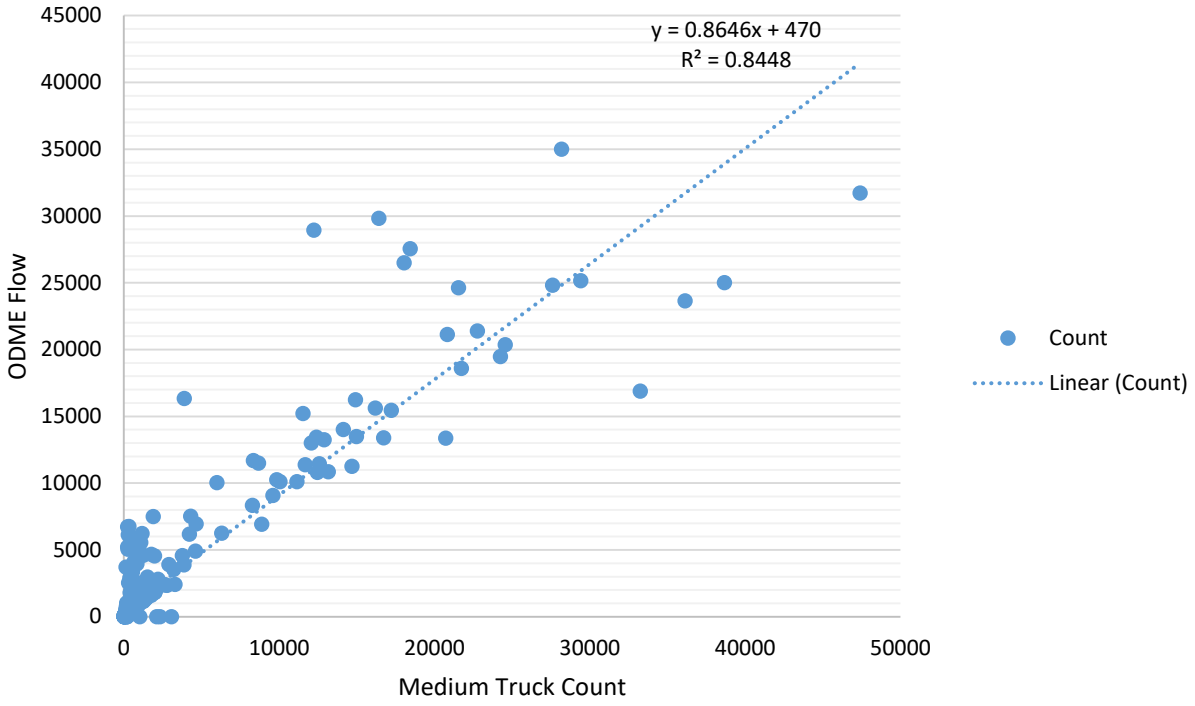
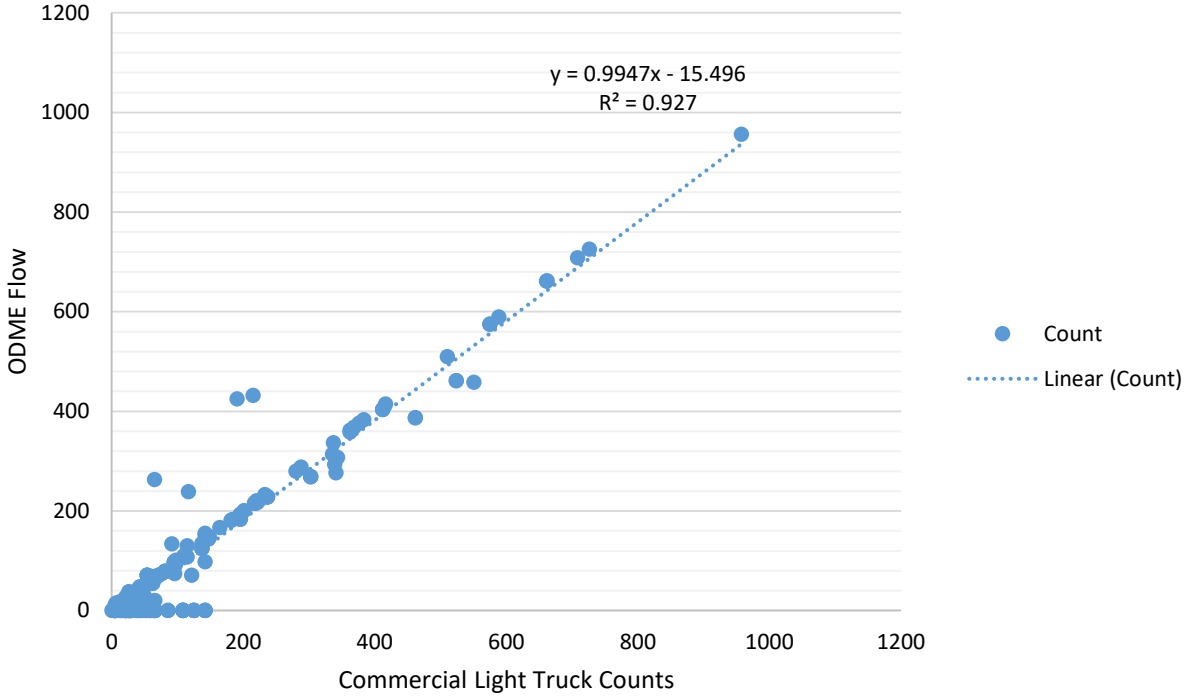


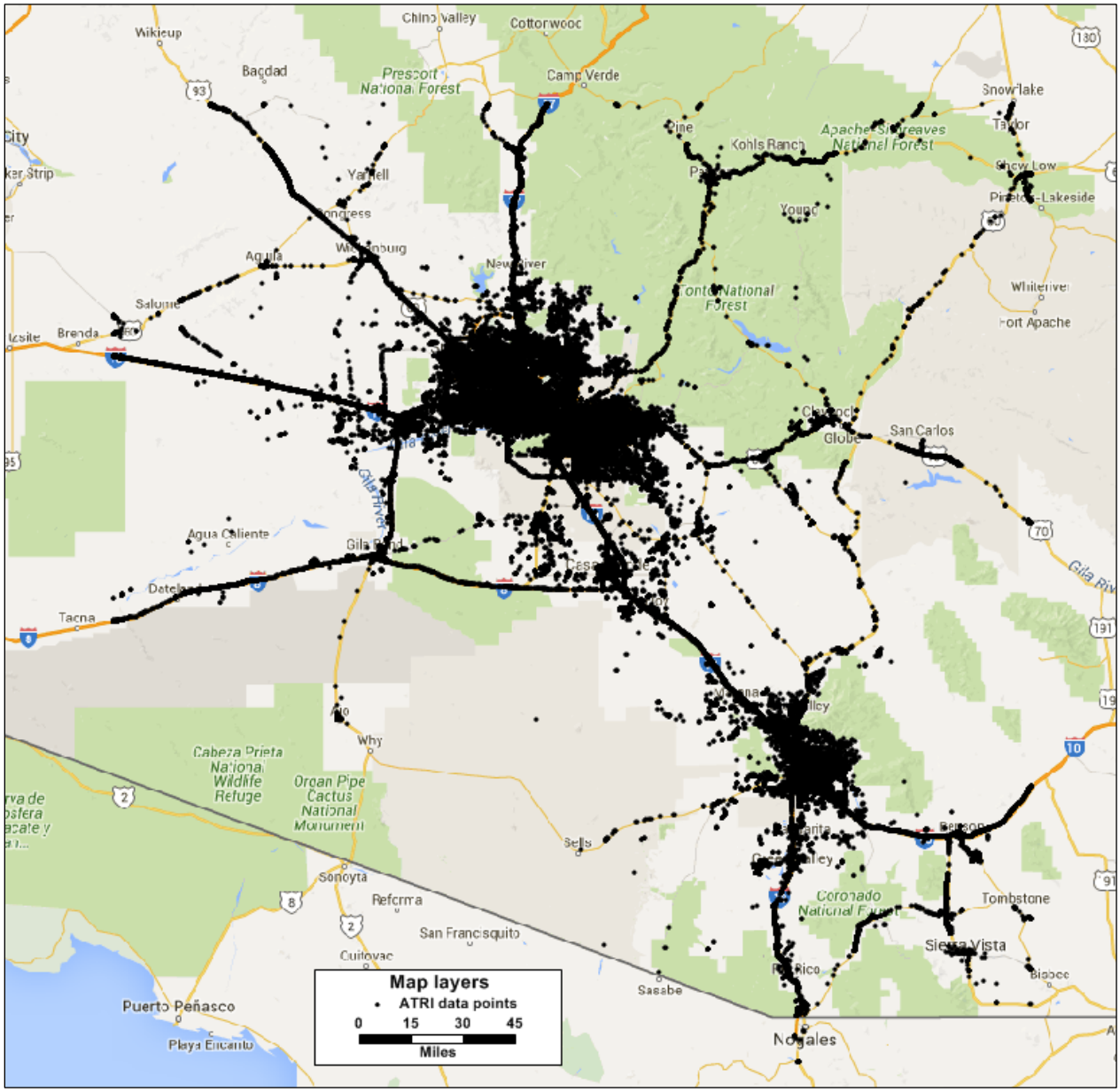
Figure 8-8 Comparison between Counts and Assigned ODME Flow of Commercial Light Trucks



The ATRI data was processed to develop the seed OD trip table at the combined MAG PAG region's TAZ level. The ATRI data includes OD trips between 139,939 OD pairs, which represents only a

fraction of all the possible OD pairs in the region. However, the geographic locations of the ATRI data points (shown in Figure 8-9) indicate a good representation of heavy truck travel within the region. Figure 8-9 also shows that the data points capture the truck travel on major roads and major activity locations.

Figure 8-9 ATRI Data Points

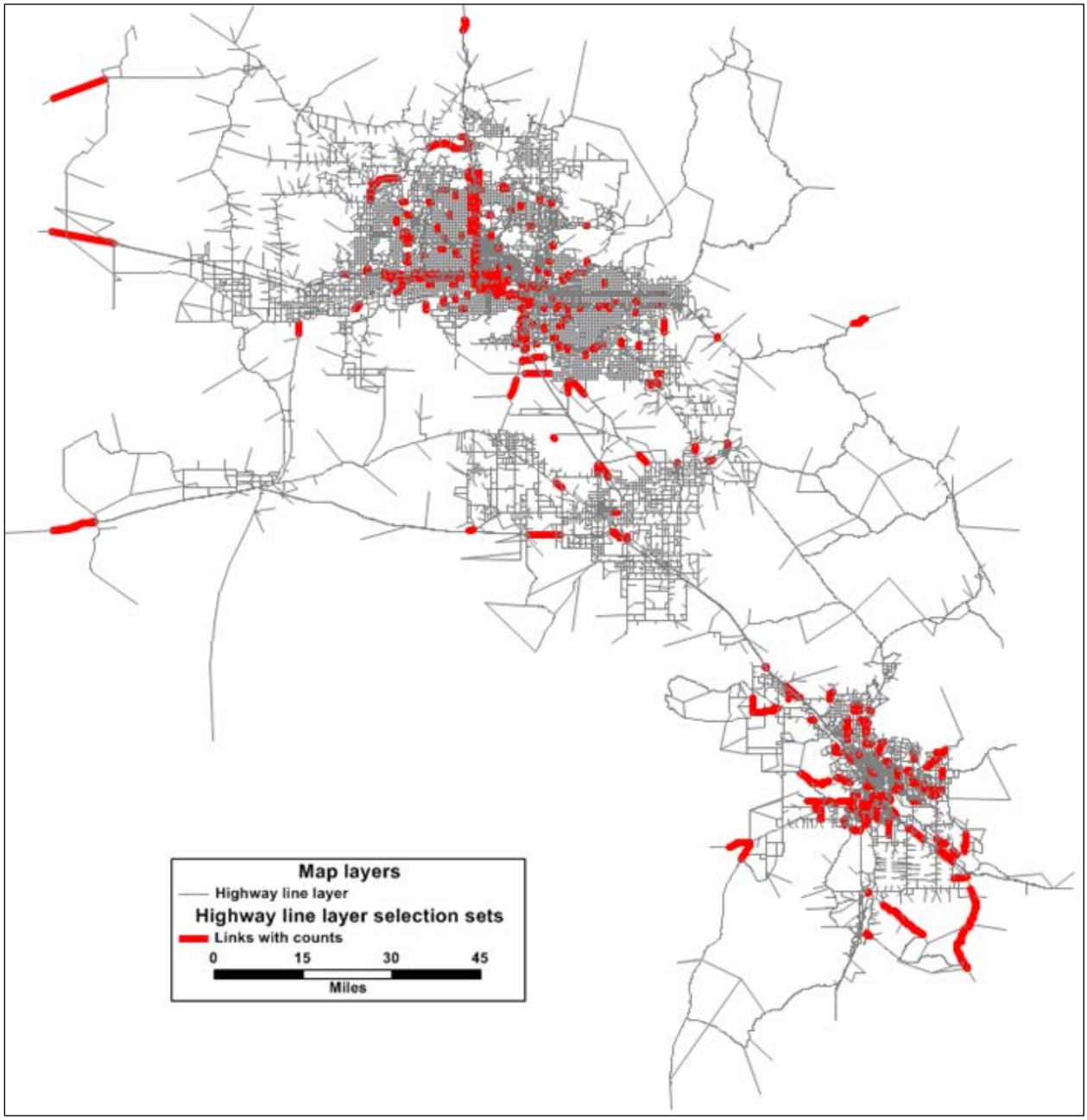


TAZ IDs were attached to each of the ATRI data points using geographic processing. Then, the data points were aggregated at the TAZ level to develop trips between origin and destination TAZ.

According to the time stamp of survey data, it was found that the data was collected for 30 days. The ATRI data captures 996,031 OD heavy truck trips within the study region for 30 days' time period. The monthly trips were converted to daily level using a factor of 30.

The heavy truck count data was used as another input to the ODME process. The count data was provided by MAG and PAG for their respective regional boundaries. The data was geocoded into the highway network, as shown in Figure 8-10 that depicts counts on highway network. The counts were available at 618 locations on the network that that were located both on freeways and nonfreeways.

Figure 8-10 Count Locations in Highway Network



The final adjusted ODME output was assigned on the highway network and compared with counts to test the performance of the ODME process. Figure 8-11 presents the results from the assignment process. As expected, there is a large volume of heavy trucks along the major freeways like I-10 connecting Phoenix and Tucson.

Figure 8-11 Assignment of ODME Heavy Trucks

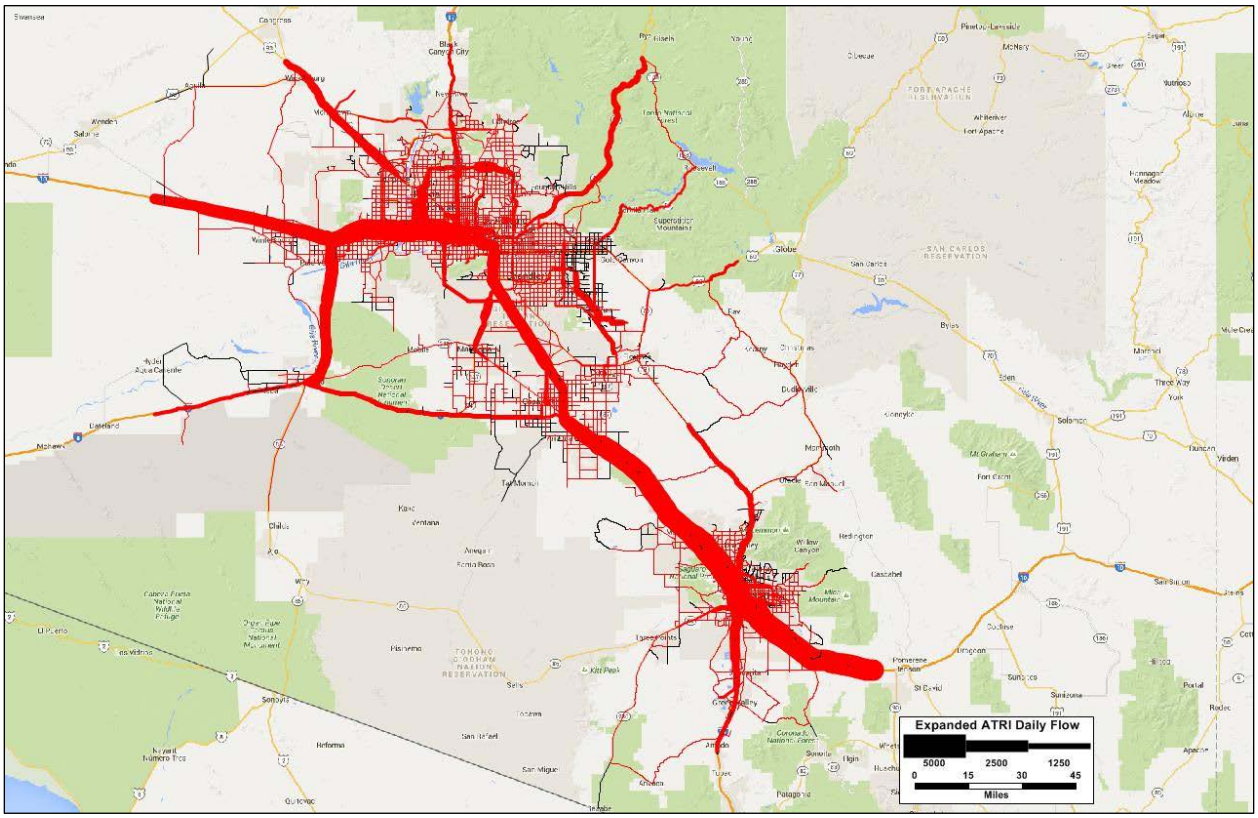
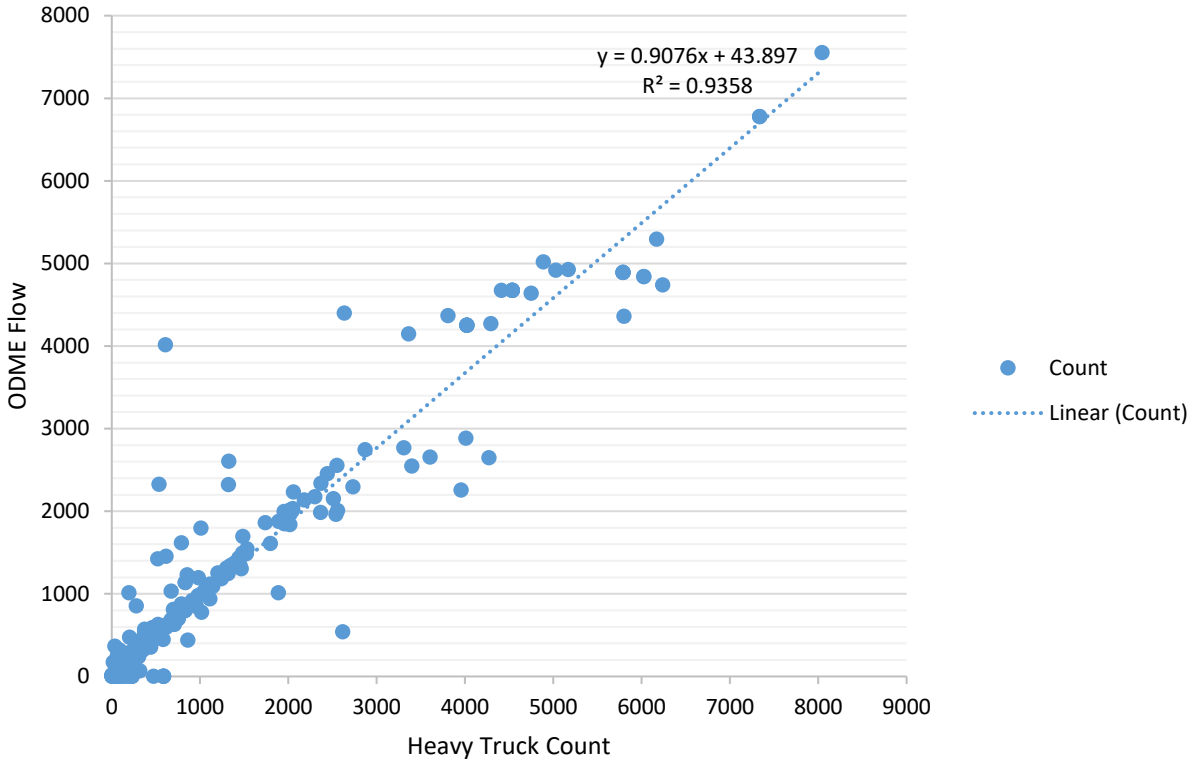


Figure 8-12 shows a scatter plot between the heavy truck counts and the assigned truck flows from the ODME process for all the 627 count locations. The almost 45 degree line shows that the ODME process had sufficient truck counts to expand the ATRI truck trip table to represent the entire region. The R-square of 0.9358 also corroborates the strong correlation of the expanded ATRI truck trip with against the heavy truck counts.

Figure 8-12 Comparison between Counts and Assigned ODME Flow of Heavy Trucks



9 Model Implementation and Deployment

Consultant Technical Lead –Zahra Pourabdollahi (RS&H) and Arun Kuppam (Cambridge Systematics)

MAG Technical Lead – Kyunghwi Jeon and Daehyun You

The behavior based freight model has been implemented using R and Java script language and deployed on MAG's server as well as Desktop. All model components were tested and worked. Java codes were compiled using 64-bit Eclipse Mars IDE for Java EE Developers which is a free software. Users may need to re-compile Java codes if there are any changes in Java codes.

9.1 Programming Language

Table 9-1 shows what script language was used for the model implementation. R scripts do not require a compiler so users could run R scripts in either Rgui, RStudio or DOS command prompt.

Table 9-1 Programming Language for Model Implementation

Model name	Programming language
Firm synthesis model	R
Supply chain model	R
Supplier selection model [67]	Java
Truck tour model	R

9.2 Software and Hardware Requirements

This section describes the software and hardware requirements to run the behavior based freight model.

9.2.1 Software Requirements

Below is a list of software to run the model system. RStudio has a user-friendly interface which can easily check the status of variables and data tables during the model run.

- 64-bit Windows 7/10 or 64-bit Windows Server 2008 R2 Enterprise
- 64-bit R-3.3.0 or later version
- 64-bit RStudio Version 0.99.902 (optional)
- 64-bit Java (jdk1.8.0_92 or later version)

9.2.2 Hardware System Requirements

Default memory settings are based on MAG's machine. So users must do the test run to find an optimal memory setting in R and Java runs on user's machine.

- Minimum of 48 GB RAM

[67] Supplier selection model is a sub model of the supply chain model.

- Minimum of 15 GB of free space
- Processor Speed: 2.9GHz or faster (recommended)
- 2 Processors (at least 12 logical cores, recommended)

9.3 Directory Structure and Model Run

This section is to describe the directory structure and global variable set up for model run.

9.3.1 Working Directory and Global Variable Set Up

Working directory is the path where R codes, all input and output directories are set up. The work directory is defined by **modelWD** variable and set up by **setwd** function in "MAG_BehaviorBasedFreightModel.R" which is a main script and should be under the main work directory. Please note that R uses forward slash ("/") to indicate directory paths.

User could define as many commodities as desired in the supply chain model. There are 42 commodities based on SCTG codes (1-43 excludes 42). The markets to be analyzed are determined by market_i and market_j variables. market_i determines the start market and market_j determines the end market. The default is to run all markets with **market_i <- 1** and **market_j <- 43**. In BehaviorBasedFreightModel.R, user have to make sure if the work directory and SCTG settings are correct:

Example (BehaviorBasedFreightModel.R):

```
#Set the work directory
projfolder="D:/BehaviorBasedFreightModel_20161101/"
setwd(projfolder)

#Determine the markets (based on SCTG codes) to be allocated in supplier selection model
#e.g. (1 - 43): SCTG 1 through 43
#start market
market_i <- 1
#end market
market_j <- 43

# Java Path - If there are multiple versions of Java installed, you may need to update the
path below i.e. "C:\PROGRA~1\Java\jdk1.8.0_92\bin"
path_java <- "C://PROGRA~1//Java//jdk1.8.0_92//bin//java -server -Xms2G -
Xmx60G -Xss12M"
```

To handle warnings, R provides the option to show or hide warnings generated during the model run. It is suggested to always show warnings generated during the run.

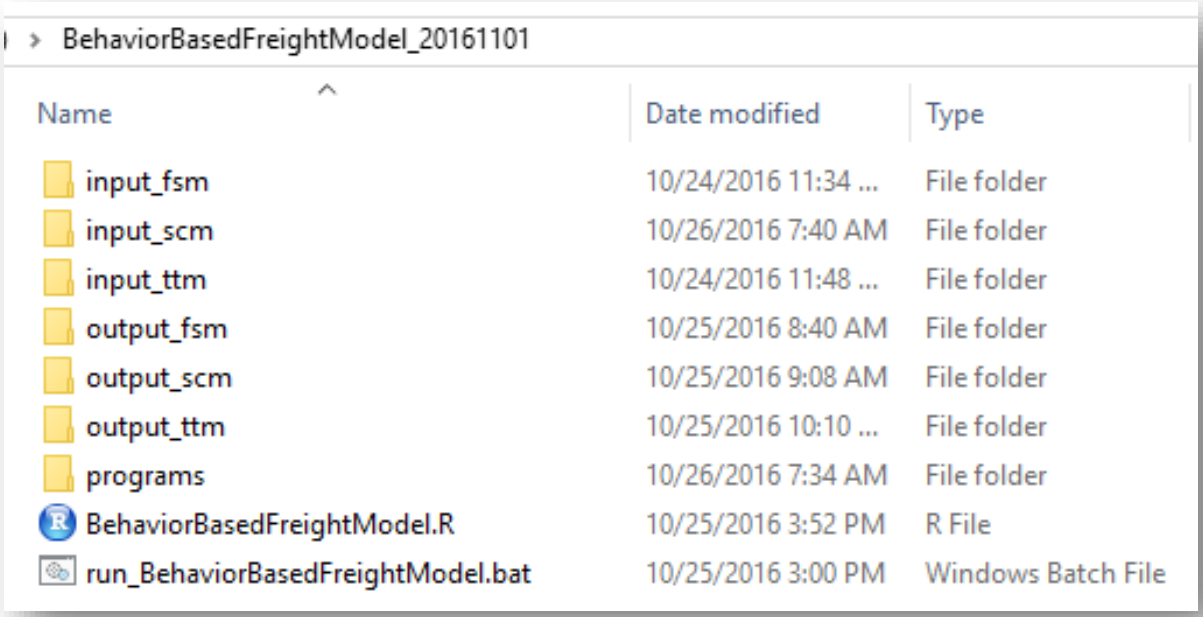
User should check the input and output directory path in ssm_run.properties under programs directory. Below shows an example of directory paths:

```
# INPUT FILES DIRECTORY
input_files_directory = "../input_scm/"
```

```
# INPUT FILES FROM SUPPLY CHAIN MODEL DIRECTORY
input_files_from_scm_step1 = "../output_scm/"
# OUTPUT FILES DIRECTORY
output_files_directory = "../output_scm/"
```

Figure 9-1 shows the working directory setup for the model. There are 7 sub directories in the main working directory which are also described below.

Figure 9-1 Working Directory Set Up



There will be sub directories for inputs and outputs of each model (see Table 9-2), and the program folder which contains all model scripts, supplier selection model property file, and Java Archive file.

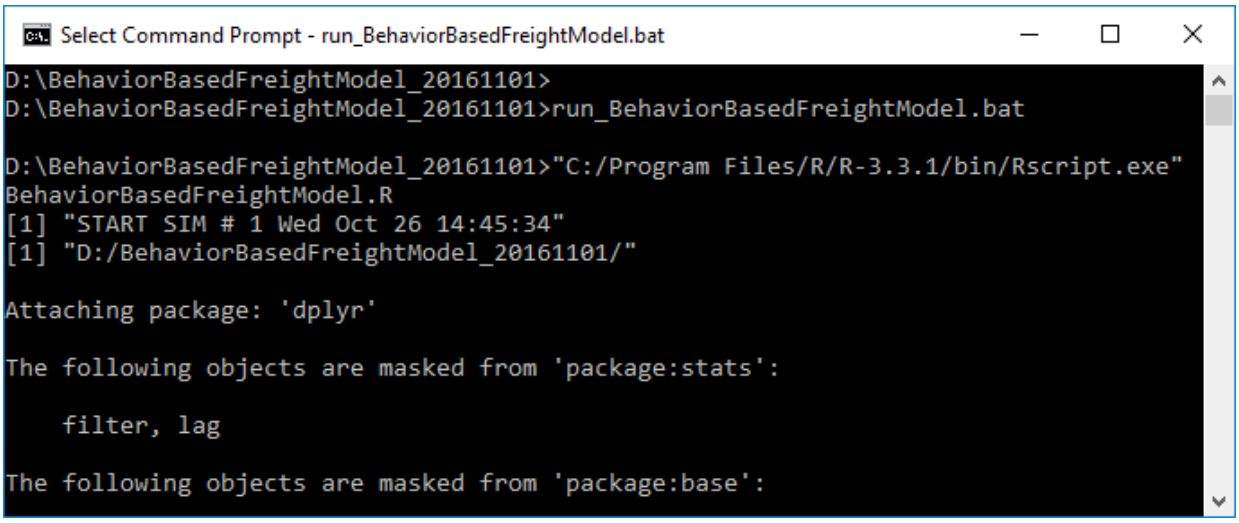
Table 9-2 Sub Directories in the Main Working Directory

Sub directory name	Description
input_fsm	Firm synthesis model input folder
input_scm	Supply chain model input folder
input_ttm	Truck tour model input folder
output_fsm	Firm synthesis model output folder
output_scm	Supply chain model output folder
output_ttm	Truck tour model output folder
programs	Model script, model properties, and Jar file folder

9.3.2 Model Run and Runtime

Before running the batch file (**run_BehaviorBasedFreightModel.bat**) in the main working directory (e.g., D:/BehaviorBasedFreightModel_20161101), users need to check and update R program path in batch file if necessary.

Figure 9-2 Model Run in Command Prompt



Test run was done on the high performance computer with Intel® Xeon® CPU E5-2690 @ 2.90GHz (2 processors, 32 Cores), 96 GB of RAM, 64-bit Operating System, and Window Server 2008 R2 Enterprise. Total model run time on this computer was about 19 hours. Table 9-3 shows more runtime information on each sub model.

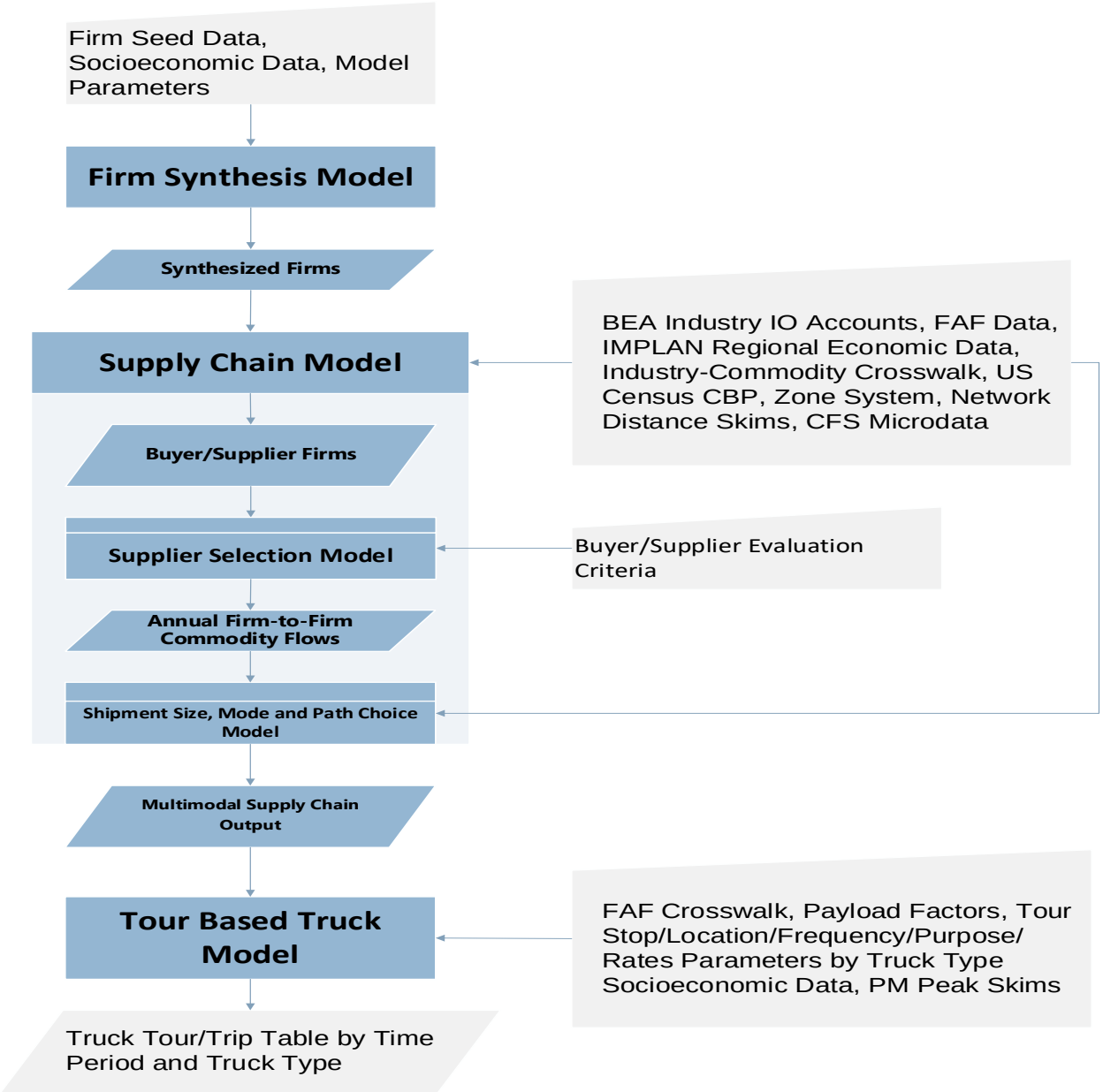
Table 9-3 Model Runtime

Model Name	Run Time
Firm Synthesizer Model	3 hours 17 minutes
Supply Chain Model	10 hours 24 minutes
Tour based truck Model	5 hours 21 minutes
Total	19 hours 2 minutes

9.4 Model Input and Output Files

This section is to describe the input and output files of the firm synthesis model, supply chain model, and truck tour model. Figure 9-3 shows the overall model structure, and its inputs and outputs.

Figure 9-3 Overall Model Structure, Inputs, and Outputs



9.4.1 Firm Synthesis Model Inputs

Figure 9-4 and Table 9-4 show the inputs of the firm synthesis model under "**input_fsm**" directory.

Figure 9-4 Firm Synthesis Model Input Directory

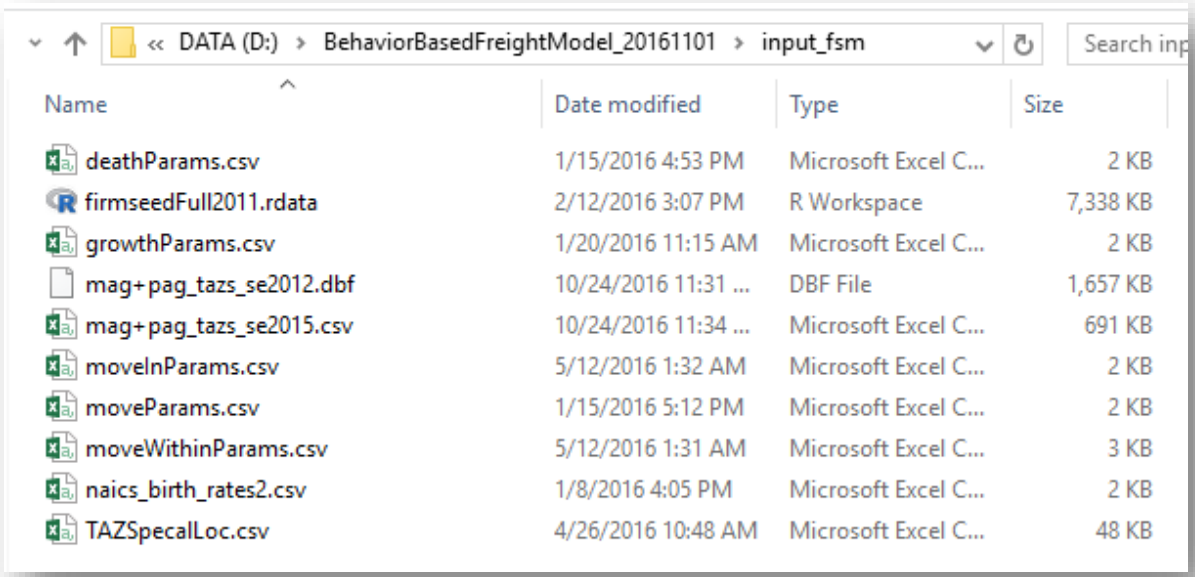


Table 9-4 Input Files of the Firm Synthesis Model

Input	Description
firmseedFull2011.rdata	Seed data (a set of all firms from a single year of the NETS database)
mag+pag_tazs_se2012.dbf	Socioeconomic data for the model input. *Note that the year of this data must be the same as the year of the NETS seed data. Due to data availability issue to PAG area, the project team decided to use the 2012 socioeconomic data as a proxy of 2011 socioeconomic data.*
mag+pag_tazs_se2015.csv	Control target socioeconomic data for post processing (e.g., 2015)
growthParams.csv	Parameters of firm growth model
deathParams.csv	Parameters of firm death model
moveInParams.csv	Parameters of firm move-in model
moveParams.csv	Parameters of firm move-out model
moveWithinParams.csv	Parameters of firm move-within modeling area
naics_birth_rates2.csv	Parameters of firm birth rate
TAZSpecalLoc.csv	Location index of special facility such as airport and major university

Table 9-5 shows the data structure of "**firmseedFull2011.rdata**" which is a set of all firms from a single year (2011) of the NETS database. This seed database was created for MAG and PAG area, Arizona.

Table 9-5 Firm Seed Data Structure

Field Name	Description
DunsNumber	NETS database IDs
firmID	Unique firm identifier (not DunsNumber)
currtaz	TAZ number of firm's location
NAICS2	2-digit NAICS code
EMP	Number of employees
Subsidiary	Indicates firm is subsidiary (1) or is not (0)
prevEmp	Number of employees in previous year
prevEmp2	Number of employees in two years prior
age	Age of firm in years
NAICS6	6-digit NAICS code
NAICS3	3-digit NAICS code
isN11	Firm is in Agriculture Industry (0/1)
isN21	Firm is in Natural Resources Industry (0/1)
isN22	Firm is in Utility Industry (0/1)
isN23	Firm is in Construction Industry (0/1)
isN31	Firm is in Manufacturing I Industry (0/1)
isN32	Firm is in Manufacturing II Industry (0/1)
isN33	Firm is in Manufacturing III Industry (0/1)
isN42	Firm is in Wholesale Industry (0/1)
isN44	Firm is in Retail I Industry (0/1)
isN45	Firm is in Retail II Industry (0/1)
isN48	Firm is in Transportation Industry (0/1)
isN49	Firm is in Transportation - Postal Industry (0/1)
isN51	Firm is in Information Industry (0/1)
isN52	Firm is in Finance and Insurance Industry (0/1)
isN53	Firm is in Real Estate Industry (0/1)
isN54	Firm is in Professional/Scientific/Technical Industry (0/1)
isN55	Firm is in Management Industry (0/1)
isN56	Firm is in Administrative/Support/Waste Management Industry (0/1)
isN61	Firm is in Educational Industry (0/1)
isN62	Firm is in Health Care Industry (0/1)
isN71	Firm is in Arts/Recreation Industry (0/1)
isN72	Firm is in Accommodation/Food Industry (0/1)
isN81	Firm is in Other Services, (except Public) Industry (0/1)
isN92	Firm is in Public Administration (0/1)
YearStart	Year of firm's birth

Table 9-6 shows the socioeconomic data structure for MAG and PAG area. "**mag+pag_tazs_se2012.dbf**" for the model input and "**mag+pag_tazs_se2015.csv**" for controlling the target year's employments have the same data structure.

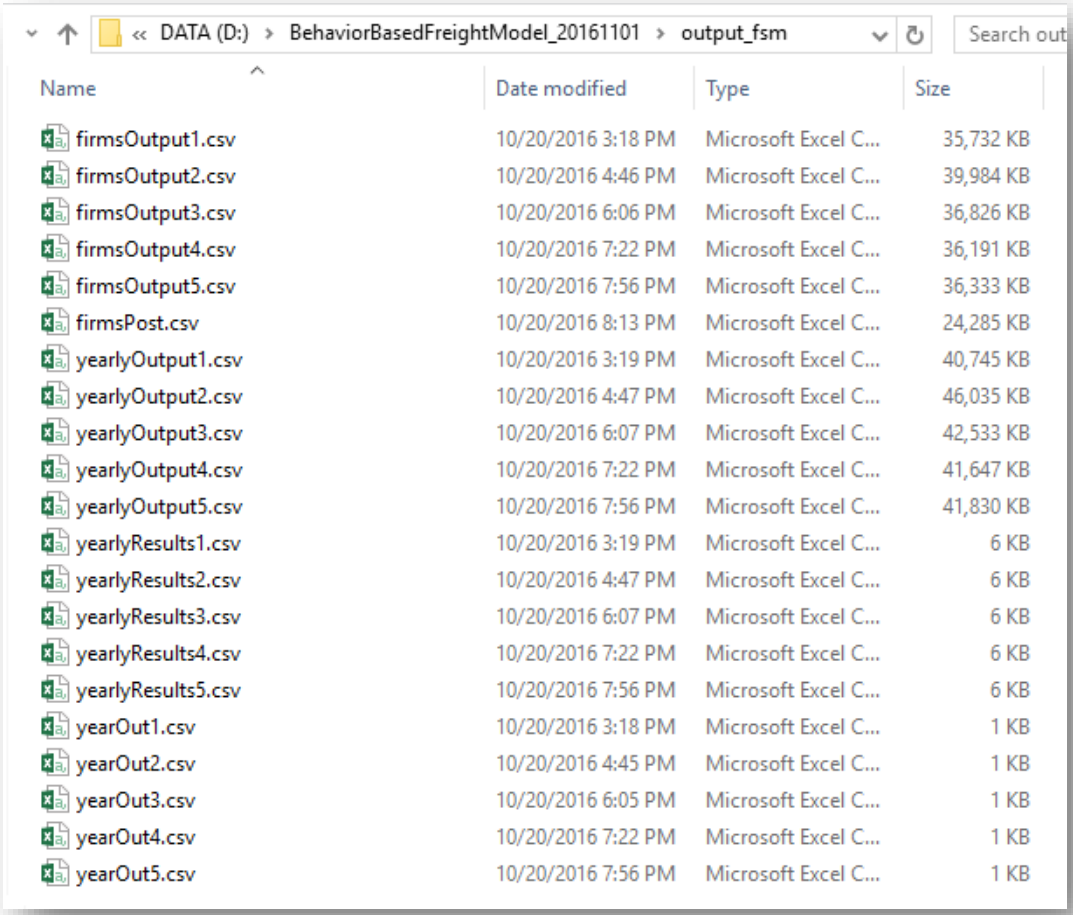
Table 9-6 Base Year Socioeconomic Data Structure

Field_Name	Description
TAZ	TAZ ID
MPA	MPA
MPA_FULLNA	MPA full name
COUNTY	County
COUNTY_NAM	County full name
MPO	MPO
YEAR	Data year
RESIDENT_P	Residential population
RESIDENT_H	Residential households
GQ_POP	Group quarter population
AVG_HH_INC	Average household income
K_8	K-8 enrollments
G9_12	Grade 9-12 enrollments
UOFA_ENROL	U of Arizona enrollments
NON_UOFA_E	Non U of Arizona enrollments
ASU_ENROLL	ASU enrollments
NON_ASU_EN	Non ASU enrollments
ASU_FLAG	ASU campus location flag
UOFA_FLAG	U of Arizona campus flag
AIRPORT_FL	Airport flag
N11	Employment in Agriculture Industry
N21	Employment in Natural Resources Industry
N22	Employment in Utility Industry
N23	Employment in Construction Industry
N31	Employment in Manufacturing I Industry
N32	Employment in Manufacturing II Industry
N33	Employment in Manufacturing III Industry
N42	Employment in Wholesale Industry
N44	Employment in Retail I Industry
N45	Employment in Retail II Industry
N48	Employment in Transportation Industry
N49	Employment in Transportation - Postal Industry
N51	Employment in Information Industry
N52	Employment in Finance and Insurance Industry
N53	Employment in Real Estate Industry
N54	Employment in Professional/Scientific/Technical Industry
N55	Employment in Management Industry
N56	Employment in Administrative/Support/Waste Management Industry
N61	Employment in Educational Industry
N62	Employment in Health Care Industry
N71	Employment in Arts/Recreation Industry
N72	Employment in Accommodation/Food Industry
N81	Employment in Other Services, (except Public) Industry
N92	Employment in Public Administration
art_mi	Arterial centerline-miles
colt_mi	Collector centerline-miles
fwy_mi	Freeway centerline-miles
AREA2	Area of TAZ (Square miles)

9.4.2 Firm Synthesis Model Outputs

Figure 9-5 and Table 9-7 show the firm synthesis model outputs from 5 simulation runs. "**firmsPost.csv**" is a final output created by the post processing procedure.

Figure 9-5 Firm Synthesis Model Output Directory



Name	Date modified	Type	Size
firmsOutput1.csv	10/20/2016 3:18 PM	Microsoft Excel C...	35,732 KB
firmsOutput2.csv	10/20/2016 4:46 PM	Microsoft Excel C...	39,984 KB
firmsOutput3.csv	10/20/2016 6:06 PM	Microsoft Excel C...	36,826 KB
firmsOutput4.csv	10/20/2016 7:22 PM	Microsoft Excel C...	36,191 KB
firmsOutput5.csv	10/20/2016 7:56 PM	Microsoft Excel C...	36,333 KB
firmsPost.csv	10/20/2016 8:13 PM	Microsoft Excel C...	24,285 KB
yearlyOutput1.csv	10/20/2016 3:19 PM	Microsoft Excel C...	40,745 KB
yearlyOutput2.csv	10/20/2016 4:47 PM	Microsoft Excel C...	46,035 KB
yearlyOutput3.csv	10/20/2016 6:07 PM	Microsoft Excel C...	42,533 KB
yearlyOutput4.csv	10/20/2016 7:22 PM	Microsoft Excel C...	41,647 KB
yearlyOutput5.csv	10/20/2016 7:56 PM	Microsoft Excel C...	41,830 KB
yearlyResults1.csv	10/20/2016 3:19 PM	Microsoft Excel C...	6 KB
yearlyResults2.csv	10/20/2016 4:47 PM	Microsoft Excel C...	6 KB
yearlyResults3.csv	10/20/2016 6:07 PM	Microsoft Excel C...	6 KB
yearlyResults4.csv	10/20/2016 7:22 PM	Microsoft Excel C...	6 KB
yearlyResults5.csv	10/20/2016 7:56 PM	Microsoft Excel C...	6 KB
yearOut1.csv	10/20/2016 3:18 PM	Microsoft Excel C...	1 KB
yearOut2.csv	10/20/2016 4:45 PM	Microsoft Excel C...	1 KB
yearOut3.csv	10/20/2016 6:05 PM	Microsoft Excel C...	1 KB
yearOut4.csv	10/20/2016 7:22 PM	Microsoft Excel C...	1 KB
yearOut5.csv	10/20/2016 7:56 PM	Microsoft Excel C...	1 KB

Table 9-7 Firm Synthesis Outputs

Output	Description
yearlyOutput1.csv, .., yearlyOutput5.csv	Firm synthesis yearly output from 5 simulation runs
yearlyResults1.csv, .., yearlyResults5.csv	Summary of yearly output by NAICS 2-digit from 5 simulation runs
yearOut1.csv, .., yearOut5.csv	Summary of yearly output from 5 simulation runs
firmsOutput1.csv, .., firmsOutput5.csv	Firm synthesis output from 5 simulation runs
firmsPost.csv	Final firm synthesis output from 5 firm outputs. This is an input of supply chain model.

9.4.3 Supply Chain Model Inputs

Table 9-8 shows the zone system for MAG and PAG, and other area used in the supply chain model. Users should update the zone system for their application.

Table 9-8 Zone system

Zone System	Description
101- 3140	MAG TAZs
5001 – 6104	PAG TAZs
7000 – 7764	Airports
8000 – 8602	Intermodal rail stations
9000 – 9012	AZ counties
9100 – 9220	FAF4 TAZs

Figure 9-6 and Table 9-9 show the list of supply chain model inputs under "**input_scm**" directory.

Figure 9-6 Supply Chain Model Input Directory

BehaviorBasedFreightModel_20161101 › input_scm

Name	Date modified	Type
CentroidsXml.csv	3/10/2016 7:13 AM	Microsoft Excel C...
corresp_io_to_sctg.csv	8/11/2015 2:03 PM	Microsoft Excel C...
corresp_naics_to_io.csv	8/11/2015 3:21 PM	Microsoft Excel C...
corresp_naics_to_sctg.csv	8/10/2015 7:53 PM	Microsoft Excel C...
corresp_naics2_to_desc.csv	8/26/2015 10:38 AM	Microsoft Excel C...
corresp_sctg_unitvalue_2007.csv	8/3/2015 12:43 PM	Microsoft Excel C...
corresp_sctg_unitvalue_az_cfs2012.csv	3/30/2016 3:56 PM	Microsoft Excel C...
corrsp_greatcircledistance_matrix.csv	6/3/2016 1:36 PM	Microsoft Excel C...
corrsp_inbound_extstation.csv	8/19/2016 4:29 PM	Microsoft Excel C...
corrsp_outbound_extstation.csv	8/19/2016 4:30 PM	Microsoft Excel C...
corrsp_railyards.csv	6/6/2016 12:10 PM	Microsoft Excel C...
corrsp_routeddistance_rail.csv	6/6/2016 10:23 AM	Microsoft Excel C...
corrsp_routeddistance_truck.csv	3/31/2016 12:15 PM	Microsoft Excel C...
deliveryindex.csv	3/10/2016 7:14 AM	Microsoft Excel C...
DistanceMatrix_Linear.csv	5/6/2016 3:23 PM	Microsoft Excel C...
emp_2007.csv	8/11/2015 3:25 PM	Microsoft Excel C...
faf_consumption.csv	8/12/2016 9:57 AM	Microsoft Excel C...
faf_production.csv	8/12/2016 10:29 AM	Microsoft Excel C...
faf_zone_desc.csv	8/26/2015 1:01 PM	Microsoft Excel C...
firms_external.csv	10/20/2016 1:26 PM	Microsoft Excel C...
io_make_detail.csv	8/12/2015 2:34 PM	Microsoft Excel C...
io_use_detail.csv	8/5/2015 5:49 PM	Microsoft Excel C...
MAG_domestic_flow.csv	8/12/2016 2:24 PM	Microsoft Excel C...
reginal_make_rate.csv	9/13/2016 1:44 PM	Microsoft Excel C...
reginal_use_rate.csv	9/13/2016 1:28 PM	Microsoft Excel C...
regional_sctg_price.csv	3/10/2016 7:15 AM	Microsoft Excel C...
zone_crosswalk.csv	3/10/2016 7:15 AM	Microsoft Excel C...

Table 9-9 Supply Chain Model Inputs

Input		Description
corresp_naics_to_sctg.csv		Correspondence table that link 6-digit NAICS codes to 2-digit SCTG codes
corresp_naics_to_io.csv		Correspondence file that link NAICS codes to IO industry codes. This file is used to determine IO industry code for each agent in the firm from previous step
corresp_naics2_to_desc.csv		2-digit NAICS code description
corresp_io_to_sctg.csv		Correspondence table that links 6-digit IO industry classes to 2-digit NAICS code
corresp_sctg_unitvalue_2007.csv		Correspondence between SCTG unite value (dollar per pound)
corresp_sctg_unitvalue_az_cfs2012.csv		SCTG value per pound based on CFS-Micro data flows to/from/within Arizona 2012
corrsp_greatcircledistance_matrix.csv		Distance matrix that is used for supplier selection and in mode choice (for air/parcel alternative). It is the great circle distance between ODs
corrsp_inbound_extstation.csv		Inbound external station of Origin FAF4 zone, origin FAF4 TAZ, Destination FAF4 zone
corrsp_outbound_extstation.csv		outbound external station of Origin FAF4 zone, origin FAF4 TAZ, Destination FAF4 zone
corrsp_railyards.csv		Access and egress rail stations for each origin and destination pair
corrsp_routeddistance_rail.csv		Rail distance for each origin and destination pair
corrsp_routeddistance_truck.csv		Truck distance for each origin and destination pair
emp_2007.csv		National employment data obtained from CBP for the same year that BEA industry IO tables is available. This data is used to calculate rates per employee
faf_production.csv		FAF zone level total annual productions obtained from FAF4
faf_consumption.csv		FAF zone level total annual consumption obtained from FAF4
faf_zone_desc.csv		FAF zone description
firms_external.csv		External firm data from CBP 2013
io_use_detail.csv		Latest (2007) benchmark industry use table (after redefinition) obtained from BEA with detailed industry classifications (6-digit 389 classes). This is used for calculation of consumption rates
io_make_detail.csv		Latest (2007) benchmark industry make table (after redefinition) obtained from BEA with detailed industry classifications (6-digit 389 classes). This is used for calculation production rates.
regional_make_rate.csv		Regional make rates per employee from IMPLAN
regional_use_rate.csv		Regional use rates per employee from IMPLAN
Supplier selection model inputs	CentroidsXml.csv	Latitude and longitude of TAZ location
	DistanceMatrix_Linear.csv	Linear distance between TAZs (feet)
	MAG_domestic_flow.csv	Domestic flows between MAG/PAG and U.S. area
	regional_sctg_price.csv	Region specific price by SCTG
	zone_crosswalk.csv	Crosswalk between MAG/PAG TAZ and FAF zone

9.4.4 Supply Chain Model Outputs

Figure 9-7 shows an example of the supply chain model outputs under "**output_scm**" directory. The main output of this model is "**CommodityFlows_firmlevel.csv**" which is an input of the truck tour model.

Figure 9-7 Supply Chain Output Directory

DATA (D:) › BehaviorBasedFreightModel_20161101 › output_scm

Name	Date modified	Type	Size
BuyersResults_42_39.csv	9/30/2016 12:34 PM	Microsoft Excel C...	2,830 KB
BuyersResults_42_40.csv	9/30/2016 12:35 PM	Microsoft Excel C...	3,566 KB
BuyersResults_42_41.csv	9/30/2016 12:35 PM	Microsoft Excel C...	83 KB
BuyersResults_42_43.csv	9/30/2016 12:35 PM	Microsoft Excel C...	22 KB
CalibrationComparisonDetailed_1_43.csv	9/30/2016 12:37 PM	Microsoft Excel C...	233 KB
CalibrationComparisonTotals_1_43.csv	9/30/2016 12:37 PM	Microsoft Excel C...	4 KB
CommodityFlows_firmlevel.csv	9/30/2016 12:50 PM	Microsoft Excel C...	346,429 KB
employment_by_naics2.csv	10/21/2016 9:51 PM	Microsoft Excel C...	2 KB
employment_MAG_PAG_by_naics2.csv	10/21/2016 9:51 PM	Microsoft Excel C...	2 KB
firms_all.csv	10/21/2016 9:51 PM	Microsoft Excel C...	8,731 KB
industrydemand.csv	10/21/2016 9:53 PM	Microsoft Excel C...	67 KB
industrysupply.csv	10/21/2016 9:53 PM	Microsoft Excel C...	18 KB
ModelsAgentData_41_1.csv	10/20/2016 3:14 PM	Microsoft Excel C...	810 KB
ModelsAgentData_41_2.csv	10/20/2016 3:10 PM	Microsoft Excel C...	1,470 KB
ModelsAgentData_41_3.csv	9/30/2016 10:53 AM	Microsoft Excel C...	3,329 KB
ModelsAgentData_41_4.csv	9/30/2016 10:55 AM	Microsoft Excel C...	917 KB
ModelsAgentData_41_5.csv	9/30/2016 10:55 AM	Microsoft Excel C...	1,632 KB

Table 9-10 shows the commodity flows between TAZ pairs and FAF zone pairs by each commodity category.

Table 9-10 CommodityFlows_firmlevel.csv

Attribute	Description
sctg	SCTG code
orig_mag_zone	Supplier TAZ ID
orig_faf	Supplier FAZ zone
orig_naics	Supplier 6-digit NAICS code
dest_mag_zone	Buyer TAZ ID
dest_faf	Buyer FAZ zone
dest_naics	Buyer 6-digit NAICS code
flow_ktns	Total commodity flow (Kilo tons)
Rail_Large	Commodity flow by rail ship size (1,500-34,999 lbs)
Rail_VeryLarge	Commodity flow by rail ship size (35,000+ lbs)
Truck_Small	Commodity flow by truck ship size (<150 lbs)
Truck_Medium	Commodity flow by truck ship size (150-1,499 lbs)
Truck_Large	Commodity flow by truck ship size (1500-34,999 lbs)
Truck_VeryLarge	Commodity flow by truck ship size (35,000+ lbs)
ParcelAir_Small	Commodity flow by air ship size (<150 lbs)
ParcelAir_Medium	Commodity flow by air ship size (150-1,499 lbs)
ParcelAir_Large	Commodity flow by air ship size (1500-34,999 lbs)
AccessRail_TAZ	Rail accessible TAZ ID
EgressRail_TAZ	Rail egressible TAZ ID
ExternalStation_inbound	Inbound external station
ExternalStation_outbound	Outbound external station

Cleaned and classified firm clusters are in **firms_all.csv**. Attributes included in the output file (**firms_all.csv**) are described in Table 9-11.

Table 9-11 firm_all.csv attributes

Attribute	Description
MAG_ZONE	MAG TAZ ID
FAF_ZONE	Corresponding FAF3 Zone ID
naics	6-digit NAICS code
empsize	Employee Size Group
emp	Total number of employees in the firm cluster
est	Total number of establishments in the firm cluster

io_emp_after_merge_2007.csv employment information by 2-digit NAICS industry class for base year 2007.

io_make_2007.csv produced values by IO industry class

Table 9-12 io_make_2007.csv attributes

Attribute	Description
io	Industry class (BEA IO code system)
Commodity	Made commodity (BEA IO code system)
value	Total annual value made (output) $million

io_use_2007.csv consumed values by IO industry class

Table 9-13 io_use_2007.csv attributes

Attribute	Description
io	Industry class (BEA IO code system)
Commodity	Used commodity (BEA IO code system)
value	Total annual value used (input) $million

BuyerFirms.csv is one of the main outputs of the supply chain model. That include required attributes for running the supplier selection model as described in the following table.

Table 9-14 BuyerFirms.csv data table attributes

Attribute	Description
MAG_Zone	MAG TAZ ID
FAF_Zone	FAF Zone ID
io	IO industry code
naics	6-digit NAICS code
Naics2	2-digit NAICS code
empsize	Employee size group
employment	Number of employees (Do not use this field for analysis)
establishments	Number of establishments (Do not use this field for analysis)
sctg	Consumed SCTG code
demandkton	Total annual consumption (kilo-ton)

Several intermediates output files are generated for model performance check:

SupplierFirms.csv: list of supplier agents with their annual supply.

Table 9-15 SupplierFirms data table attributes

Attribute	Description
MAG_Zone	MAG TAZ ID
FAF_Zone	FAF Zone ID
io	IO industry code
naics	6-digit NAICS code
naics2	2-digit NAICS code (used for summary report)
empsize	Employee size group
employment	Number of employees
establishments	Number of establishments
sctg	Produced SCTG code
supplykton	Total annual supply (Kilo-ton)

zonesupply.csv: aggregated zone level supply

Table 9-16 zonesupply.csv data table attributes

Attribute	Description
FAF_Zone	FAF Zone ID
sctg	Commodity type
Supply_Kton	Total annual supply (Kilo-ton)

zonedemand.csv: aggregated zone level demand

Table 9-17 zonedemand.csv data table attributes

Attribute	Description
FAF_Zone	FAF Zone ID
sctg	Commodity type
Demand_Kton	Total annual demand

industrysupply.csv: aggregated supply by FAF zone and by 2-digit industry type (NAICS 2)

Table 9-18 industrysupply.csv data table attributes

Attribute	Description
FAF_Zone	FAF Zone ID
naics2	2-digit NAICS code
Supply_Kton	Total annual Supply

industrydemand.csv: aggregated demand by FAF zone and by 2-digit industry class (NAICS 2)

Table 9-19 industrydemand.csv data table attributes

Attribute	Description
FAF_Zone	FAF Zone ID
naics2	2-digit NAICS code
Demand_Kton	Total annual Demand

totalsupply.csv: total supply by commodity type

Table 9-20 totalsupply.csv data table attributes

Attribute	Description
sctg	Commodity type
Supply_Kton	Total annual supply

totaldemand.csv: total demand by commodity type

Table 9-21 totaldemand.csv data table attributes

Attribute	Description
sctg	Commodity type
Demand_Kton	Total annual Demand

annualtonnage_firmlevel.csv: commodity flows between supplier and buyer firms to/from and within MAG-PAG region.

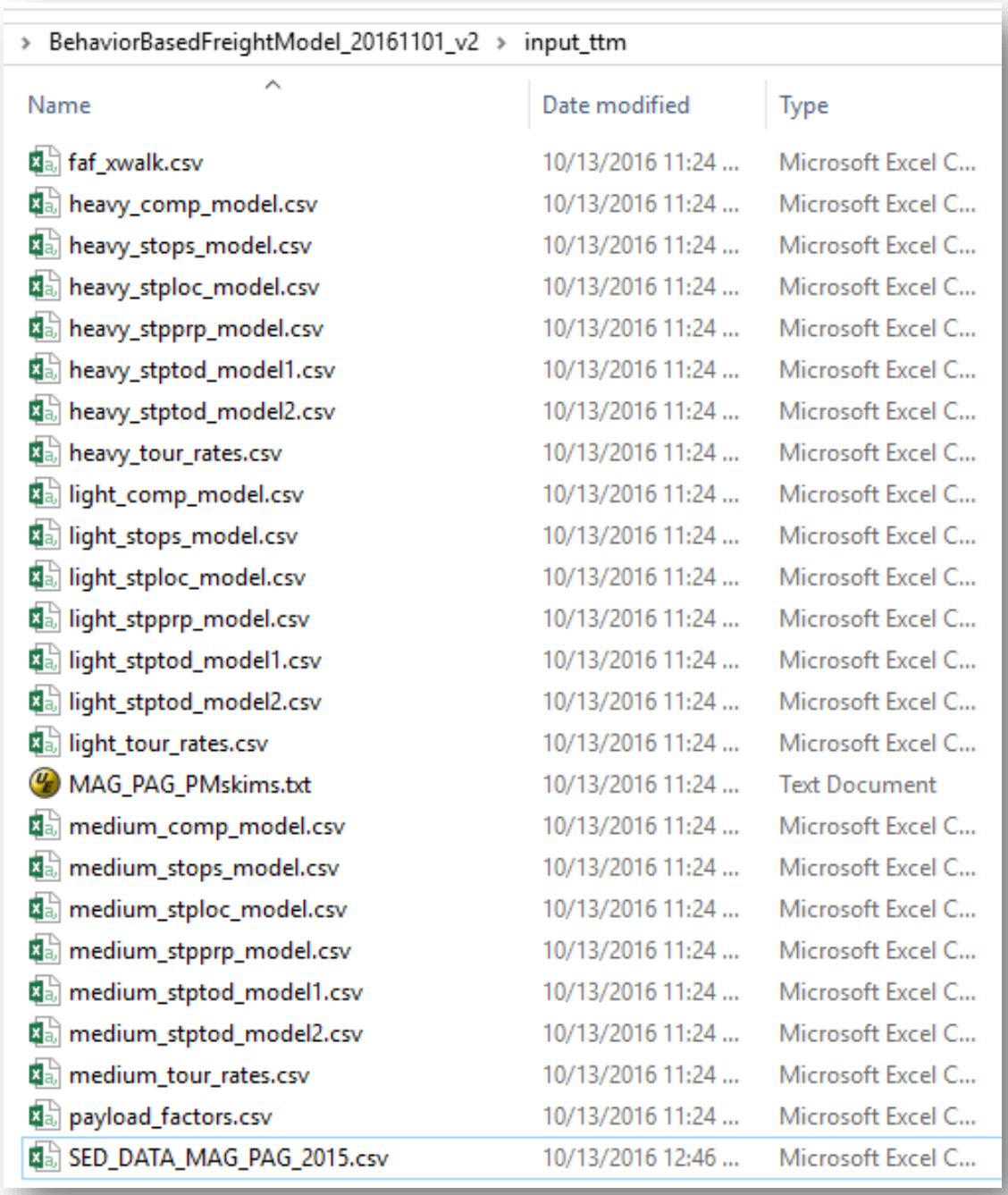
Table 9-22 annualtonnage_firmlevel.csv data table attributes

Attribute	Description
sctg	commodity type
orig_naics	Supplier Industry class (NAICS 6-digit)
orig_mag_zone	Supplier MAG zone
orig_faf	Supplier FAF zone
orig_empSize	Supplier employee size category
orig_employees	Supplier employee#
orig_estabs	Number of establishments at supplier firm cluster
dest_naics	Buyer Industry Class
dest_mag_zone	Buyer MAG zone
dest_faf	Buyer FAF zone
dest_empSize	Buyer employee size category
dest_employees	Buyer employee#
dest_estabs	Number of establishments at buyer firm cluster
flow_ktns	total annual commodity flow between supplier and buyer (Kilo-ton)
export_canada_kt	total export amount to Canada (Kilo-ton)
export_mexico_kt	total export amount to Mexico (Kilo-ton)
export_other_kt	total export amount to other countries (Kilo-ton)

9.4.5 Truck Tour Model Inputs

Figure 9-8 shows the truck tour model input directory and Table 9-23 is a list of input files under "**input_ttm**" directory and its description.

Figure 9-8 Truck tour model input directory



BehaviorBasedFreightModel_20161101_v2 › input_ttm

Name	Date modified	Type
faf_xwalk.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_comp_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_stops_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_stploc_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_stpprp_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_stptod_model1.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_stptod_model2.csv	10/13/2016 11:24 ...	Microsoft Excel C...
heavy_tour_rates.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_comp_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_stops_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_stploc_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_stpprp_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_stptod_model1.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_stptod_model2.csv	10/13/2016 11:24 ...	Microsoft Excel C...
light_tour_rates.csv	10/13/2016 11:24 ...	Microsoft Excel C...
MAG_PAG_PMskims.txt	10/13/2016 11:24 ...	Text Document
medium_comp_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
medium_stops_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
medium_stploc_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
medium_stpprp_model.csv	10/13/2016 11:24 ...	Microsoft Excel C...
medium_stptod_model1.csv	10/13/2016 11:24 ...	Microsoft Excel C...
medium_stptod_model2.csv	10/13/2016 11:24 ...	Microsoft Excel C...
medium_tour_rates.csv	10/13/2016 11:24 ...	Microsoft Excel C...
payload_factors.csv	10/13/2016 11:24 ...	Microsoft Excel C...
SED_DATA_MAG_PAG_2015.csv	10/13/2016 12:46 ...	Microsoft Excel C...

Table 9-23 Truck Tour Model Inputs

Input	Description
faf_xwalk.csv	MAG and PAG external station crosswalk table
heavy_comp_model.csv	Heavy truck tour complete model parameters
heavy_stops_model.csv	Heavy truck tour stop frequency model parameters
heavy_stploc_model.csv	Heavy truck tour stop location model parameters
heavy_stpprp_model.csv	Heavy truck tour stop purpose model parameters
heavy_stptod_model1.csv	Heavy truck tour first stop TOD model parameters
heavy_stptod_model2.csv	Heavy truck tour stop TOD model parameters after first stop
heavy_tour_rates.csv	Heavy truck tour generation rate model parameters
light_comp_model.csv	Light truck tour complete model parameters
light_stops_model.csv	Light truck tour stop frequency model parameters
light_stploc_model.csv	Light truck tour stop location model parameters
light_stpprp_model.csv	Light truck tour stop purpose model parameters
light_stptod_model1.csv	Light truck tour first stop TOD model parameters
light_stptod_model2.csv	Light truck tour stop TOD model parameters after first stop
light_tour_rates.csv	Light truck tour generation rate model parameters
MAG_PAG_PMskims.txt	PM peak highway time skim for MAG and PAG
medium_comp_model.csv	Medium truck tour complete model parameters
medium_stops_model.csv	Medium truck tour stop frequency model parameters
medium_stploc_model.csv	Medium truck tour stop location model parameters
medium_stpprp_model.csv	Medium truck tour stop purpose model parameters
medium_stptod_model1.csv	Medium truck tour first stop TOD model parameters
medium_stptod_model2.csv	Medium truck tour stop TOD model parameters after first stop
medium_tour_rates.csv	Medium truck tour generation rate model parameters
payload_factors.csv	The carrying capacity of medium and heavy trucks
SED_DATA_MAG_PAG_2015.csv	MAG and PAG socioeconomic data for 2015. This can be any forecasting socioeconomic data.

9.4.6 Truck Tour Model Outputs

Figure 9-9 shows the truck tour model output directory and Table 9-24 is a list of output files under "**output_ttm**" directory and its description.

Figure 9-9 Truck Tour Model Output Directory

« DATA (D:) › BehaviorBasedFreightModel_20161101 › output_ttm

Name	Date modified	Type	Size
am_hv_trips.csv	10/13/2016 4:34 PM	Microsoft Excel C...	33,564 KB
am_lt_trips.csv	10/13/2016 3:43 PM	Microsoft Excel C...	33,568 KB
am_md_trips.csv	10/13/2016 4:22 PM	Microsoft Excel C...	33,565 KB
heavy_ext_truck_trips.csv	10/13/2016 1:31 PM	Microsoft Excel C...	34,780 KB
hv_tours_output.csv	10/13/2016 4:32 PM	Microsoft Excel C...	2,179 KB
lt_tours_output.csv	10/13/2016 3:41 PM	Microsoft Excel C...	101,184 KB
md_hv_trips.csv	10/13/2016 4:36 PM	Microsoft Excel C...	33,564 KB
md_lt_trips.csv	10/13/2016 3:45 PM	Microsoft Excel C...	33,575 KB
md_md_trips.csv	10/13/2016 4:24 PM	Microsoft Excel C...	33,566 KB
md_tours_output.csv	10/13/2016 4:20 PM	Microsoft Excel C...	14,443 KB
medium_ext_truck_trips.csv	10/13/2016 1:29 PM	Microsoft Excel C...	34,784 KB
nt_hv_trips.csv	10/13/2016 4:40 PM	Microsoft Excel C...	33,564 KB
nt_lt_trips.csv	10/13/2016 3:49 PM	Microsoft Excel C...	33,570 KB
nt_md_trips.csv	10/13/2016 4:28 PM	Microsoft Excel C...	33,565 KB
pm_hv_trips.csv	10/13/2016 4:38 PM	Microsoft Excel C...	33,564 KB
pm_lt_trips.csv	10/13/2016 3:47 PM	Microsoft Excel C...	33,570 KB
pm_md_trips.csv	10/13/2016 4:26 PM	Microsoft Excel C...	33,565 KB
sc_truck_tours.csv	10/13/2016 1:26 PM	Microsoft Excel C...	506 KB

Table 9-24 Truck Tour Model Outputs

Input	Description
am_hv_trips.csv	AM heavy truck trip table
am_lt_trips.csv	AM light truck trip table
am_md_trips.csv	AM medium truck trip table
heavy_ext_truck_trips.csv	Heavy external truck trip table
hv_tours_output.csv	Heavy truck tour table
lt_tours_output.csv	Light truck tour table
md_hv_trips.csv	Midday heavy truck trip table
md_lt_trips.csv	Midday light truck trip table
md_md_trips.csv	Midday medium truck trip table
md_tours_output.csv	Medium truck tour table
medium_ext_truck_trips.csv	Medium external truck trip table
nt_hv_trips.csv	Night heavy truck trip table
nt_lt_trips.csv	Night light truck trip table
nt_md_trips.csv	Night medium trip table
pm_hv_trips.csv	PM heavy truck trip table
pm_lt_trips.csv	PM light truck trip table
pm_md_trips.csv	PM medium truck trip table
sc_truck_tours.csv	Supply chain truck tour table

9.4.7 Program Directory

Figure 9-10 shows "**programs**" directory which contains all the model scripts such as firm synthesis model, supply chain model, and truck tour models. Model script and properties file naming conventions are as follows:

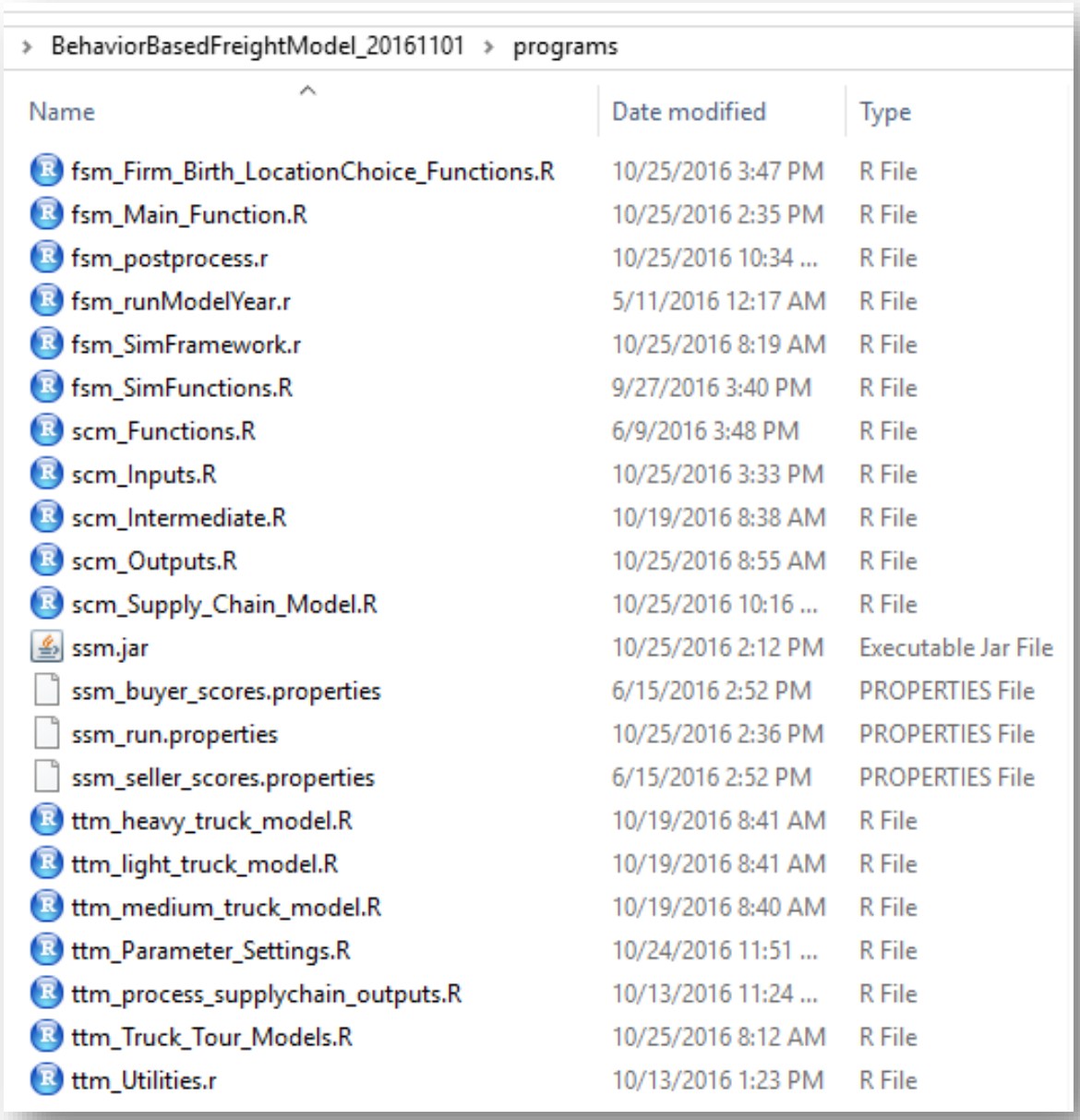
(1) fsm_*: firm synthesis model script,

(2) scm_*: supply chain model script,

(3) ssm.jar or ssm_*: supply selection model executable file and its properties, and

(4) ttm_*: truck tour model script.

Figure 9-10 Model Scripts and Properties Files in Programs Directory

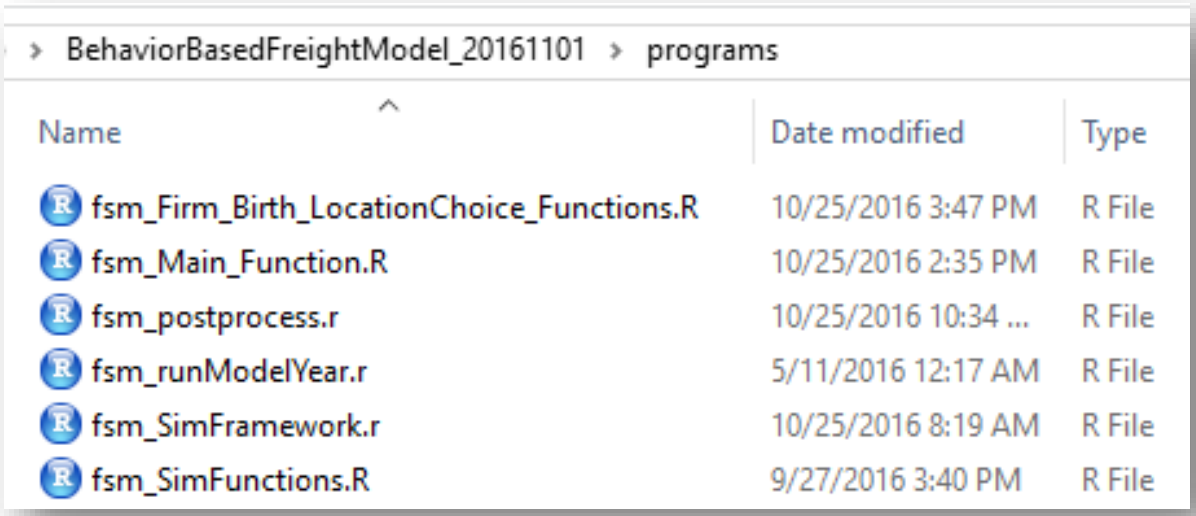


BehaviorBasedFreightModel_20161101 › programs

Name	Date modified	Type
fsm_Firm_Birth_LocationChoice_Functions.R	10/25/2016 3:47 PM	R File
fsm_Main_Function.R	10/25/2016 2:35 PM	R File
fsm_postprocess.r	10/25/2016 10:34 ...	R File
fsm_runModelYear.r	5/11/2016 12:17 AM	R File
fsm_SimFramework.r	10/25/2016 8:19 AM	R File
fsm_SimFunctions.R	9/27/2016 3:40 PM	R File
scm_Functions.R	6/9/2016 3:48 PM	R File
scm_Inputs.R	10/25/2016 3:33 PM	R File
scm_Intermediate.R	10/19/2016 8:38 AM	R File
scm_Outputs.R	10/25/2016 8:55 AM	R File
scm_Supply_Chain_Model.R	10/25/2016 10:16 ...	R File
ssm.jar	10/25/2016 2:12 PM	Executable Jar File
ssm_buyer_scores.properties	6/15/2016 2:52 PM	PROPERTIES File
ssm_run.properties	10/25/2016 2:36 PM	PROPERTIES File
ssm_seller_scores.properties	6/15/2016 2:52 PM	PROPERTIES File
ttm_heavy_truck_model.R	10/19/2016 8:41 AM	R File
ttm_light_truck_model.R	10/19/2016 8:41 AM	R File
ttm_medium_truck_model.R	10/19/2016 8:40 AM	R File
ttm_Parameter_Settings.R	10/24/2016 11:51 ...	R File
ttm_process_supplychain_outputs.R	10/13/2016 11:24 ...	R File
ttm_Truck_Tour_Models.R	10/25/2016 8:12 AM	R File
ttm_Utilities.r	10/13/2016 1:23 PM	R File

Figure 9-11 shows the firm synthesis model scripts under "**programs**" folder.

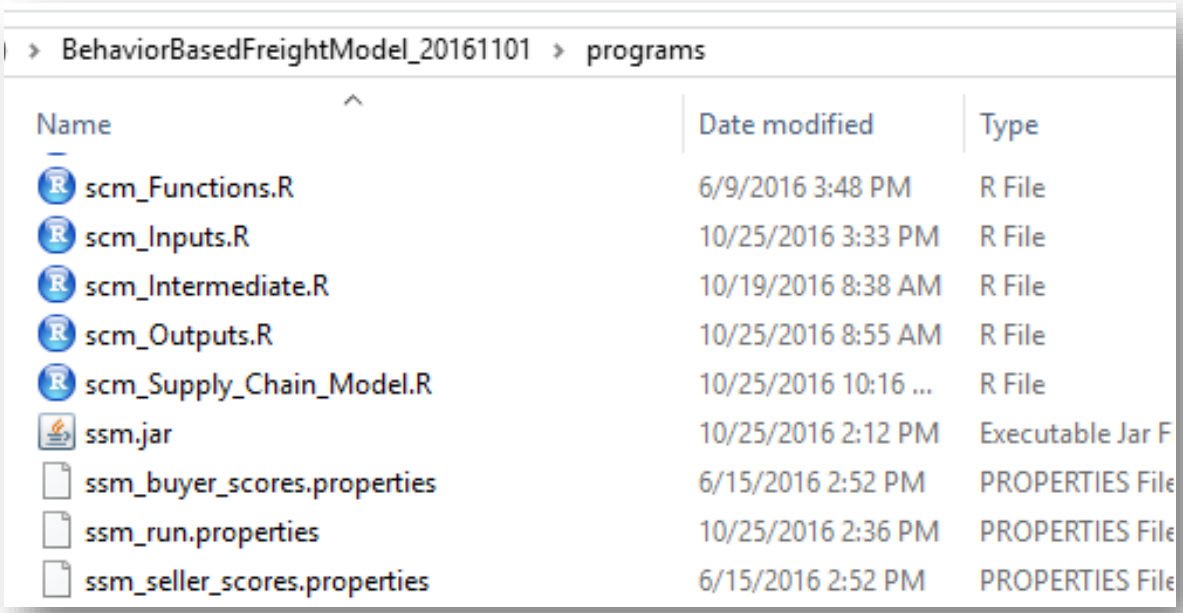
Figure 9-11 Firm Synthesis Model Scripts (fsm_xxxxxxxxxx.R)



- **fsm_Firm_Birth_LocationChoice_Functions.R**: This contains other functions specifically related to location choice and firm births.
- **fsm_Main_Function.R**: This is a main firm synthesis model script. Forecasting year value will be taken from "YEAR" field of "mag+pag_tazs_se2015.csv" for example.
- **fsm_postprocess.r**: This script is creating firmsPost.csv by adjusting 5 simulation run outputs using the forecasting year socioeconomic control data
- **fsm_runModelYear.r**: This script is called once for every year of the simulation. It sets up and executes each submodel.
- **fsm_SimFramework.r**: This file will do an entire simulation, including loading inputs, running the given number of years, and creating 5 output files.
- **fsm_SimFunctions.R**: This contains functions used repeatedly throughout the program.

Figure 9-12 shows the supply chain model scripts, supplier selection executable file, and its properties files under "programs" directory.

Figure 9-12 Supply chain model scripts (scm_xxxxxxxxxx.R, ssm.jar, ssm_xxxxxx.properties)



- **scm_Supply_Chain_Model.R**: Supply chain model main code stream.
  - ✓ Important note: Default memory settings in the script **"java –Xms48G –Xmx48G -Xss100M -jar ssm.jar ssm_run.properties"**. User might need to find the optimal memory settings. Minimum memory requirement is determined by **memory.limit** function in R script.
- **scm_Functions.R**: Include all libraries, user defined functions and scripts used in the main code stream.
- **scm_Inputs.R**: Input files are called from here.
- **scm_Intermediate.R**: Generate the intermediate files. Users need to change "intrmoutput = FALSE" into "intrmoutput = TRUE" to see intermediate outputs.
- **scm_Outputs.R**: Outputs are written to the Output directory from here.
- **ssm.jar**: Suppler selection model (ssm) Java Archive (executable) file
- **ssm_run.properties**: Define the path of the ssm model inputs and outputs. Users may need to read and revise this file for their application.
- **ssm_buyer_scores.properties**: Buyer score properties. Users may use this as a default setting.
- **ssm_seller_scores.properties**: Seller score properties. Users may use this as a default setting.

Figure 9-13 shows the truck tour model scripts under "programs" directory.

Figure 9-13 Truck tour model scripts (ttm_xxxxxxxxxx.R)

> BehaviorBasedFreightModel_20161101 > programs

Name	Date modified	Type
ttm_heavy_truck_model.R	10/19/2016 8:41 AM	R File
ttm_light_truck_model.R	10/19/2016 8:41 AM	R File
ttm_medium_truck_model.R	10/19/2016 8:40 AM	R File
ttm_Parameter_Settings.R	10/24/2016 11:51 ...	R File
ttm_process_supplychain_outputs.R	10/13/2016 11:24 ...	R File
ttm_Truck_Tour_Models.R	10/25/2016 8:12 AM	R File
ttm_Utilities.r	10/13/2016 1:23 PM	R File

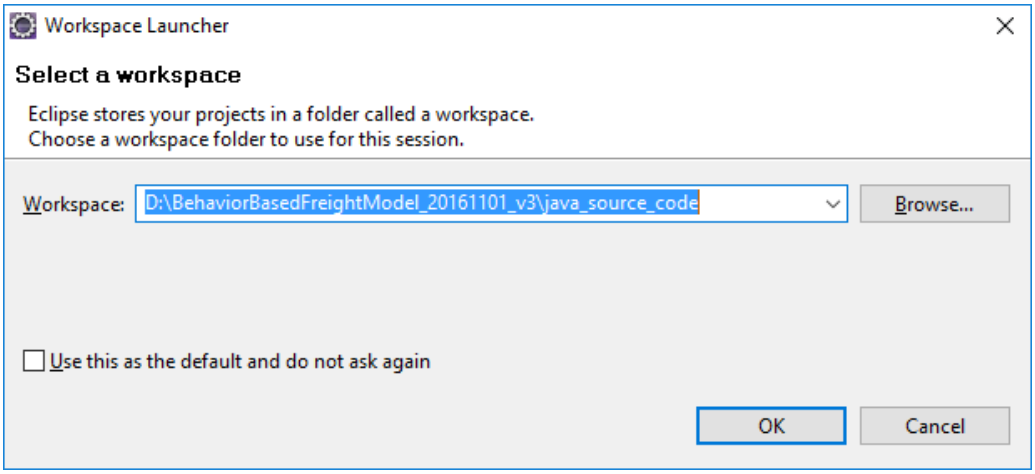
- **ttm_heavy_truck_model.R**: This is the heavy truck tour model script and generates heavy truck tour and trip table by time period.
- **ttm_light_truck_model.R**: This is the light truck tour model script and generates light truck tour and trip table by time period.
- **ttm_medium_truck_model.R**: This is the medium truck tour model script and generates medium truck tour and trip table by time period.
- **ttm_Parameter_Settings.R**: This scripts take model input parameters, socioeconomic data, supply chain model output, skims, payload factors, and external station id
- **ttm_process_supplychain_outputs.R**: This script is to process the supply chain output.
- **ttm_Truck_Tour_Models.R**: This is a main truck tour model script.
- **ttm_Utilities.R**: This includes all functions such as (1) Read SED data, (2) Read Commodity Flow Inputs, (3) Assign OD Industry Sector Codes, (4) Assign Shipment Industry Sector Code, (5) Determine Shipment Type, (6) Determine Truck Type, (7) Conversion to Daily Trucks, (8) Truck Tours Generated by Supply Chains, and (9) Generate External Truck Trip Matrices

9.5 How to Set Up the Supplier Selection Model Workspace and Project Folder in Eclipse

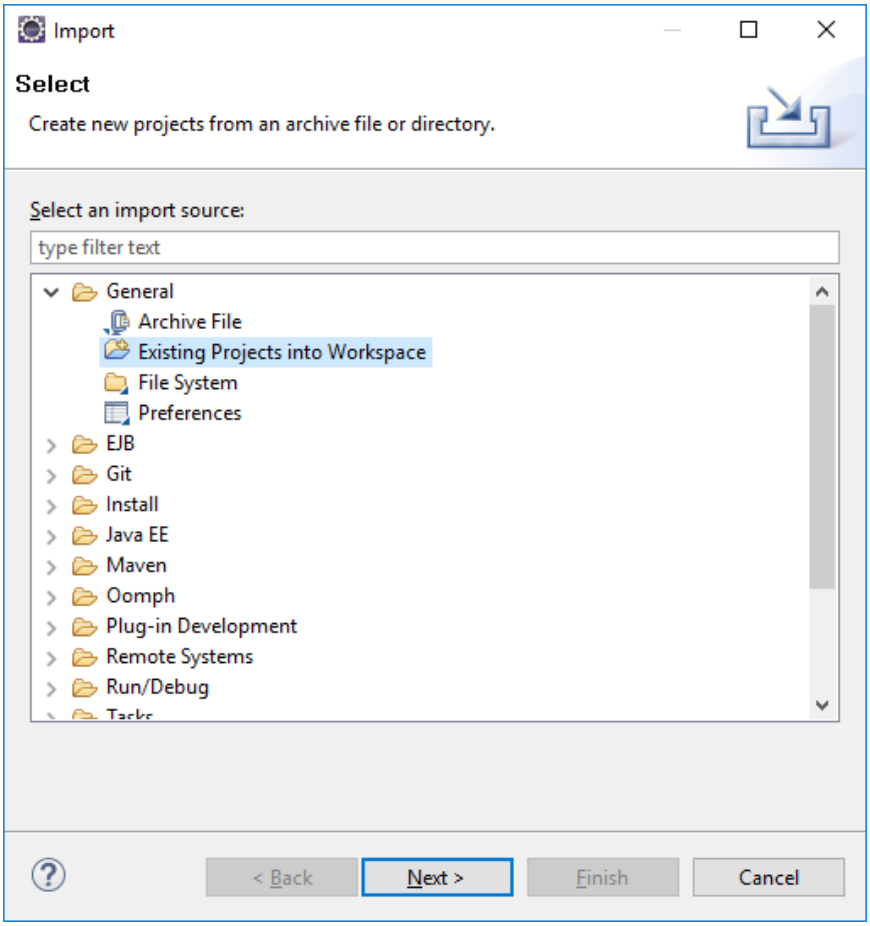
This section is to show how to set up the supplier selection model workspace and project folder in ECLIPSE. The steps to open the project in ECLIPSE are listed such as below:

(1) Run **Eclipse** and select **Workspace** (see Figure below)

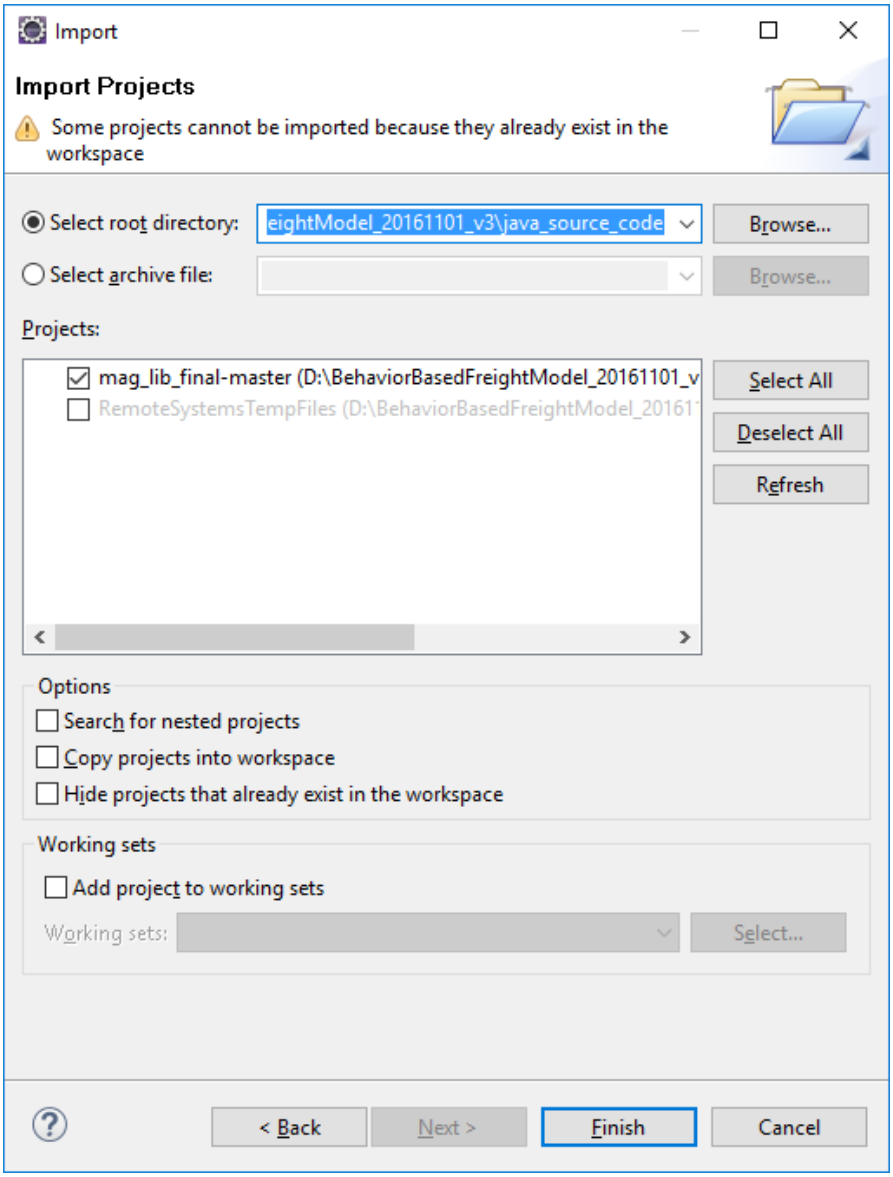
Note: **Workspace** should be the folder that include the supplier selection model project folder. For example, suppose "**mag_lib_final-master**" is the project folder. If this project folder is located under the folder "**java_source_code**", **Workspace** should be "**java_source_code**"



(2) Go to **File** -> **Import** and then select **Existing Projects into Workspace** under **General** and click **Next>** button (see Figure below)



(3) Select **Select root directory:**, click **Browse…** button, and select the **Workspace** folder. Then the project name will be shown under **Projects:** box (see Figure below)



(4) Check combo box in **Project** box and click **Finish** button

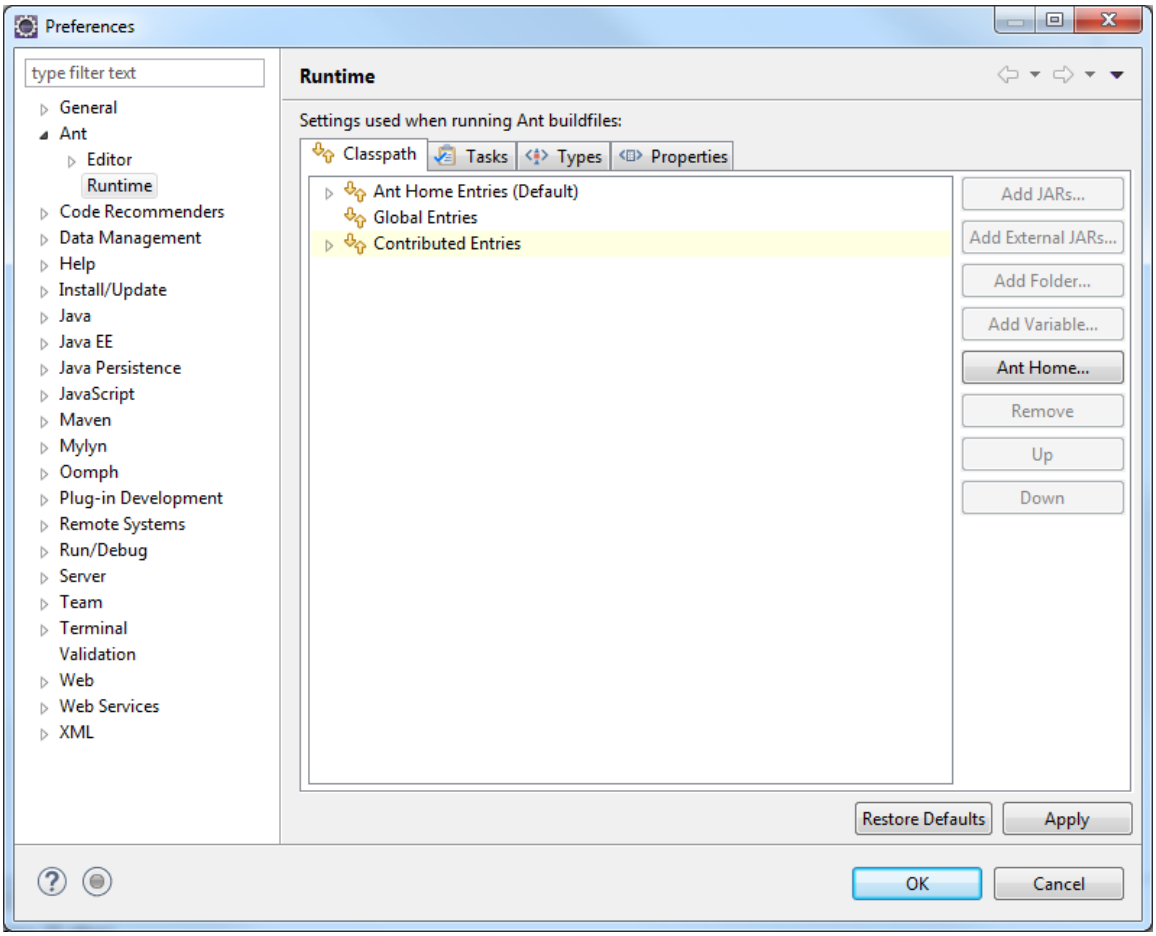
9.6 Eclipse Settings for Building JAR File

This section is to illustrate how to create JAR file in ECLIPSE. Supplier selection model was developed using Java. The Java codes of the model should be wrapped up as JAR file for convenient use and simulation of supplier selection model. That is, a JAR file is only needed to simulate the model without all Java scripts. The steps to create new JAR file for supplier selection model are listed such as below:

(1) Open **Eclipse**

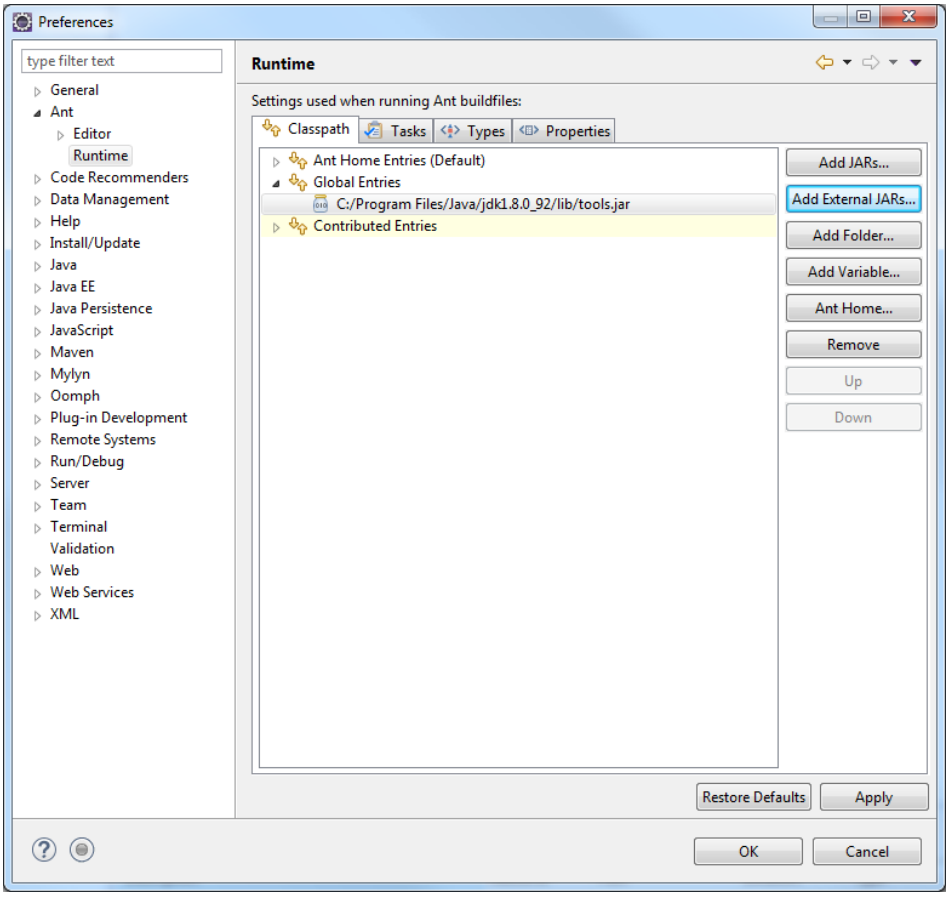
(2) **Window** -> **Preferences**

(3) Expand **Ant** and then click **Runtime** (see Figure below)



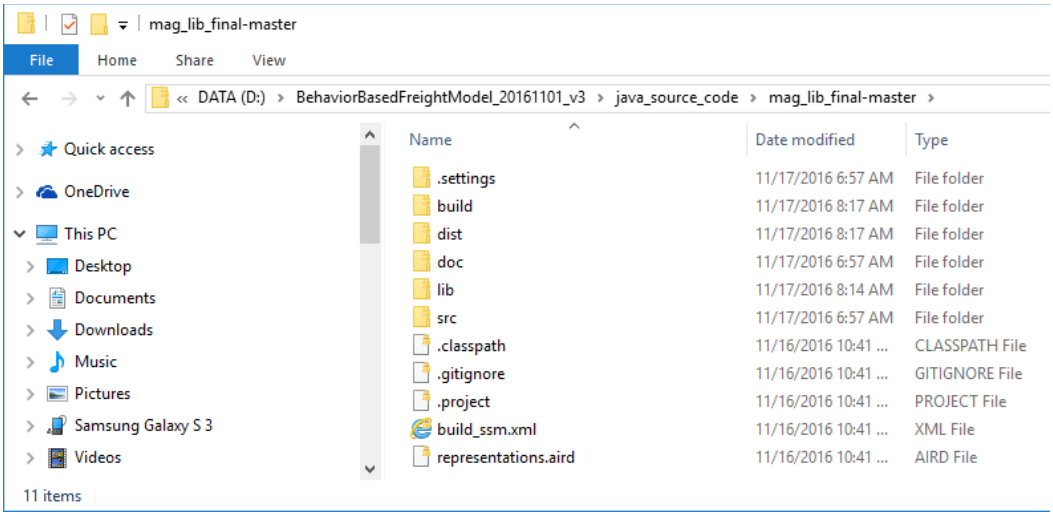
(5) Click **Classpath** tab and then select **Global Entries**

(6) Click **Add External JARs…** button and Add **tools.jar** from the folder "C:/Program Files/Java/jdk1.8.0_92/lib" (see Figure below)

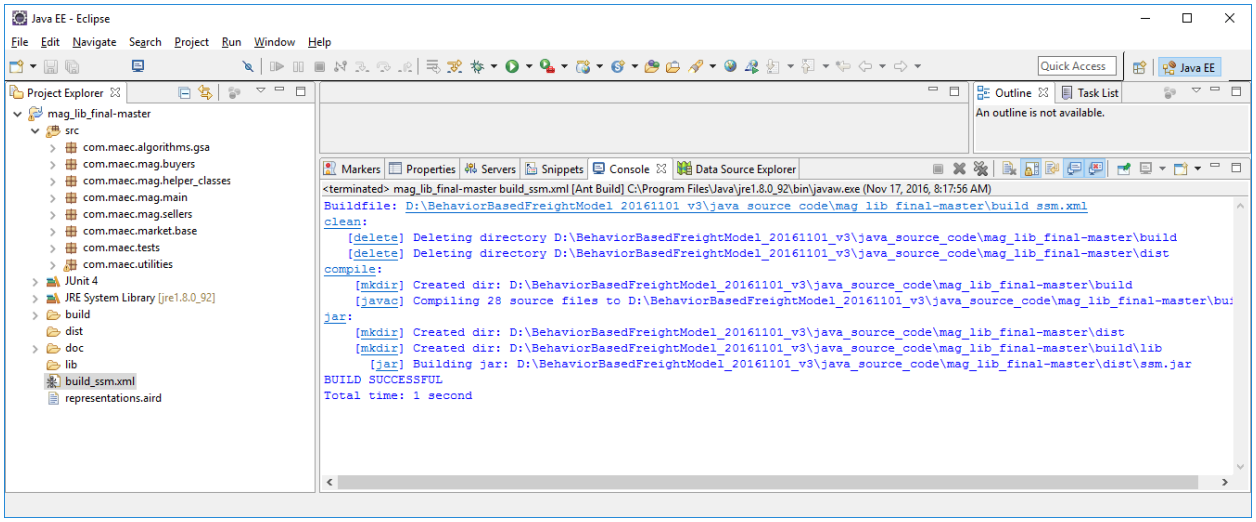


(7) Click **Apply** button and then click **OK** button

(8) Create folders **lib** and **dist** under the project folder (**mag_lib_final-master**)



(9) Right click on **build_ssm.xml**, then **Run As->Ant Build.** If message shows "**BUILD SUCCESSFUL**", new **ssm.jar** will be created under the **dist** folder of the project folder



User may need to see **doc** folder where help files on **ssm** code are stored

10 Overall Freight Model Validation

MAG Technical Lead –Haidong Zhu

PAG Technical Lead - Aichong Sun, Hyunsoo Noh

The ultimate output from the Freight models are internal truck trips by class and time period, as well as daily external-internal medium and heavy trips for the entire freight modeling area and its external stations. The trips are stored in fourteen CSV files in a format of 4161 rows and 4162 columns, which include one ID column. In order to combine these text-based matrices with travel demand model matrices for assignment, a GISDK script was developed to convert the files from CSV to MTX format used in TransCAD. Since MAG and PAG maintain separate travel demand models, the incorporation and validation processes for the two agencies were carried out separately for their modeling areas.

10.1 Incorporation of Freight Models to MAG Travel Demand Model

The next step is to extract the truck trips pertaining to MAG modeling area from the converted matrix file. To do so, a matrix ID index crosswalk was created. The internal zone IDs of MAG network were automatically carried over. The external stations of MAG network and Mega-regional network are associated based on their matching locations. External zones # 12-17 of the Mega-regional network are assigned to MAG external zone # 2 (Figure 10-1 and Figure 10-2 on the following pages).

Figure 10-1 External stations for MAG network

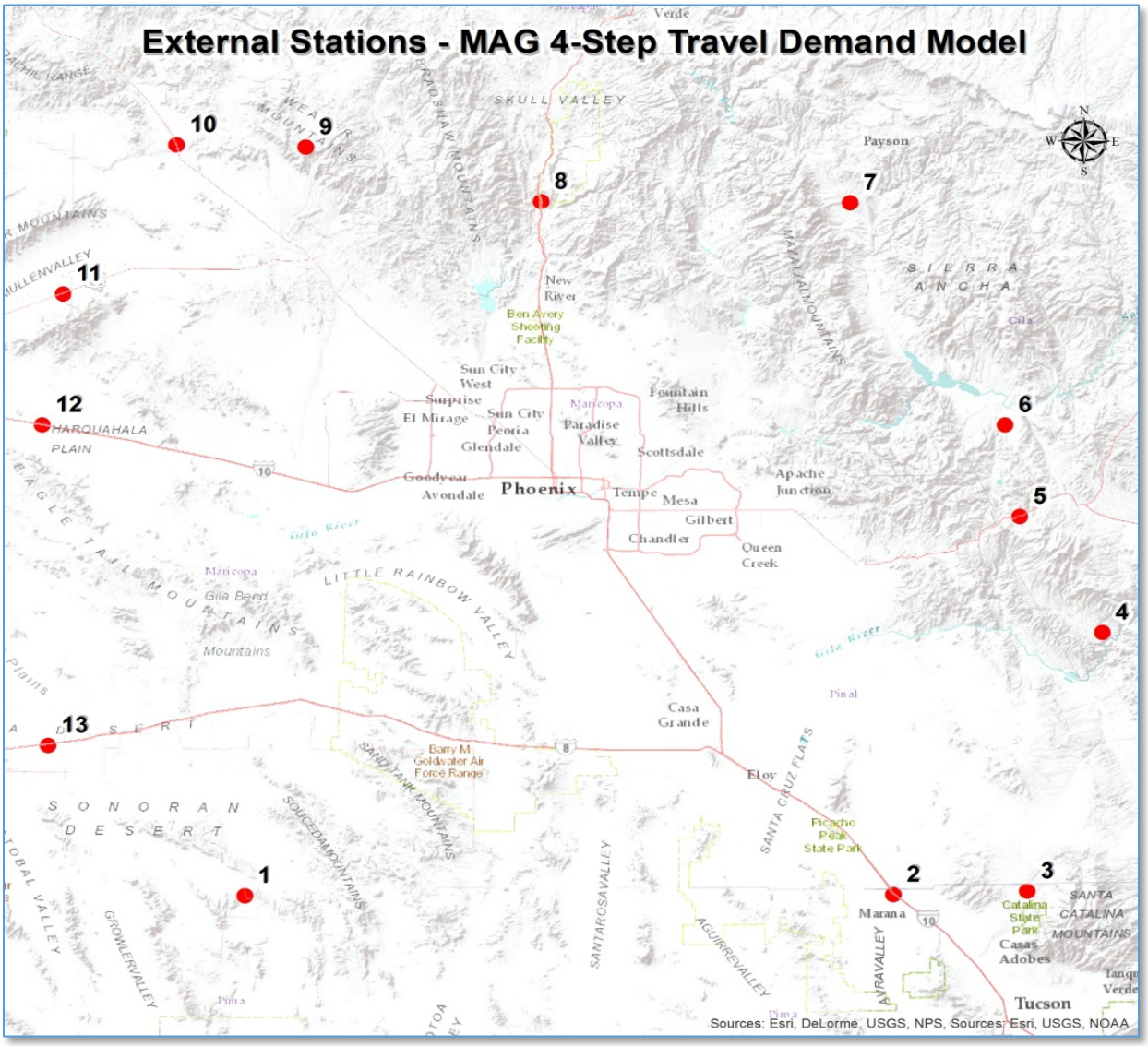
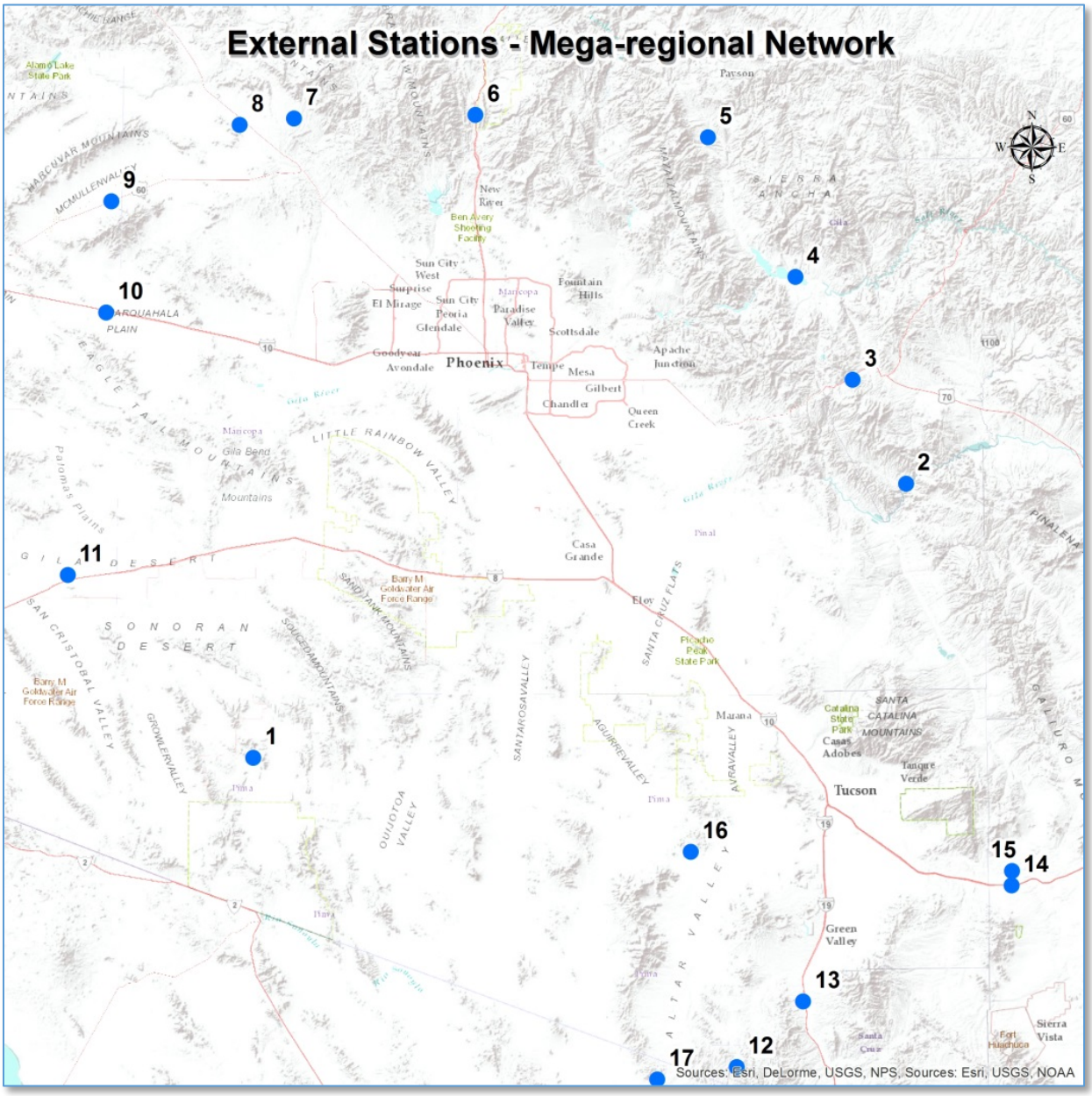
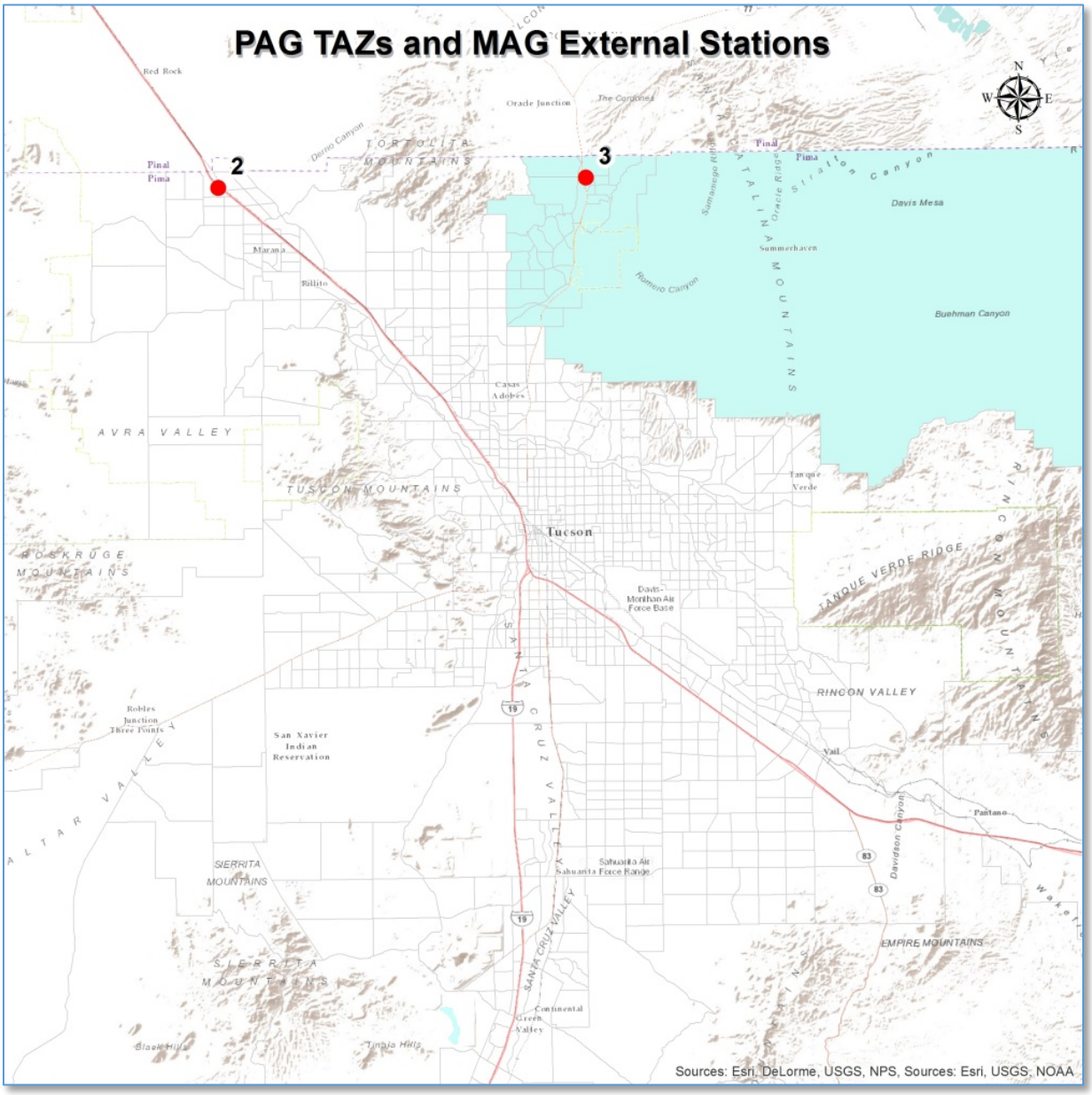


Figure 10-2 External stations for Mega-regional network



PAG zones are all external to MAG modeling area. So the shaded zones in Figure 10-3 were assumed to be associated with MAG external zone # 3, and the rest of PAG zones to external zone # 2.

Figure 10-3 External zone allocation for PAG area



The internal and external truck trip tables were combined but the truck classes – light, medium and heavy – were retained. These three classes of vehicle trip tables were then assigned with the passenger auto trip tables in a multiclass equilibrium assignment. Several assignment runs were performed until the total truck volumes were validated against observed data. The validation performance measures of the model are described in the following sections.

Trip assignment of the truck trips was completed using a user equilibrium highway assignment. Truck trips were assigned simultaneously with the passenger trips, because congestion has a significant impact on travel times experienced by trucks. Truck trips are assigned separately by type using the multiclass assignment technique for five vehicle types:

1. Single-occupant passenger vehicles,
2. High-occupant passenger vehicles with two or more occupants,
3. Light commercial trucks,
4. Medium trucks, and
5. Heavy trucks.

10.2 MAG Validation

The assignment validation is done at two different levels of geography. Screenlines analysis includes some of the major freeways that pass through the region and carry a large volume of trucks in the region. These freeways include I-17 W, I-17 E, I 10 N and I-10 S. Also included in these screenlines are Agua Fria, which is the river that flows through the west side of the MAG region while Salt River flows just south of the Phoenix metro area.

Table 10-1 presents the percent differences between the model and the observed screenline counts. These results are shown by screenline, truck type and direction. So of the differences are within the allowable range, while for some other locations further research is necessary.

Table 10-1 Screenline validation summary

Heavy Truck

Screenline	# of Counts	Count	Model	Difference	% Diff
I-17 W	28	27,578	26,066	-1,512	-5.5%
I-17 E	27	17,503	21,884	4,381	25.0%
I-10 N	44	38,778	41,512	2,734	7.0%
I-10 S	42	35,888	38,695	2,807	7.8%
Agua Fria	12	19,661	13,380	-6,281	-31.9%
Salt River	20	25,622	35,236	9,614	37.5%

Medium Truck

Screenline	# of Counts	Count	Model	Difference	% Diff
I-17 W	28	102,182	81,327	-20,855	-20.4%
I-17 E	27	80,219	72,741	-7,478	-9.3%
I-10 N	44	192,454	154,046	-38,408	-20.0%
I-10 S	42	82,608	113,557	30,949	37.5%
Agua Fria	12	49,214	25,859	-23,355	-47.5%
Salt River	20	80,231	106,685	26,454	33.0%

The sum of truck volumes at all locations were compared against the observed counts by truck class and area type. These are presented in Table 10-2. The medium and heavy trucks are within 25 percent for most combinations while for Suburban and Rural areas, medium truck needs additional refinement.

Table 10-2 Truck volumes vs. counts

Medium Truck

Area Type	Freeway	Arterial	All
CBD	-6.1%	-24.4%	-10.3%
Outlying CBD	6.7%	-14.1%	-0.7%
Mixed Urban	-3.6%	-34.7%	-17.2%
Suburban	-48.4%	-31.4%	-44.0%
Rural	111.5%	74.0%	101.8%
All	-1.4%	-21.6%	-8.2%

Heavy Truck

Area Type	Freeway	Arterial	All
CBD	12.1%	26.9%	13.9%
Outlying CBD	30.2%	20.2%	27.7%
Mixed Urban	-3.7%	-43.0%	-16.7%
Suburban	-9.4%	-14.2%	-10.9%
Rural	5.3%	-10.0%	3.9%
All	12.2%	-7.5%	7.7%

The light trucks are combined with passenger cars in this as there is no way to separate out light non-commercial vehicles from light commercial vehicles. The overall validation for all traffic is shown in Table 10-3.

Table 10-3 All volumes vs. counts

R2	Freeway	Arterial	All
CBD	0.971	0.863	0.987
Outlying CBD	0.964	0.822	0.977
Mixed Urban	0.959	0.845	0.965
Suburban	0.982	0.846	0.945
Rural	0.868	0.810	0.872
All	0.968	0.868	0.971

RMSE %	Freeway	Arterial	All
CBD	11.5%	22.6%	20.2%
Outlying CBD	12.0%	25.5%	22.2%
Mixed Urban	9.4%	23.1%	20.9%
Suburban	15.6%	32.0%	30.5%
Rural	45.8%	49.7%	53.8%
All	13.2%	27.6%	25.2%

% Diff	Freeway	Arterial	All
CBD	7.2%	0.7%	4.3%
Outlying CBD	5.7%	5.9%	5.8%
Mixed Urban	-4.8%	-7.3%	-6.7%
Suburban	-7.4%	-3.5%	-4.3%
Rural	21.6%	2.3%	6.6%
All	2.7%	-1.5%	-0.1%

10.3 Incorporation of Freight Models to PAG Travel Demand Model

The incorporation of the developed freight models into PAG's travel demand model for validation was a very similar task to what MAG has undertaken, which included the following steps:

- Develop a lookup table to associate the TAZs in mega-regional model to those in PAG's travel demand model.
- Use the developed lookup table to convert the mega-region truck trip matrices to the matrices that are compatible with the TAZ system in PAG's travel demand model

- Consolidate the heavy and medium truck trips in the converted matrices into the existing auto trip tables in the travel demand model chosen for this particular validation work. Specifically, the existing commercial truck trips in the travel demand model that are internal to PAG model area were replaced by the total of heavy and medium truck trips in the converted matrices, but the Light truck data in the converted matrices was ignored as PAG's travel demand model lacks the mechanism to separate light commercial truck trips from the other auto trips. Also, external zone related commercial truck trips in the travel demand model remained the same.
- Load two classes of auto trips, commercial trucks (heavy and medium combined) and other vehicles, onto PAG model network using Multi-class user equilibrium traffic assignment procedure.
- Compare the traffic assignment results with collected vehicle classification counts.

10.4 PAG Validation

The estimated commercial truck traffic volumes were compared with 91 vehicle classification counts recently collected in PAG area, with 16 counts on freeways and 75 on urban arterials

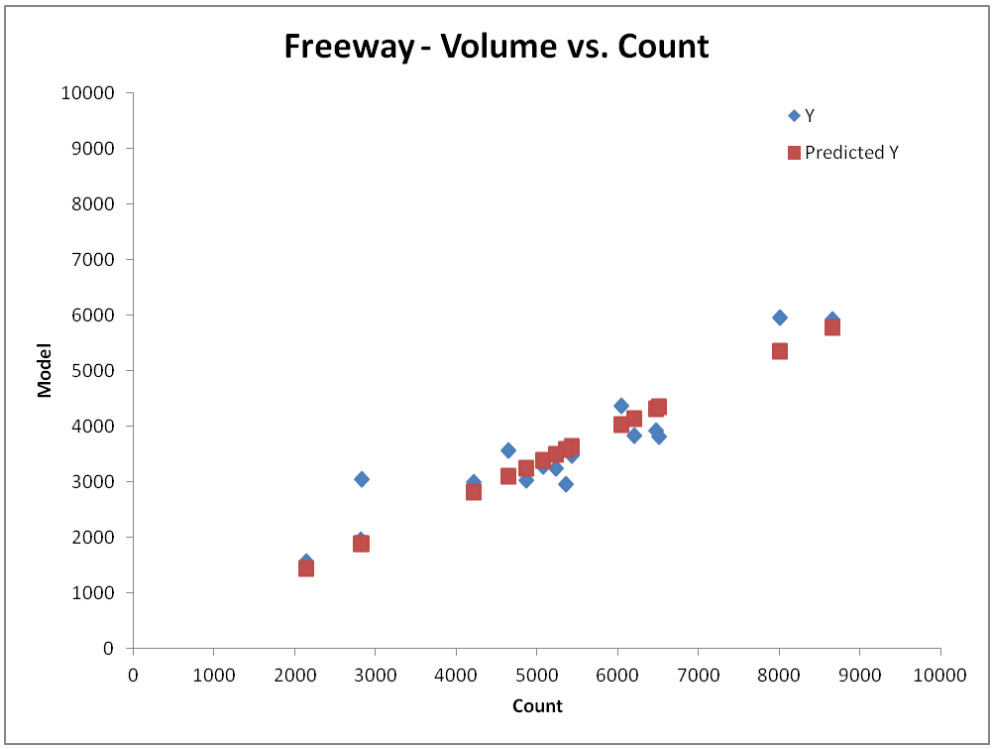
Table 10-4 presents how the modeled truck traffic volumes are compared with the observed counts for freeways and urban arterials. Overall, the model underestimates the truck traffic on freeways by 32.6% while it overestimates by 74.7% for urban arterials. This is also evidenced by the scatter plots in Figure 10-4 (a) and (b). However, such difference may be partially attributed to the existing external truck trips in PAG's travel demand model which were estimated largely based on 2009 external travel survey and assumed annual traffic growth rates. They may not be accurate enough for this validation work, but these trips may dominate other truck trips due to relatively small size of PAG area.

Table 10-4 Volume and Count Statistical Analysis of Freeway and Arterial

	Freeway	Arterial
%RMSE	36.0%	201.4%
Difference	-27,553	13,782
% Diff	-32.6%	74.7%

Figure 10-4 Volume-Count Statistical Result for Freeway and Arterial

(a) Freeway (R-Square: 0.986, X variable: 0.667)



(b) Arterial (R-Square: 0.589, X variable: 1.228)

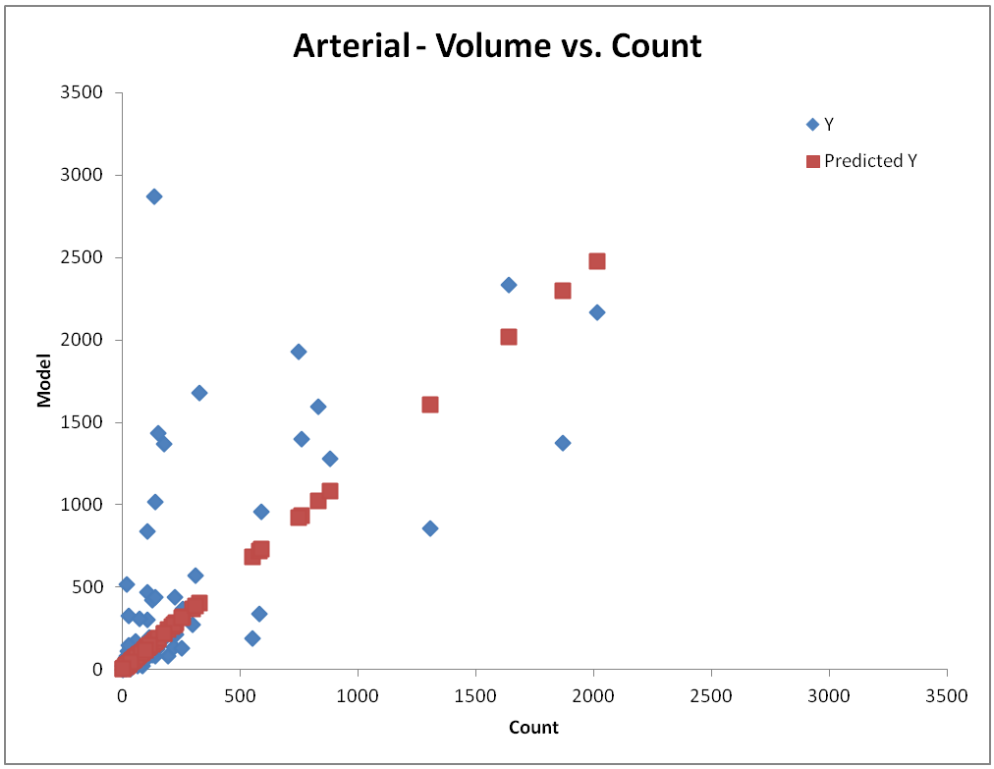


Table 10-5 presents more detailed comparison of estimated truck traffic volumes and observed counts. The comparison varies with the area type and facility type. However, good correlation of model estimates and observed counts has been observed for freeways and freeway and urban arterials combined no matter where they are while urban arterials alone tell a different version of story.

Table 10-5 Statistical Analysis by Area Type

Area Type	Freeway				
	Count	% Diff	%RMSE	R Square	X Variable
High Urban	0				
Medium Urban	5	-30.9%	34.4	1.00	0.69
Low Urban	7	-37.3%	38.5	0.99	0.62
Suburban	2	-16.2%	26.8	0.97	0.79
Rural	2	-33.7%	34.4	1.00	0.66

Area Type	Arterial				
	Count	% Diff	%RMSE	R Square	X Variable
High Urban	1	306.6%	306.6		
Medium Urban	22	125.6%	196.7	0.54	1.31
Low Urban	24	26.9%	84.7	0.86	1.11
Suburban	11	4.6%	149.6	0.38	0.84
Rural	17	-28.5%	104.0	0.68	0.46

Area Type	All				
	Count	% Diff	%RMSE	R Square	X Variable
High Urban	1	306.6%	306.6		
Medium Urban	27	9.9%	82.0	0.84	0.73
Low Urban	31	-28.6%	71.1	0.97	0.63
Suburban	13	-14.3%	63.9	0.97	0.79
Rural	19	-32.9%	91.7	0.99	0.66

11 Visualization

MAG Technical Lead – Shuyao Hong

11.1 Freight Model Visualization

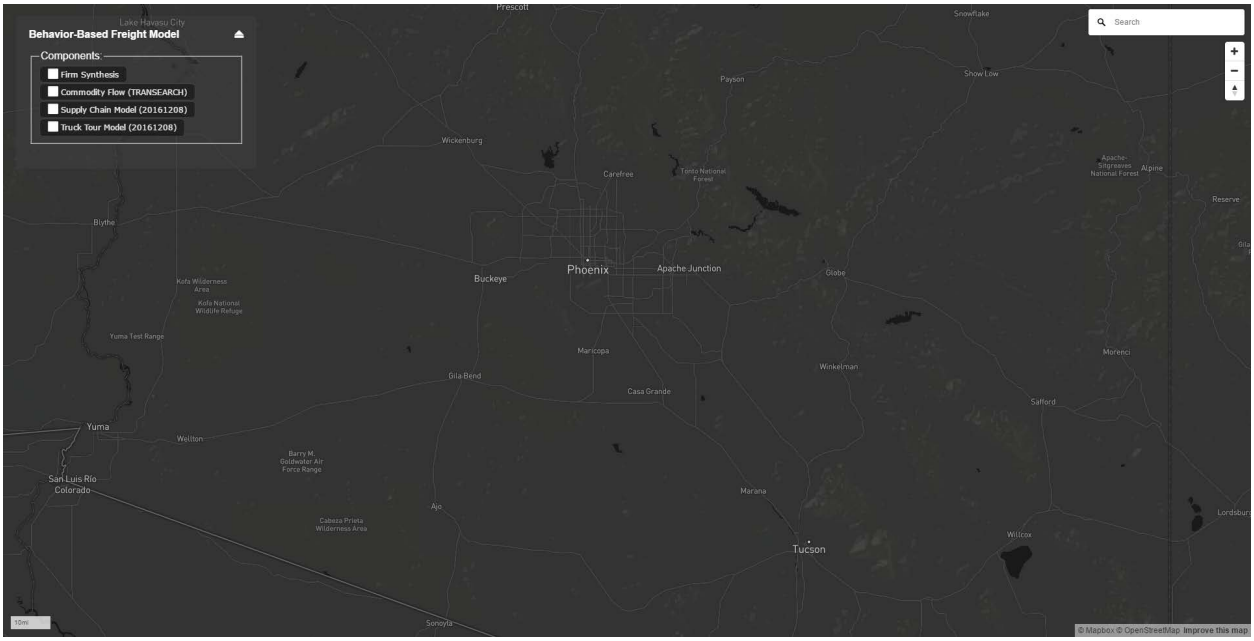
The main visualization tool of this effort combines several model components and utilizes Mapbox GL JS[68], an open-source JavaScript web mapping library, to display various datasets at different levels of geography from either a 2-dimensional or 3-dimensional perspective. Mapbox GL JS uses WebGL[69] technology to render interactive maps on the user's web browser and is widely supported in any modern web browser including those on mobile devices. Its rich built-in API[70] functions give developers a greater control over displaying various geospatial features (i.e. point, polyline or polygon) with specified color, width/radius, and/or extrusion height, based on associated data. It also dynamically changes mapped features based on the viewer's actions, such as clicking, mouse hovering, or zooming in/out. This key advantage makes the library particularly suitable for visualizing the freight model datasets developed in this project because various types of data involved are at different levels of geography and from different modeling components.

The visualization interface is accessible via a web browser. Figure 11-1 shows the interface. The viewer has a dark-theme base map, provided by Mapbox, and is by default centered on the mega-region study area. Users are able to find common web map control widgets like address search bar, zoom in/out buttons on the upper-right corner, and an adaptive scale bar on the lower-left corner, which are all provided natively by Mapbox GL JS library. Mouse control functions are similar to other web maps where the user could pan the map by dragging the left mouse button or zoom in/out by clicking the left mouse button twice or scrolling mouse wheel. In addition, holding the right mouse button while dragging the map will change the viewing perspective freely between 2-dimensional and 3-dimensional.

The main control panel is on the upper left corner, which allows users to select different model component(s) to be visualized. Based on the user's selection, additional settings and parameters will appear to enable further analyses on the selected datasets. Each component will be discussed separately in following subsections and followed by a description about the workflow and tools behind the scene.

[68] https://www.mapbox.com/mapbox-gl-js/api/
[69] Web Graphics Library is a JavaScript API for rendering interactive 3D and 2D graphics within any compatible web browser without the use of plug-ins. (https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API)
[70] Application Program Interface: a set of clearly defined methods of communication between various software components. (https://en.wikipedia.org/wiki/Application_programming_interface)

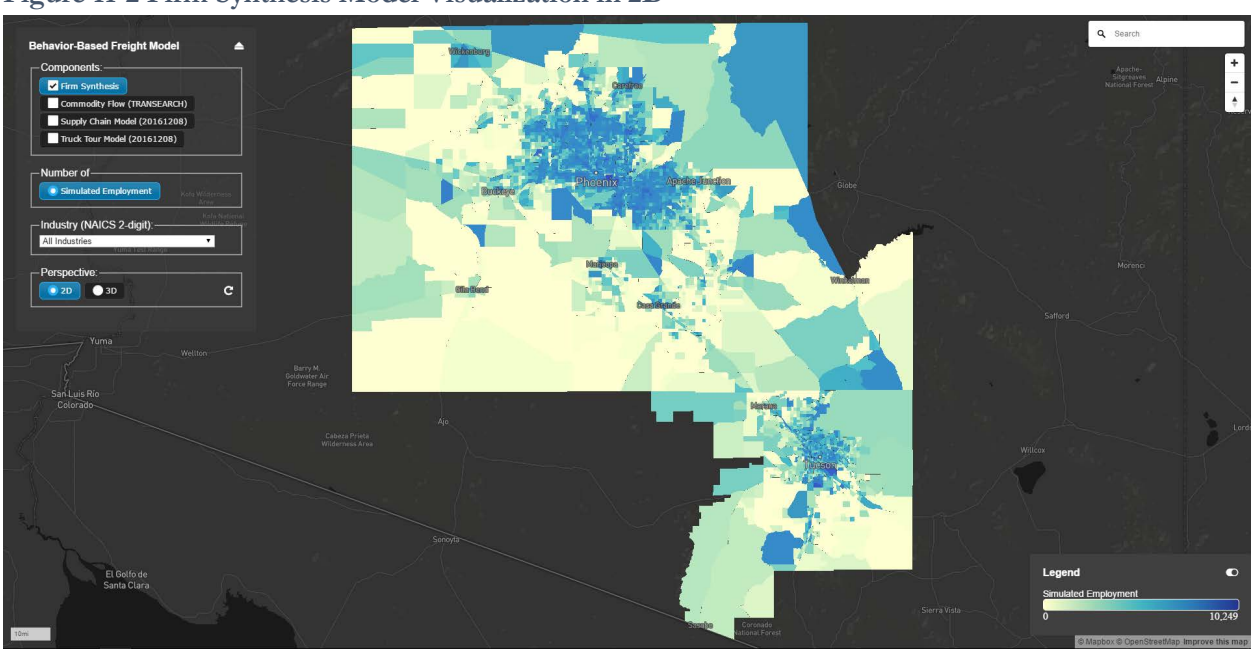
Figure 11-1 Freight Model Visualization Interface at Launch



11.1.1 Firm Synthesis Model

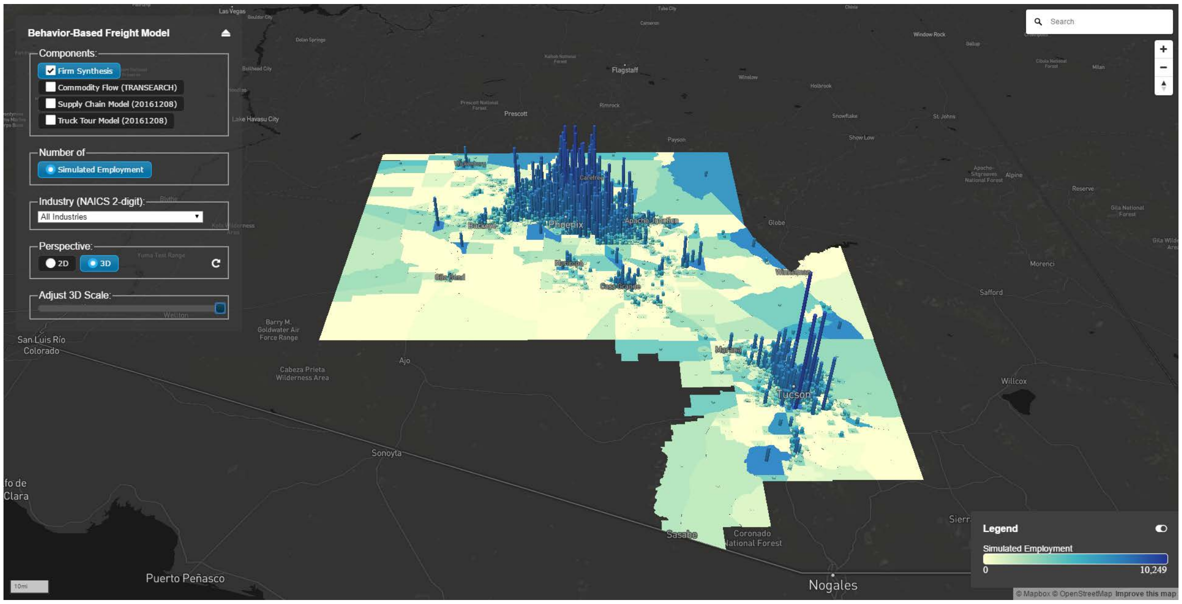
Once "Firm Synthesis" is selected in the main control panel, the number of simulated employment for all industry at the TAZ level will be visualized in the map in 2D as shown in Figure 11-2. By default, it is a 2D choropleth map where each TAZ is colored based on the amount of simulated employment as indicated in the legend. Users can specify which specific industry's simulated employment from the Firm Synthesis Model they would like to see by choosing a specific NAICS 2-digit industry category from the drop-down menu on the control panel.

Figure 11-2 Firm Synthesis Model Visualization in 2D



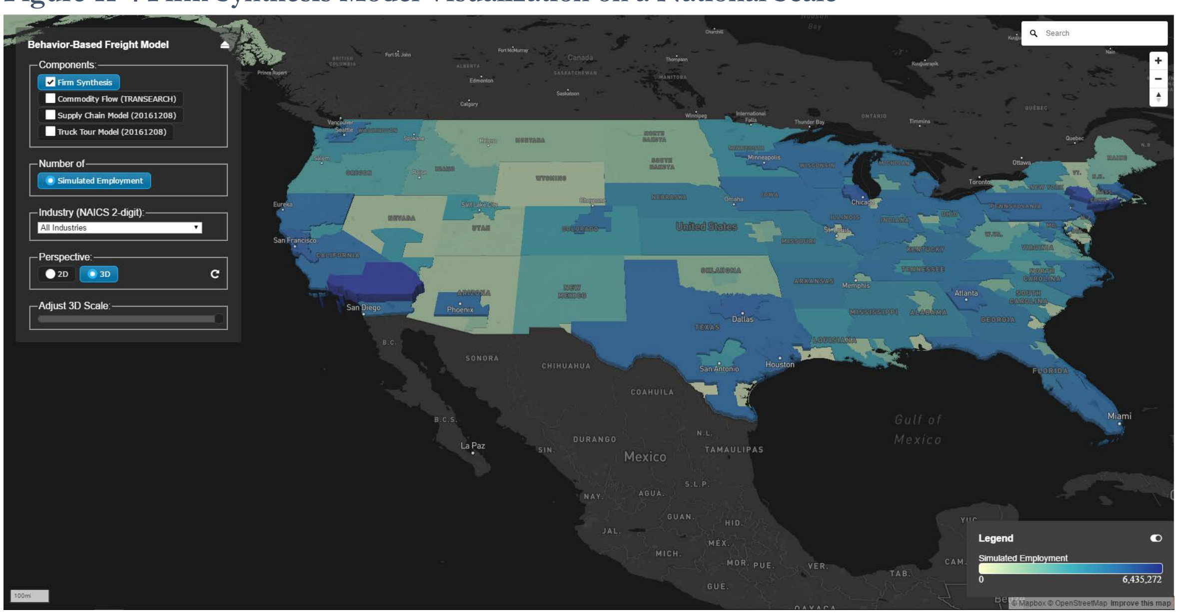
As shown in Figure 11-3, users can also visualize this data in 3-dimension by choosing the "3D" perspective in the control panel, which then presents the data in a more easy-to-digest way compared to the 2D choropleth map shown earlier. The extrusion height of each cylinder is proportional to the number of simulated employees at each TAZ. Users are able to control the 3D extrusion scale via the slider in the control panel, which could be handy when they would like to focus on a small area.

Figure 11-3 Firm Synthesis Model Visualization in 3D



When users zoom out the map to the national scale, the viewer will then display the simulated employment of firms related to Arizona-related freight activities by FAF zone at a national scale as shown in Figure 11-4. Data within the study area are from Firm Synthesis Model and for external firms, the sources is County Business Pattern data from U.S. Census Bureau.

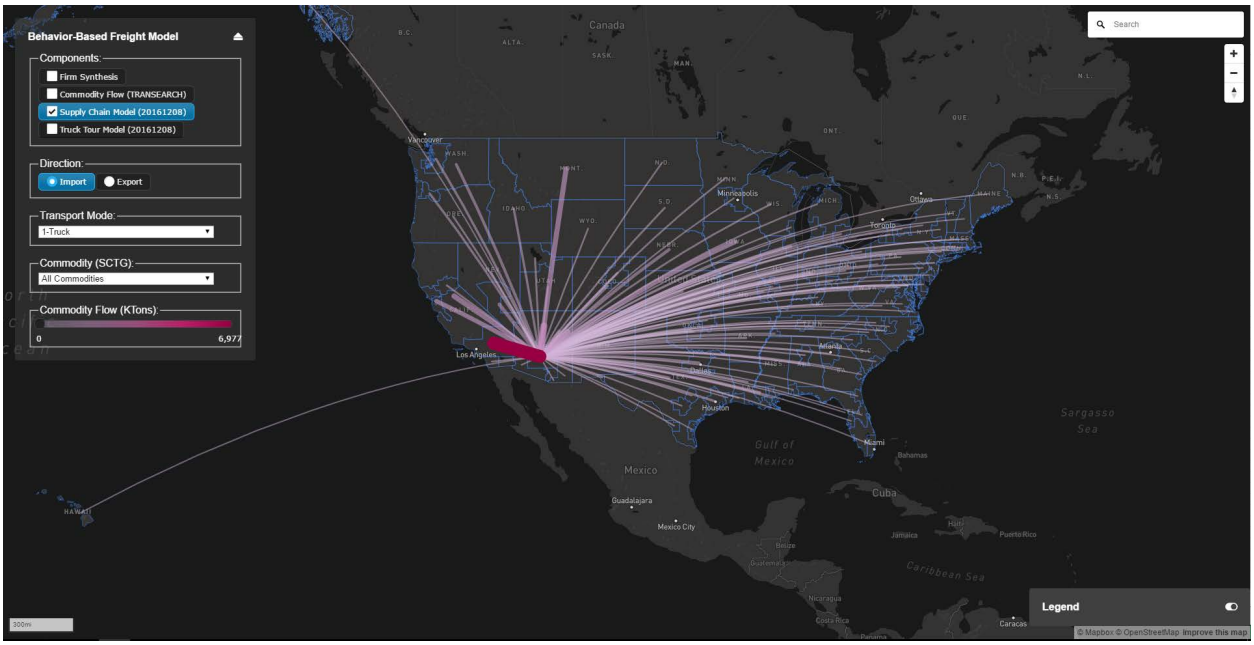
Figure 11-4 Firm Synthesis Model Visualization on a National Scale



11.1.2 Commodity Flow from Supply Chain Model

The next visualization component uses commodity flow outputs from the Supply Chain Model. The commodity flow records were aggregated to FAF region level and the FAF zone boundaries are drawn as blue in the visualization (Figure 11-5). Great circle routes from the mega-region to all other regions are drawn on the map. By default, the width and color of the lines are proportional to the total amount of commodity flow in tonnage. Users can specify the type of commodity group (in SCTG 2-digit code), commodity flow direction (import or export to the mega-region), transport mode, and unit of measurements (weights in tons or dollar value) in the control panel, and the visualization and the legend will update accordingly. When users hover the mouse pointer over a specific route on the map, the information box on the upper-right corner will display the origin and destination of that particular flow with the corresponding amount, as shown in. Additionally, the control panel allows users to filter out some of the small-volume flows by using the slider function, in order to focus on the major commodity flows coming in/out the mega-region.

Figure 11-5 Commodity Flow Visualization (Supply Chain Model)



11.1.3 Commodity Flow from TRANSEARCH

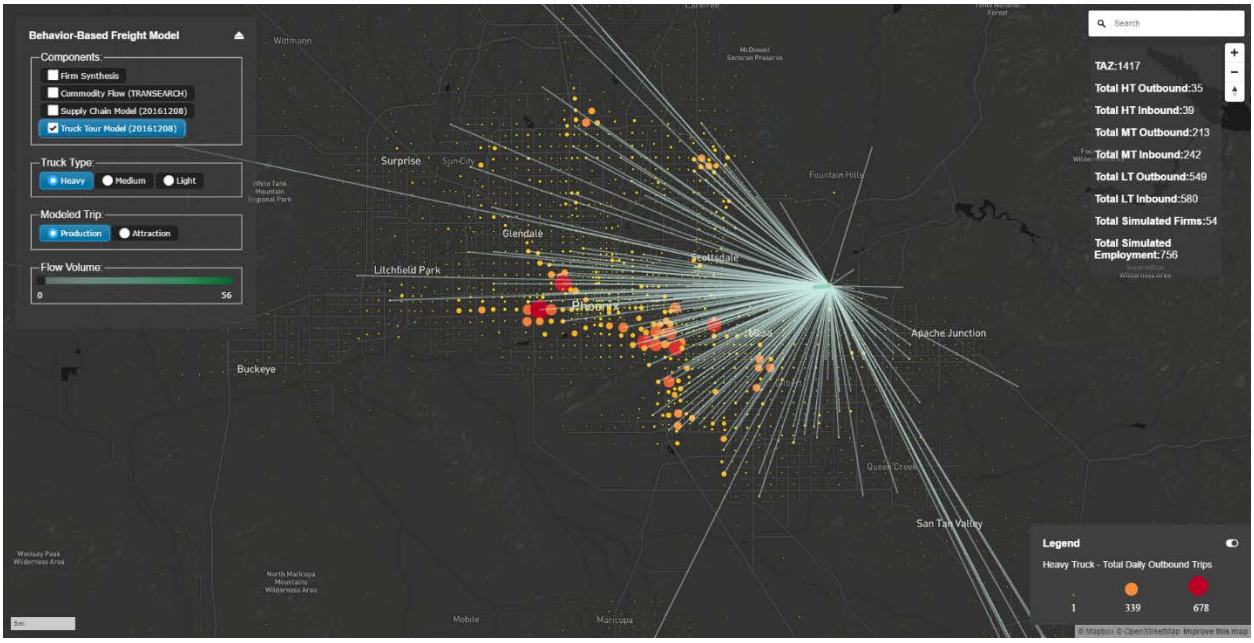
The next visualization aggregates commodity flow data from the TRANSEARCH database, which is the validation data source for the Supply Chain Model. Similarly to the previous component, users can also specify which mode they would like to visualize, commodity group (in STCC 2-digit code), flow direction (import or export to the mega-region), and unit of measurements (weights in tons or dollar value) on the control panel (Figure 11-6). The geography is Bureau of Economic Analysis Economic Area (BEA) region level. Because of the difference between commodity grouping of Supply Chain Model (SCTG) and TRANSEARCH database (STCC), users need to select matching commodity groups when comparing these two sources.

Figure 11-6 Commodity Flow Visualization (TRANSEARCH)

11.1.4 Truck Tour Model

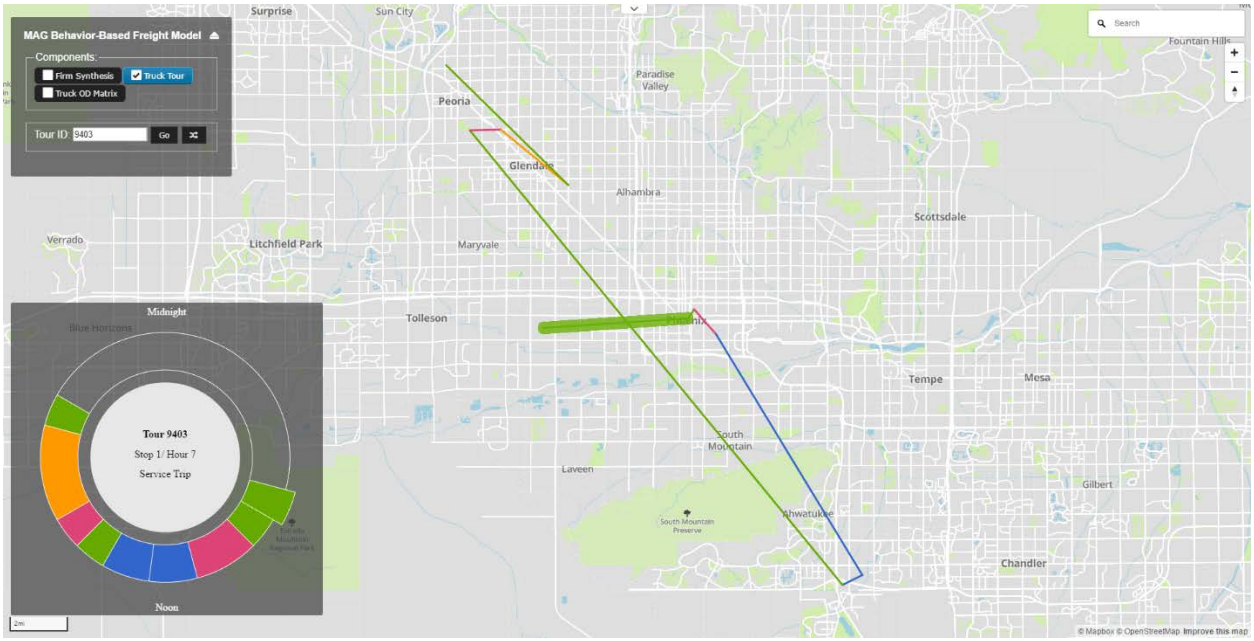
Several visualizations were developed based on outputs from the Truck Tour Model. The trip matrix from the Truck Tour Model is used to show the truck travel pattern at TAZ level. As shown in Figure 11-7, trip matrices produced by the model were used in two different ways. Each circle's radius and color are proportional to the number trips going out from/in to a TAZ for the selected truck types, which allows users to examine the spatial pattern of trip production/attraction for the selected truck type at TAZ level. When users hover the mouse pointer over a specific circle/TAZ on the map, an interactive representation of desire lines will appear where the width and color transparency of each line are proportional to the number of trips going outbound/inbound to the TAZ on the other end of the line. Desire line will not be drawn between TAZs with no trips between them. An information box will also appear on the upper right corner showing the numerical values about this TAZ. Additionally, the control panel allows users to filter out some of the small-volume desire line by using a slider.

Figure 11-7 Truck Tour Model Visualization - Trip Matrix



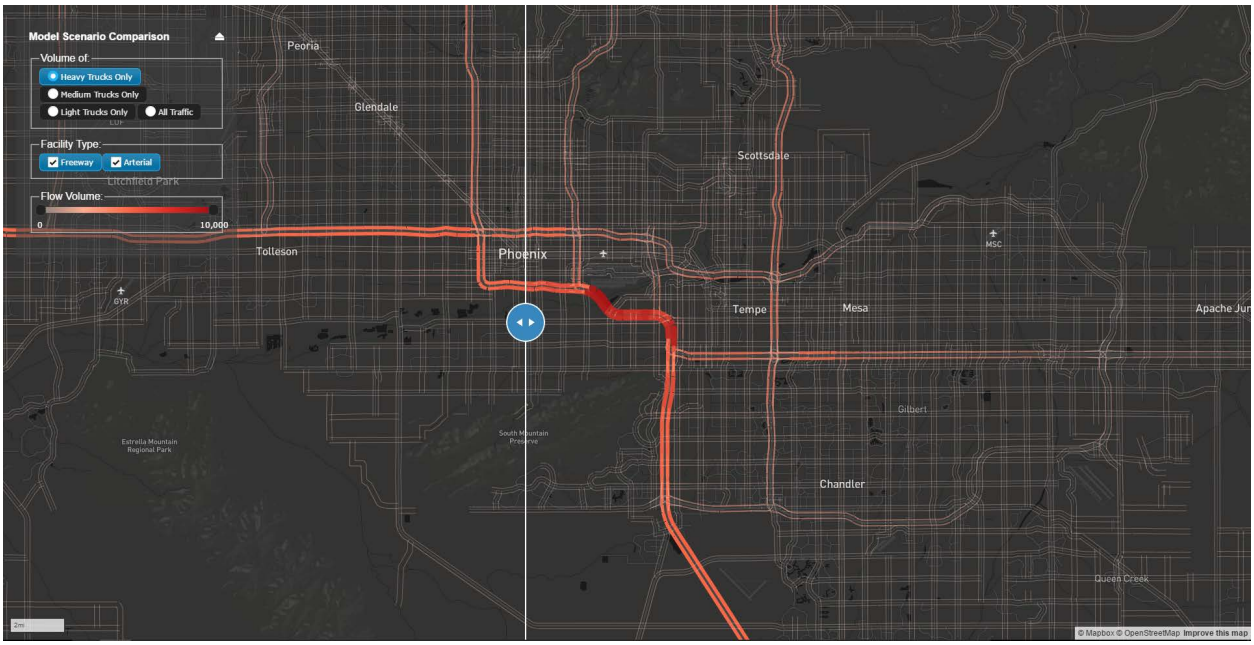
Another important output from Truck Tour Model is detailed records of each modelled tour, containing information about tour purpose, tour completion status, and start time and stop purpose of each trips within the tour. As shown in Figure 11-8, an interactive visualization is developed to display information at individual tour level. Once a tour id is submitted on the control panel, the time of day profile of the selected tour will be displayed in a 24-hour clock on the lower left corner, while stops of the tour (at TAZ level) will be connected sequentially via straight lines on the map. By default, the center of the clock will show general information about the tour such as how many stops are included, and whether the truck comes back to the starting TAZ. Each segment on the clock and the corresponding trip leg on the map are colored based on the stop purpose. The duration of each segment is determined by the start time of this trip and the next. When users hover the mouse pointer over a specific segment on the clock, the corresponding leg of trip will be highlighted on the map. If a user clicks a segment on the clock, a 3D fly-over animation will happen on the map for a closer look at both trip ends. This visualization can overlay other information for QC purposes.

Figure 11-8 Truck Tour Model Visualization - Individual Tour



After assignment of the truck trips on the network was completed, one way to compare truck volumes between different scenarios or different truck types is shown below in Figure 11-9. Two views of the map are displayed and separated by a slider on the center. Users could sweep the slider left and right to compare maps on the left and right sides. In addition, user's activities such as zooming and panning will be synced between two maps, which makes interactive comparison possible.

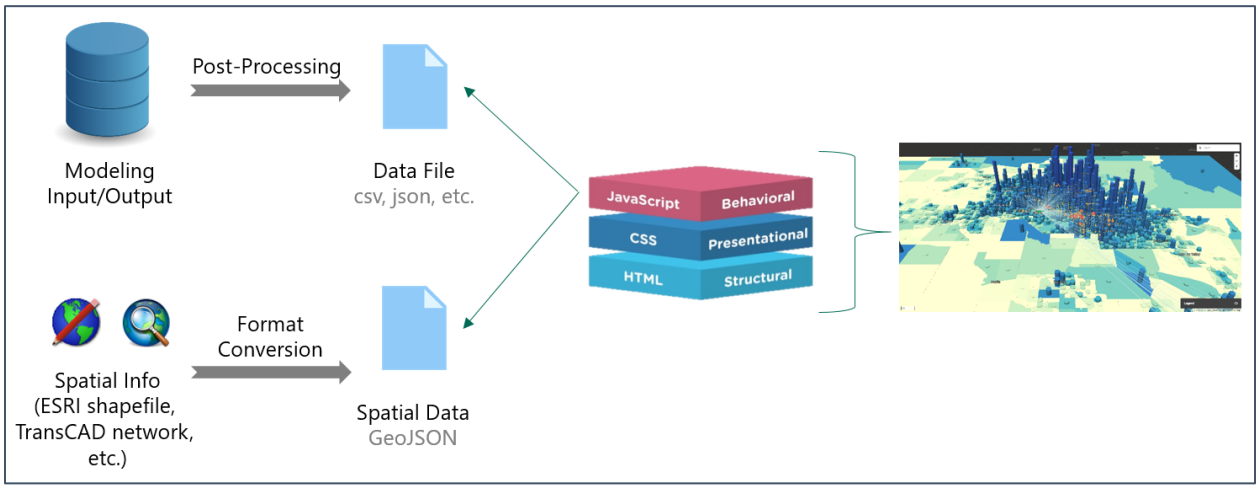
Figure 11-9 Truck Tour Model Visualization - Truck Volume Comparison



11.1.5 Workflow and Tools

The workflow behind the freight model visualization is simple but scalable (Figure 11-10). Firstly, the raw data were converted/aggregated using Python to web-ready format like csv or JSON[71] files and converted ESRI shapefiles or TransCAD network to GeoJSON[72] format. This procedure can be replaced if the data is already served through an API on a server or the geography information is served through a Web Map Services.

Figure 11-10 Freight Model Visualization Workflow



Once the data preparation is finished, web development is needed in three major aspects. A web page in HTML5 is needed to define the structure of the visualization, such as how the page layout looks like and how each control element is placed in the visualization. Corresponding CSS[73] files are needed for the appearance of all the elements. Last but most importantly, JavaScript scripts are used to (1) control the interaction between different visual components, (2) load the datasets into the browser tab's memory space, and (3) call Mapbox GL JS APIs to display and modify the web map visualization per users' actions. Some useful links for Mapbox GL JS library are listed below:

- Github: https://github.com/mapbox/mapbox-gl-js
- API document: https://www.mapbox.com/mapbox-gl-js/api/
- Demo: https://www.mapbox.com/blog/population-inspector/
- Demo: https://www.mapbox.com/blog/track-your-jetlag-with-data-driven-line-color/

[71] JavaScript Object Notation
[72] A geospatial data interchange format based on JavaScript Object Notation (https://tools.ietf.org/html/rfc7946)
[73] Cascading Style Sheets

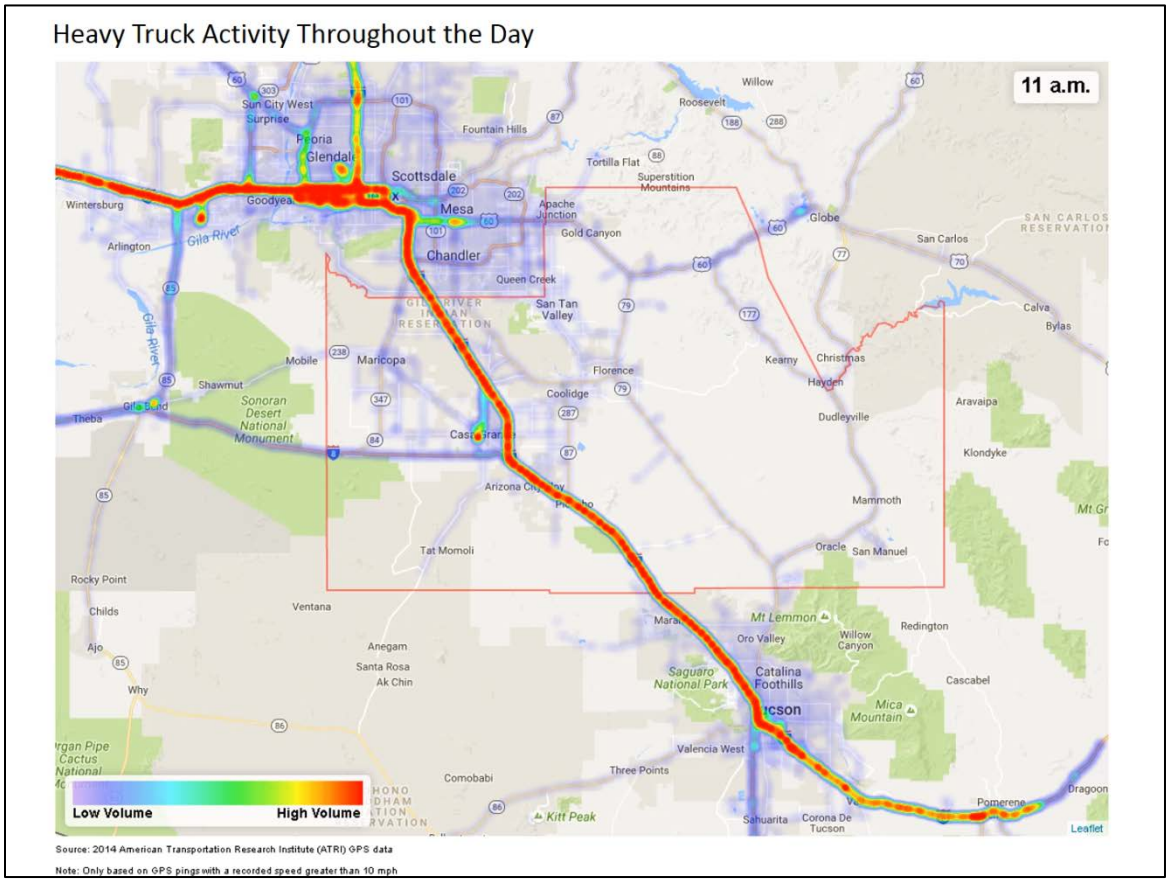
11.2 GPS Data Visualizations

Truck GPS data obtained in this project contain a large amount of spatial-temporal information. In order to get the most out of it, additional data mining and visual analytics applications are needed to transform data into meaningful insights about truck activities. This section will discuss MAG's internal data visualization efforts using truck GPS data.

11.2.1 Heat Map

As shown in Figure 4-3 and Figure 4-9 in Chapter 4, heat map method was used in this project to visualize the spatial concentration of GPS pings as it shows truck activity concentration in a clearer way than a point cluster map. The visualization was developed in an interactive web map using Leaflet web map library's *Heat* plugin[74] . As marked in the legend, the red color indicates a high concentration of points. It is a scale-dependent representation which means when users zoom in/out, the map will re-render accordingly. Depending on what message one would like to convey, a temporal dimension could be added to the heat map visualization and make it an animation (e.g. a screenshot of an animation is shown in Figure 11-11), or one can visualize pings of selected trucks that go to/leave from a particular area of interest.

Figure 11-11 Truck Heat Map Animation Screenshot



[74] https://github.com/Leaflet/Leaflet.heat

The maximum memory that a browser tab can handle limits how many points you can simultaneously load into the heat map. For example, if displaying more than 3 million pairs of coordinates in a 64-bit Chrome browser, selecting a random subset of all pings is recommended. Another caveat of using heat map is that there are several parameters that will affect the appearance of a heat map, such as the minimum opacity the heat will start at, or radius of each "point" of the heat map. In order to compare two heat maps visually in a meaningful way, it is recommended to keep these parameters identical or as close as possible. Heat maps are most suited for getting the big picture at a glance, as a supplement to quantitative analyses, rather than a replacement.

11.2.2 Vehicle Trajectory Animation

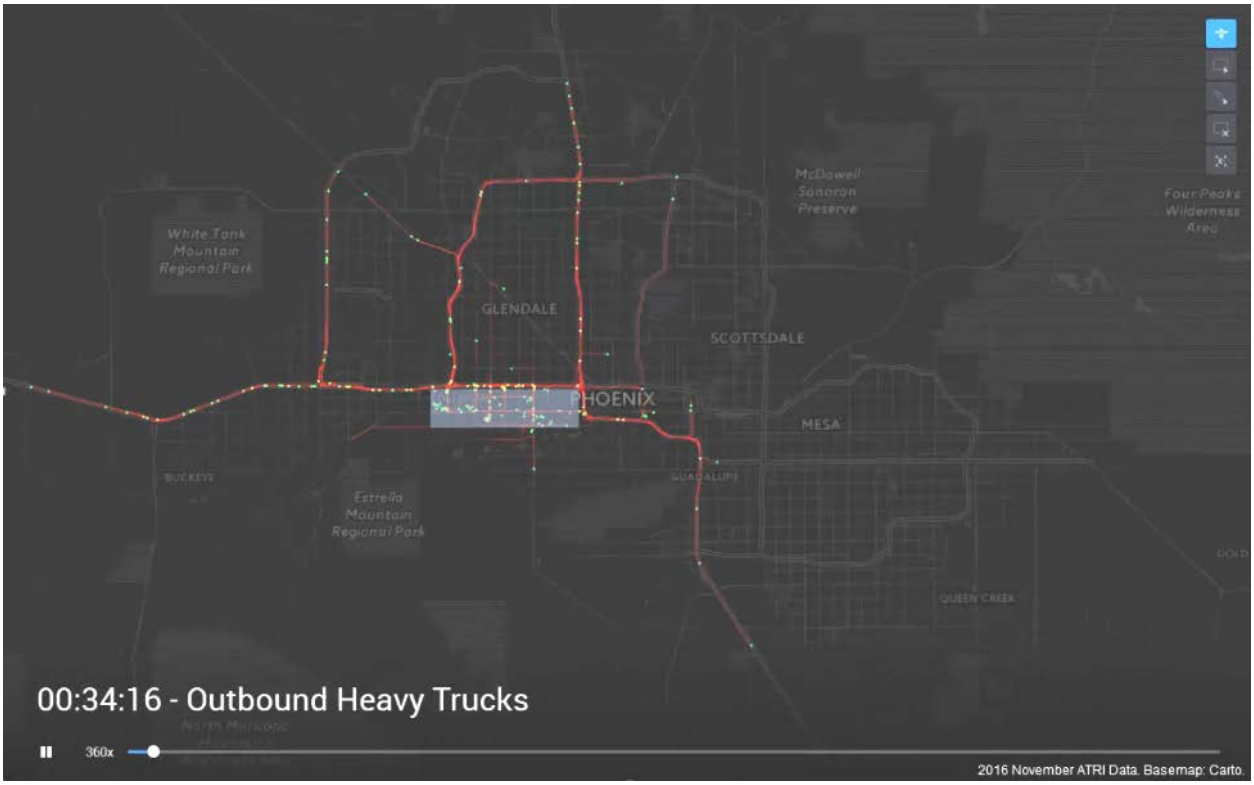
In order to show the dynamic aspect of the GPS data, a series of experimental animation has been developed to visualize vehicle trajectories. An innovative map-matching algorithm[75] was first used to reconstruct the vehicle's trajectory from low-frequency GPS pings. Leaflet's *MovingMarker* plug-in[76] was used to develop the trajectory animation as shown in Figure 11-12. The visualization has a playback progress bar with adjustable playback speed and a clock on the upper right corner to represent the time dimension. Selected inbound heavy trucks are represented as points on the map which are colored based on where were they came from. Once start playing, selected vehicles will start moving based on their location at the given timestamp from GPS data. In this particular case, two weeks of data was aggregated into a typical 24-hour animation. This kind of visualization/animation heavily depends on how the GPS data was processed and selected for displaying and is best suited for showing the big picture.

Figure 11-12 Vehicle Trajectory Animation using Leaflet

[75] Camargo, P., Hong, S., Livshits, V. (2017). Expanding the Uses of Truck GPS Data in Freight Modeling and Planning Activities. Transportation Research Record, No. 2646. *In press*. DOI:10.3141/2646-08.
[76] https://github.com/ewoken/Leaflet.MovingMarker

A similar truck trajectory animation was created using CityPhi software. Any kind of trajectory information (as long as there are id, timestamp, and coordinates) including GPS data can be imported to the software through Jupyter Notebook using Python. As shown in Figure 11-13, the animated point layer is used to show the current location of one vehicle id at a given timestamp, while the motion trailer layer is used to retain a transparent trace for a specified period of time after the point is moving away. The time dimension is added to the visualization interface in the form of a playback slider with adjustable playback speed. Select link analysis is also possible. Users could draw a select link in CityPhi's visual interface and vehicles that have passed or will pass that link could be highlighted in a different color along with their motion trailers.

Figure 11-13 Vehicle Trajectory Animation using CityPhi



11.2.3 Summary

License agreements prohibit the sharing of raw GPS data as they contain personally identifiable information (PII). A number of measures has been taken to avoid any violation. First of all, visualizations are based on highly aggregated groups of GPS pings rather than from an individual vehicle. For example, several days of data were aggregated into a 24-hour period so it would not be able to trace back to a specific date. The zoom-level of the visual interface was also limited so that the precise location is not available to the users. Most importantly, even though some animations were developed in a web interface, video clips were used to present the visualizations, rather than making the visualization available through the web. This is to ensure that no data was leaked externally and to allow careful examinations of clips to make sure no PII is compromised. As mentioned earlier, GPS data animations are best suited for getting the big picture at a glance, as a supplement to quantitative analyses.

Appendix A. CFS Microdata Megaregions

MA	STATE	*MA Type*	Description	names
142	01	C	Birmingham-Hoover-Talladega, AL CFS Area	Southeast U.S.
380	01	C	Mobile-Daphne-Fairhope, AL CFS Area	Southeast U.S.
99999	01	R	Remainder of Alabama	Southeast U.S.
99999	02	S	Remainder of Alaska	Alaska and Hawaii
536	04	C	Tucson-Nogales, AZ CFS Area	Sun Corridor - Phoenix and Tucson
38060	04	M	Phoenix-Mesa-Glendale, AZ CFS Area	Sun Corridor - Phoenix and Tucson
99999	04	R	Remainder of Arizona	Sun Corridor - Phoenix and Tucson
99999	05	S	Remainder of Arkansas	Southeast U.S.
260	06	C	Fresno-Madera, CA CFS Area	California - Other
348	06	C	Los Angeles-Long Beach, CA CFS Area	California - Southern
472	06	C	Sacramento-Roseville, CA CFS Area	California - Other
488	06	C	San Jose-San Francisco-Oakland, CA CFS Area	California - Other
41740	06	M	San Diego-Carlsbad-San Marcos, CA CFS Area	California - Southern
99999	06	R	Remainder of California	California - Other
216	08	C	Denver-Aurora, CO CFS Area	Colorado
99999	08	R	Remainder of Colorado	Colorado
408	09	C	New York-Newark, NY-NJ-CT-PA CFS Area (CT Part)	Northeast U.S.
25540	09	M	Hartford-West Hartford-East Hartford, CT CFS Area	Northeast U.S.
99999	09	R	Remainder of Connecticut	Northeast U.S.
428	10	C	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area (DE Part)	Northeast U.S.
99999	10	R	Remainder of Delaware	Northeast U.S.
47900	11	M	Washington-Arlington-Alexandria, DC-VA-MD-WV CFS Area (DC Part)	Northeast U.S.
300	12	C	Jacksonville-St. Marys-Palatka, FL-GA CFS Area (FL Part)	Southeast U.S.
370	12	C	Miami-Fort Lauderdale-Port St. Lucie, FL CFS Area	Southeast U.S.
422	12	C	Orlando-Deltona-Daytona Beach, FL CFS Area	Southeast U.S.
45300	12	M	Tampa-St. Petersburg-Clearwater, FL CFS Area	Southeast U.S.
99999	12	R	Remainder of Florida	Southeast U.S.
122	13	C	Atlanta-Athens-Clarke County-Sandy Springs, GA CFS Area	Southeast U.S.

MA	STATE	MA Type	Description	names
496	13	C	Savannah-Hinesville-Statesboro, GA CFS Area	Southeast U.S.
99999	13	R	Remainder of Georgia	Southeast U.S.
46520	15	M	Urban Honolulu, HI CFS Area	Alaska and Hawaii
99999	15	R	Remainder of Hawaii	Alaska and Hawaii
99999	16	S	Remainder of Idaho	Central-Upper Midwest
176	17	C	Chicago-Naperville, IL-IN-WI CFS Area (IL Part)	Great Lakes
476	17	C	St. Louis-St. Charles-Farmington, MO-IL CFS Area (IL Part)	Great Lakes
99999	17	R	Remainder of Illinois	Great Lakes
176	18	C	Chicago-Naperville, IL-IN-WI CFS Area (IN Part)	Great Lakes
258	18	C	Fort Wayne-Huntington-Auburn, IN CFS Area	Great Lakes
294	18	C	Indianapolis-Carmel-Muncie, IN CFS Area	Great Lakes
99999	18	R	Remainder of Indiana	Great Lakes
99999	19	S	Remainder of Iowa	Central-Upper Midwest
312	20	C	Kansas City-Overland Park-Kansas City, MO-KS CFS Area (KS Part)	Central-Upper Midwest
556	20	C	Wichita-Arkansas City-Winfield, KS CFS Area	Central-Upper Midwest
99999	20	R	Remainder of Kansas	Central-Upper Midwest
178	21	C	Cincinnati-Wilmington-Maysville, OH-KY-IN CFS Area (KY Part)	Southeast U.S.
350	21	C	Louisville/Jefferson County-Elizabethtown-Madison, KY-IN CFS Area (KY Part)	Southeast U.S.
99999	21	R	Remainder of Kentucky	Southeast U.S.
406	22	C	New Orleans-Metairie-Hammond, LA-MS CFS Area (LA Part)	Southeast U.S.
12940	22	M	Baton Rouge, LA CFS Area	Southeast U.S.
29340	22	M	Lake Charles, LA CFS Area	Southeast U.S.
99999	22	R	Remainder of Louisiana	Southeast U.S.
99999	23	S	Remainder of Maine	Northeast U.S.
12580	24	M	Baltimore-Columbia-Towson, MD CFS Area	Northeast U.S.
47900	24	M	Washington-Arlington-Alexandria, DC-VA-MD-WV CFS Area (MD Part)	Northeast U.S.
99999	24	R	Remainder of Maryland	Northeast U.S.
148	25	C	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (MA Part)	Northeast U.S.
99999	25	R	Remainder of Massachusetts	Northeast U.S.
220	26	C	Detroit-Warren-Ann Arbor, MI CFS Area	Great Lakes
266	26	C	Grand Rapids-Wyoming-Muskegon, MI CFS Area	Great Lakes

MA	STATE	MA Type	Description	names
99999	26	R	Remainder of Michigan	Great Lakes
378	27	C	Minneapolis-St. Paul, MN-WI CFS Area (MN Part)	Great Lakes
99999	27	R	Remainder of Minnesota	Great Lakes
99999	28	S	Remainder of Mississippi	Southeast U.S.
312	29	C	Kansas City-Overland Park-Kansas City, MO-KS CFS Area (MO Part)	Central-Upper Midwest
476	29	C	St. Louis-St. Charles-Farmington, MO-IL CFS Area (MO Part)	Central-Upper Midwest
99999	29	R	Remainder of Missouri	Central-Upper Midwest
99999	30	S	Remainder of Montana	Central-Upper Midwest
420	31	C	Omaha-Council Bluffs-Fremont, NE-IA CFS Area (NE Part)	Central-Upper Midwest
99999	31	R	Remainder of Nebraska	Central-Upper Midwest
332	32	C	Las Vegas-Henderson, NV-AZ CFS Area (NV Part)	Nevada
99999	32	R	Remainder of Nevada	Nevada
148	33	C	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (NH Part)	Northeast U.S.
99999	33	R	Remainder of New Hampshire	Northeast U.S.
408	34	C	New York-Newark, NY-NJ-CT-PA CFS Area (NJ Part)	Northeast U.S.
428	34	C	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area (NJ Part)	Northeast U.S.
99999	35	S	Remainder of New Mexico	New Mexico
160	36	C	Buffalo-Cheektowaga, NY CFS Area	Northeast U.S.
408	36	C	New York-Newark, NY-NJ-CT-PA CFS Area (NY Part)	Northeast U.S.
464	36	C	Rochester-Batavia-Seneca Falls, NY CFS Area	Northeast U.S.
99999	36	R	Remainder of New York	Northeast U.S.
172	37	C	Charlotte-Concord, NC-SC CFS Area (NC Part)	Southeast U.S.
268	37	C	Greensboro--Winston-Salem--High Point, NC CFS Area	Southeast U.S.
450	37	C	Raleigh-Durham-Chapel Hill, NC CFS Area	Southeast U.S.
99999	37	R	Remainder of North Carolina	Southeast U.S.
99999	38	S	Remainder of North Dakota	Central-Upper Midwest
178	39	C	Cincinnati-Wilmington-Maysville, OH-KY-IN CFS Area (OH Part)	Great Lakes
184	39	C	Cleveland-Akron-Canton, OH CFS Area	Great Lakes
198	39	C	Columbus-Marion-Zanesville, OH CFS Area	Great Lakes
212	39	C	Dayton-Springfield-Sidney, OH CFS Area	Great Lakes
99999	39	R	Remainder of Ohio	Great Lakes

MA	STATE	MA Type	Description	names
416	40	C	Oklahoma City-Shawnee, OK CFS Area	Texas and Oklahoma
538	40	C	Tulsa-Muskogee-Bartlesville, OK CFS Area	Texas and Oklahoma
99999	40	R	Remainder of Oklahoma	Texas and Oklahoma
440	41	C	Portland-Vancouver-Salem, OR-WA CFS Area (OR Part)	Pacific Northwest
99999	41	R	Remainder of Oregon	Pacific Northwest
408	42	C	New York-Newark, NY-NJ-CT-PA CFS Area (PA Part)	Northeast U.S.
428	42	C	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area (PA Part)	Northeast U.S.
430	42	C	Pittsburgh-New Castle-Weirton, PA-OH-WV CFS Area (PA Part)	Northeast U.S.
99999	42	R	Remainder of Pennsylvania	Northeast U.S.
148	44	C	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (RI Part)	Northeast U.S.
273	45	C	Greenville-Spartanburg-Anderson, SC CFS Area	Southeast U.S.
16700	45	M	Charleston-North Charleston-Summerville, SC CFS Area	Southeast U.S.
99999	45	R	Remainder of South Carolina	Southeast U.S.
99999	46	S	Remainder of South Dakota	Central-Upper Midwest
314	47	C	Knoxville-Morristown-Sevierville, TN CFS Area	Southeast U.S.
368	47	C	Memphis, TN-MS-AR CFS Area (TN Part)	Southeast U.S.
400	47	C	Nashville-Davidson--Murfreesboro, TN CFS Area	Southeast U.S.
99999	47	R	Remainder of Tennessee	Southeast U.S.
204	48	C	Corpus Christi-Kingsville-Alice, TX CFS Area	Texas and Oklahoma
206	48	C	Dallas-Fort Worth, TX-OK CFS Area (TX Part)	Texas and Oklahoma
238	48	C	El Paso-Las Cruces, TX-NM CFS Area (TX Part)	Texas and Oklahoma
288	48	C	Houston-The Woodlands, TX CFS Area	Texas and Oklahoma
12420	48	M	Austin-Round Rock, TX CFS Area	Texas and Oklahoma
13140	48	M	Beaumont-Port Arthur, TX CFS Area	Texas and Oklahoma
29700	48	M	Laredo, TX CFS Area	Texas and Oklahoma
41700	48	M	San Antonio-New Braunfels, TX CFS Area	Texas and Oklahoma
99999	48	R	Remainder of Texas	Texas and Oklahoma
482	49	C	Salt Lake City-Provo-Orem, UT CFS Area	Utah
99999	49	R	Remainder of Utah	Utah
99999	50	S	Remainder of Vermont	Northeast U.S.
545	51	C	Virginia Beach-Norfolk, VA-NC CFS Area (VA Part)	Southeast U.S.

MA	STATE	*MA Type*	Description	names
40060	51	M	Richmond, VA CFS Area	Southeast U.S.
47900	51	M	Washington-Arlington-Alexandria, DC-VA-MD-WV CFS Area (VA Part)	Southeast U.S.
99999	51	R	Remainder of Virginia	Southeast U.S.
440	53	C	Portland-Vancouver-Salem, OR-WA CFS Area (WA Part)	Pacific Northwest
500	53	C	Seattle-Tacoma, WA CFS Area	Pacific Northwest
99999	53	R	Remainder of Washington	Pacific Northwest
99999	54	S	Remainder of West Virginia	Southeast U.S.
376	55	C	Milwaukee-Racine-Waukesha, WI CFS Area	Great Lakes
99999	55	R	Remainder of Wisconsin	Great Lakes
99999	56	S	Remainder of Wyoming	Central-Upper Midwest
0000	4		Arizona - region unspecified	Arizona-Unknown Region
0000	6		California - region unspecified	California-Unknown Region

Appendix B. Nationwide Network Development

Introduction

This section covers the development of nationwide freight transport networks by road, rail and air.

Data sources

Despite the abundance of publicly available data on transportation infrastructure, there is no single source for networks for all freight transportation modes, which precludes the development of completely consistent networks.

It was possible, however, to create independent networks for all main transportation modes that operate in Arizona (truck/roadways, railways & airways) with the following sources:

- MAG modeling network
- PAG modeling network
- FAF 3.4 roadway network
- National Transportation Atlas Database (NTAD) rail network
- NTAD rail yards database
- http://openflights.org/data.html airports database
- Bureau of Transportation Statistics (BTS) T100 segments database

Zoning system

The centroid connectors included in the roadway, railway and airway networks created in this effort are divided in zone centroids (MAG, PAG, AZ Counties and remaining FAF zones), intermodal stations (truck-rail and airports) and Border crossings, which are all coded with numbers up to 10,000, as described below:

MAG TAZs: 101- 3140

PAG TAZs: 5001 – 6104

Airports: 7000 – 7764

Intermodal rail stations: 8000 – 8602

AZ counties: 9000 – 9012

FAF 3.4 TAZs: 9100 – 9220

Border crossings: 9500 – 9577 (Mexico: 9500-9515, Canada 9516-9577)

Road network

The road network is comprised of three major components: MAG 2015 modeling network, PAG 2015 modeling network and FAF 3.4 Nationwide network.

Database fields

The network includes a number of data fields that were composed by merging data from the different sources, and the resulting fields are the following (* stands for AB and BA):

- *_Lanes – Number of lanes in the given direction (Directly taken from corresponding fields in each one of the data sources)
- *_Link_Class – Link functional class. Since the link classification used in MAG, PAG and FAF networks are considerably distinct, it was necessary to build a crosswalk between such classifications in order to construct a single link classification system.

Therefore, a unified link classification was created based on the MAG link classification standard and augmented to include link classes that do not exist in the MAG network, such as ferries and local roads, as shown on Table 6.

Table B1 - Link classification

Facility Type	Description	Corresponding MAG types	Corresponding PAG types	Corresponding FAF 3.4 types
0	Freeway HOV lane	Freeway HOV lane		
1	Freeway GP lane	Freeway GP lane	Freeway (1)	Urban Interstate (11) Urban Freeway or Expressway (12)
2	Expressway	Expressway	Parkway (2)	
3	Collector/Frontage Road	Collector/Frontage Road	Collector (5) Frontage road (7)	Urban Collector (17) Rural Major Collector (07) Rural Minor Collector (08)
4	6 legged arterial	6 legged arterial		
5	Centroid connector	Centroid connector	Centroid connector (9)	
6	Arterial	Arterial	Major arterial (3) Minor arterial (4)	Urban Principal Arterial (14) Urban Minor Arterial (16) Rural Principal Arterial (02) Rural Minor Arterial (06)
7	Ramp	Ramp	Ramp (6)	
8	Metered Ramp	Metered Ramp		
9	CD road	CD road		
10	Arizona Parkway	Arizona Parkway		
11	Unpaved Road	Unpaved Road		
12	Interstate System			Rural Interstate (01)
13	Local roads			Rural Local (09) Urban Local (19)
14	Ferry			Ferry (0)

Rail network

The generation of the rail network was divided into three phases. The first phase was eliminate all links not corresponding to class1 railroads, which were the only ones considered for this high-level countrywide network.

The second phase was to guarantee the connectivity of the NTAD network and to eliminate short lines that were not connected to the overall network.

The third and last phase was to create dedicated fields identifying the links corresponding to each class1 railroad, considering both track ownership and trackage rights, which was necessary in order to compute the skims in a meaningful way, as it will be explained further in a next section of this report.

A the speed to be considered for this network is uniform for all links (22 mph), and given the large size of the network and corresponding skim matrix, only distance was included in the network, and the transportation time can be trivially derived in a later stage of the modeling work.

Air network

The definition of an air network was based on the BTS T100 segments dataset. It was assumed that all direct connection between airports for which some freight transportation had occurred (regardless of being done by freight planes or by passenger planes) corresponded to a viable link for freight transportation, and that link was added to the network.

It was also assumed that the distance between two airports was the geodesic distance, as there is no available information on actual airways between airports. As for travel time, it was assumed a constant speed of 500 mph for all links.

Intermodal facilities

Two types of intermodal facilities were added: Airports and rail yards.

A total of 765 airports in the United States were found in the BTS T100, of which 590 were found to be at most 50 miles from the roadway network established for this project (other airports were in remote areas of Alaska, Hawaii, Porto Rico, Guam, etc.), being added to the network through the addition of a single centroid connector to the roadway intersection closest to the airport.

The NTAD intermodal facilities databased included a total of 2,270 railroad yards signaled as truck-rail yards. A large portion of these yards, however, are located in large clusters of yards around major metropolitan areas and rail interchanges. These agglomerations, allied to the rather coarse zoning utilized for the other States of the Country make the consideration of all these yards absolutely unnecessary.

For this reason, a single rail yard was selected at random for each County in the US, guaranteeing a level of aggregation still greater than the zoning system. The only exception to this rule was the inclusion in the network of all the rail yards within the modeling area that perform transfers of containers between rail cars and trucks.

At this point, however, it is necessary to make it explicit that no commodity-specific networks were generated, regardless of the fact that there is a number of rail and air terminals specialized in individual commodities.

Skim matrices

After building the nationwide networks, three skim matrices were generated: Truck, truck-rail and truck-air.

As a principle, all skims were built by minimizing travel distance rather than travel time, which is standard in freight transportation models.

Truck only:

The truck only skim matrix was computed for all transportation nodes in the network, including centroids, border crossings, airports and rail yards. Along with the travel distance matrix, the corresponding travel time matrix was also computed and considered free-flow speeds for all links in the network.

Truck-rail:

The truck-rail skim matrix computation was more complex than the truck-only skim matrix, as it required several assumptions to be made in order to establish an intermodal operation rationale. The first assumption made was that the rail stations used to transfer the freight were those closest to the origin and the destination, as verified through the analysis of the truck-only skim matrix.

The second assumption was that the path utilized for rail is the one that results in the longest stretch of railroad to be performed by a single Class 1 railroad, as transferring rail cars from one railroad to another implies additional time, hence increasing the cost for the shipper.

This assumption was carried out but computing one skim matrix for each class 1 railroad, penalizing all other operators' links. When analyzing the skim results, the results chosen for each OD pair were those that resulted in the shortest distance.

As transfers between railroads are an important factor, the skims for the truck-rail mode are reported with the access truck distance, egress truck distance, total rail distance and rail distance per rail-road operator, which allows for the inference of how many rail transfers were made for each OD pair.

The computation of transfers between railroads needs to be done carefully, however, as the random selection of a single rail yard per county might be the only factor actually requiring transfers for the simulation network built in this effort.

Truck-air:

The procedure to build the truck-air skim matrix was analogous to the procedure used to the produce the truck-rail skim matrices, as the airports chosen were those closest to the origin and the destination being analyzed. For the air portion of the trip, however, only the distance and time are computed.