MID-STATES CORRIDOR

APPENDIX T

TRAVEL FORECASTING MODEL DOCUMENTATION

Mid-States Corridor
Tier 1 Environmental Impact Statement

Prepared for

Indiana Department of Transportation
Mid-States Corridor Regional Development Authority

FEBRUARY 10, 2022 *UPDATED MARCH 8, 2023*

Prepared by

Mid-States Corridor Project Consultant

INDIANA DEPARTMENT OF TRANSPORTATION INDOT

LOCHMUELLER GROUP
Reinvent Your Future

TABLE OF CONTENTS

FIGURES

TABLES

1 INTRODUCTION

This report documents the Travel Demand Model (TDM) development, application, and validation process for the Mid-State Corridor Study. This regional travel demand model is a key tool to analyze travel patterns, origin-destination (O-D) trip patterns and project benefits. It emphasizes travel within the 12-county Study Area. It also forecasts travel between the Study Area and significant portions of Indiana, Kentucky, and Tennessee. **Figure 1-1** shows the 12-county Study Area and the model boundaries.

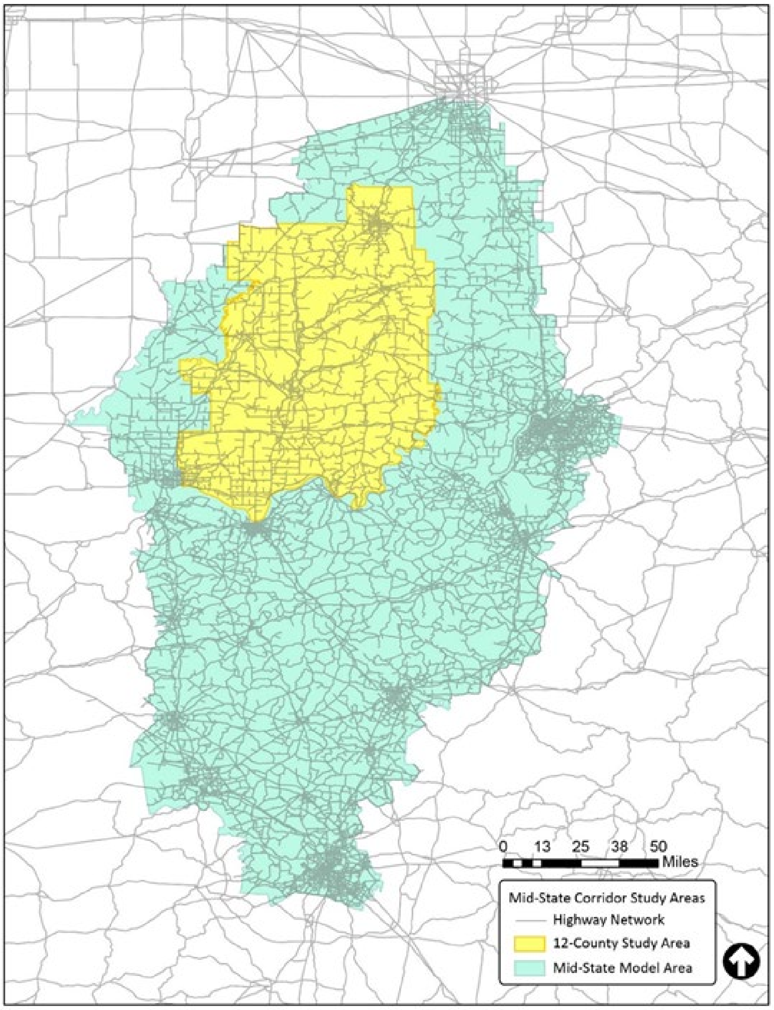


Figure 1-1: Travel Model and Study Area Boundaries

The model was developed as a three-step travel demand model. A three-step travel model is an abbreviated version of the traditional four-step travel demand model. Primary steps of a four-step travel model include: *trip generation*, *trip distribution, mode choice,* and *traffic assignment*. In a three-step travel model, *mode choice* step from the four-step model is omitted. The study area is nearly entirely rural. Automobile is the highly predominant travel mode. There is negligible use of other travel modes (e.g., walking, biking, and transit).

2 TRAVEL DEMAND MODEL ASSUMPTIONS AND COORDINATION

This section highlights major assumptions in the TDM development. It also documents model development coordination among INDOT, FHWA, local municipalities, and regional stakeholders.

Any TDM requires two large sets of input data. These data sets provide demand input and supply input. A travel model basically is an economic forecasting model. It uses standard economic relationships to forecast travel flows as the equilibrium between transportation supply and transportation demand.

Demand information includes socioeconomic data (e.g., population, household sizes, employment, income levels, etc.). These socioeconomic data determine demand for trips (types and number) within the modeled region.

Supply information includes the transportation facilities and conveyances on which travel occurs. In multi-modal models, these facilities and conveyances can include transit facilities and bus routes, rail facilities, etc. in addition to roads.

2.1 TDM Base Year

The base year model forecasts existing travel. A primary purpose for providing a base year travel model is to assess the ability of the travel model to accurately replicate travel flows. A travel model includes many detailed mathematical relationships to forecast travel demand based upon the region's socio-economic makeup (demand), the capacity of the transportation network to accommodate desired travel (supply), and achieving an equilibrium between supply and demand. These mathematical relationships include variable parameters (such as coefficients in a mathematical model). During the model development process (described in **Section 5**) these parameters are adjusted so that the base year model "predicts the present" within accepted model development standards. Once a model provides acceptable predictions of base year flows, it can serve as a basis for predicting future year travel flows.

Selecting the TDM base year considers the most current reliable socio-economic data and travel/traffic flows. As described in **Section 5**, traffic counts are used to assess the ability of the model to "predict the present." The base year for Mid-State TDM is 2017. This was the most recent year with suitable availability of socioeconomic data and traffic counts from federal and state level sources.

2.2 TDM Forecast Year

The Mid-State TDM forecast year is 2045. Traffic projections for forecast (horizon) year are used to evaluate network and traffic operational conditions and to identify future capacity needs in the regional highway network.

2.3 Existing and Committed Projects

For the forecast year 2045, a no-build highway network is defined as the base year highway network plus committed projects. “Committed” projects are funded transportation projects programmed for construction in the state DOTs’ fiscally constraint transportation plans. For the Mid-State TDM no-build network, committed projects were added from the 2045 highway networks of Indiana, Kentucky, and Tennessee statewide model highway networks. The 2045 horizon year model for the Evansville MPO (EMPO) and the 2040 horizon year model for the Kentuckiana Regional Planning and Development Agency (KIPDA) also were checked. Details of future year tolls and highway capacities crossing the Ohio River were added from the EMPO model.

2.4 Induced Growth Allocation Panel

TREDIS forecasts for each alternative forecasted induced households and employment in the 2045 forecast year for each alternative. These forecasts were for the entire 12-county study area. TREDIS runs for each alternative provided the amount of induced growth attributable to each alternative.

The allocation exercise was performed by Land Use Review Team. The Team consisted of staff from a spectrum of engineering and planning disciplines. An important reference was **Appendix U – Land Use Plan Review**. **Appendix U** reviewed and summarized local and county land-use plans within the Study Area. It identified areas targeted for future development.

The Team’s first step to identify counties where induced growth would occur. Generally, induced growth was forecasted to occur in counties where alternatives were located. For Alternative B and Alternative C, adjoining counties were identified as having the potential to receive induced growth. Counties identified as candidates to receive induced growth for each alternative were as follows:

- **Alternative B** - Spencer, Dubois, Daviess, spillover to Pike
- **Alternative C** - Spencer, Dubois, Daviess, spillover to Martin
- **Alternative M** - Spencer, Dubois, Martin and Lawrence
- **Alternative O** – Spencer, Dubois, Orange, Lawrence, Crawford
- **Alternative P** – Spencer, Dubois, Martin, Daviess, Greene
- **Alternative R** – Spencer, Dubois, Martin, Daviess, Greene

Induced households and employment were allocated in increments of 10 to candidate counties. The information in **Table 2-1** shows the results of this allocation. In some cases, no growth was allocated to counties identified as having potential to receive induced growth. **Table 2-1** shows a “0” for these counties.

For most alternatives, the number of induced jobs and households by county was similar. For these counties, a single allocation was made using the average number of induced households and jobs for the two facility types (Super-2 and Expressway). For **Alternative P**, the difference in induced households and jobs was different enough that a separate allocation was made for the two facility types.

Table 2-1: Allocation of Induced Growth by Alternative and County

	Households Allocated by Alternative							Jobs Allocated by Alternative						
County	**B**	**C**	**M**	**O**	**P2/RPA P2**	**P3/RPA P3**	**R**	**B**	**C**	**M**	**O**	**P2/RPA P2**	**P3/RPA P3**	**R**
Dubois	10	10	20	10	30	20	10	10	20	50	20	80	70	30
Pike	0							0						
Martin		0	0		20	10	5		0	0		10	10	5
Orange				0							0			
Daviess	0	10			0	0	0	10	10			0	0	0
Monroe														
Greene					20	20	10					20	10	5
Warrick														
Spencer	0	0	10	0	10	10	5	0	10	10	0	20	10	5
Perry														
Crawford				0							0			
Lawrence			20	10						20	10			
Total	**10**	**20**	**50**	**20**	**80**	**60**	**30**	**20**	**40**	**80**	**30**	**130**	**100**	**45**

After this county level allocation was made, the Team made suballocations of induced growth to individual TAZs. These were based upon local and county land use plans cited above.

2.5 Agency Coordination

Project staff conferred with INDOT, FHWA and several MPOs throughout the TDM development process. For the Mid-States TDM TAZs in Indiana, 2045 socioeconomic projections (total population in 2045 and growth in employment from 2017 to 2045) were obtained from the ISTDM TAZs 2045 projections. The Mid-States TDM base highway network for Indiana was also based on ISTDM base highway network which was received from INDOT in October, 2019.

Mid-States Tier 1 EIS staff also conferred with the following MPOs during the TDM development process. These contacts occurred between October of 2019 and January of 2020. Each MPO provided model network and TAZ information from its current travel forecasting model.

- Evansville Metropolitan Planning Organization (EMPO)
- Bloomington Monroe County Metropolitan Planning Organization (BMCMPO)
- Kentuckiana Regional Planning and Development Agency (KIPDA)

Initial contacts with INDOT Planning and Programming (P & P) occurred in early November, 2019. This was during the model development stage. At that time, Lochmueller Group stated its intent to assume I-69 Ohio River crossings in Evansville would be non-tolled for the Screening of Alternatives, only. These assumptions would be revisited as model development continued.

For the Mid-State TDM TAZs within the MPO boundaries cited above, 2045 ISTDM projections were compared with the MPOs' socioeconomic projections and necessary adjustments were made when MPO projections were more reasonable. Staff communicated with INDOT P & P regarding updates to socio-economic data for TAZs using MPO socio-economic information. A memo was sent to INDOT on February 7, 2020 highlighting updated socio-economic information for some selected TAZs where socio-economic data from the MPOs were used instead of ISTDM information. INDOT P & P acknowledged and approved these modifications on February 24, 2020.

Project staff conferred with INDOT while finalizing socio-economic forecasts for the TAZs within Indiana. Staff prepared a June 26, 2020 memo to INDOT's P & P Division documenting the population and employment growth forecasts for the 12 counties within the Study Area. Staff received INDOT's approval on July 20, 2020. The final population forecasts were taken from the ISTDM TAZ information, as modified to incorporate some MPO model forecasts. The final employment forecasts were based on employment data from the Longitudinal Employer-Household Dynamics dataset in the American Community Survey. The employment growth to 2045 was forecasted using the employment growth in the ISTDM between the base and forecast years.

Project staff initially contacted the Kentucky Transportation Cabinet (KYTC) in May 2019 and requested TAZ, highway network and traffic count files from its statewide model. KYTC provided these files in August 2019.

Project staff contacted the Tennessee Department of Transportation (TDOT) in May 2019 and requested TAZ, highway network and traffic count files from its statewide model. TDOT provided this information in June 2019.

Mid-States Tier 1 EIS staff conferred with FHWA regarding TDM development approach and methodology. Staff provided a June 25, 2019 memo to FHWA highlighting the TDM approach and methodology. Staff received FHWA's reply on July 2, 2019. Mid-States EIS staff and FHWA had a follow up conference call on August 5, 2020 and reviewed the TDM approach and methodology in detail. FHWA Indiana Division and FHWA Resource Center staff participated in the meeting. TDM approach, methodology and land use forecasts were discussed. FHWA staff stated that the proposed approach would adequately address land use impacts of the proposed Mid-States alternatives.

3 TRAVEL DEMAND MODEL STRUCTURE

3.1 Network Development

This section presents the development of the roadway network and attributes for the TDM. The areas of the Mid-State TDM geographic structure includes the following:

- Twelve county Study Area in Indiana
- Rest of Indiana within the TDM boundary
- Portion of Kentucky within the TDM boundary
- Portion of Tennessee within the TDM boundary

Figure 3-1 on the next page shows the Mid-State TDM network structure for the base (2017) year.

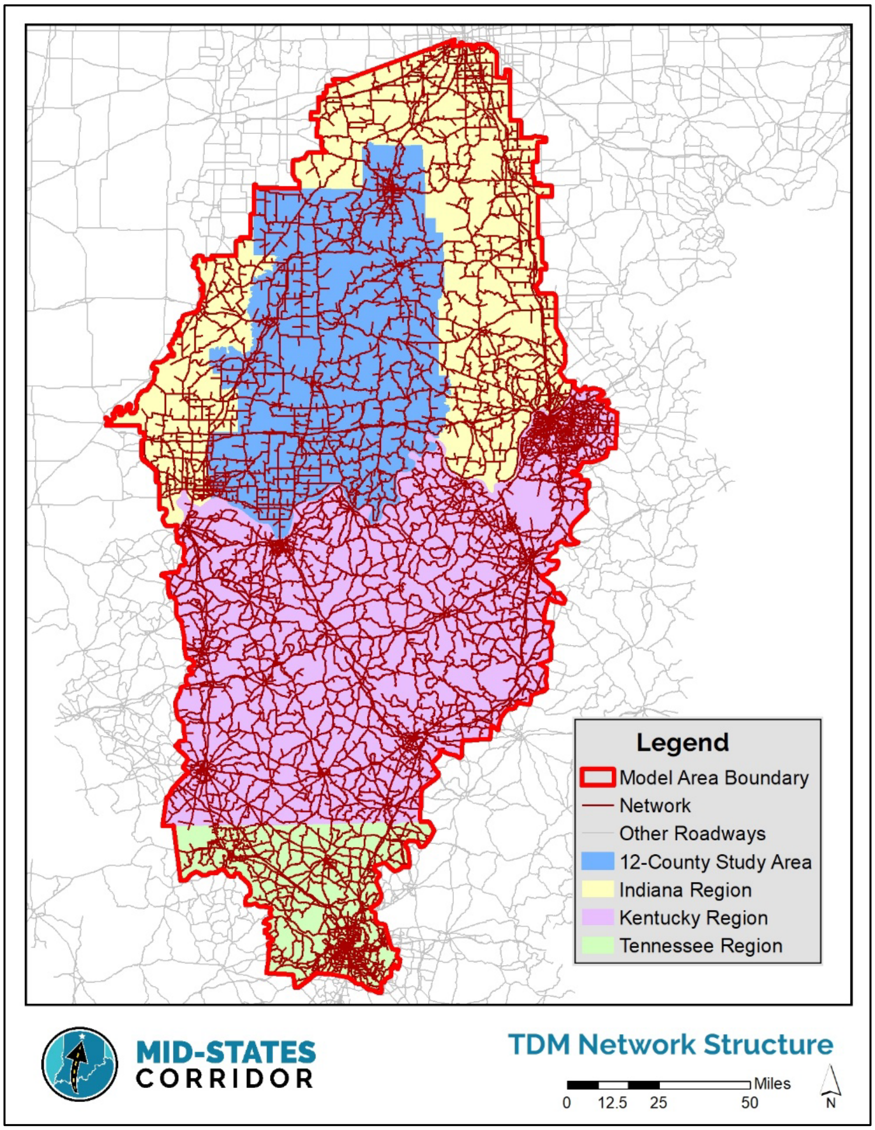


Figure 3-1: Mid-State TDM Base Year Network

3.1.1 12 County Study Area in Indiana

For the 12 counties in the Mid-State study area, the roadway network was developed by using the INDOT's Statewide Model (INSWM) roadway network (updated in 2019). Additional roadway links were considered for major population areas through careful evaluation of regional roadway networks in each county. Figure 3.2 shows new roadway links (in red) added for Jasper. Network not shown in the INSWM also was added in Huntingburg and Bedford.

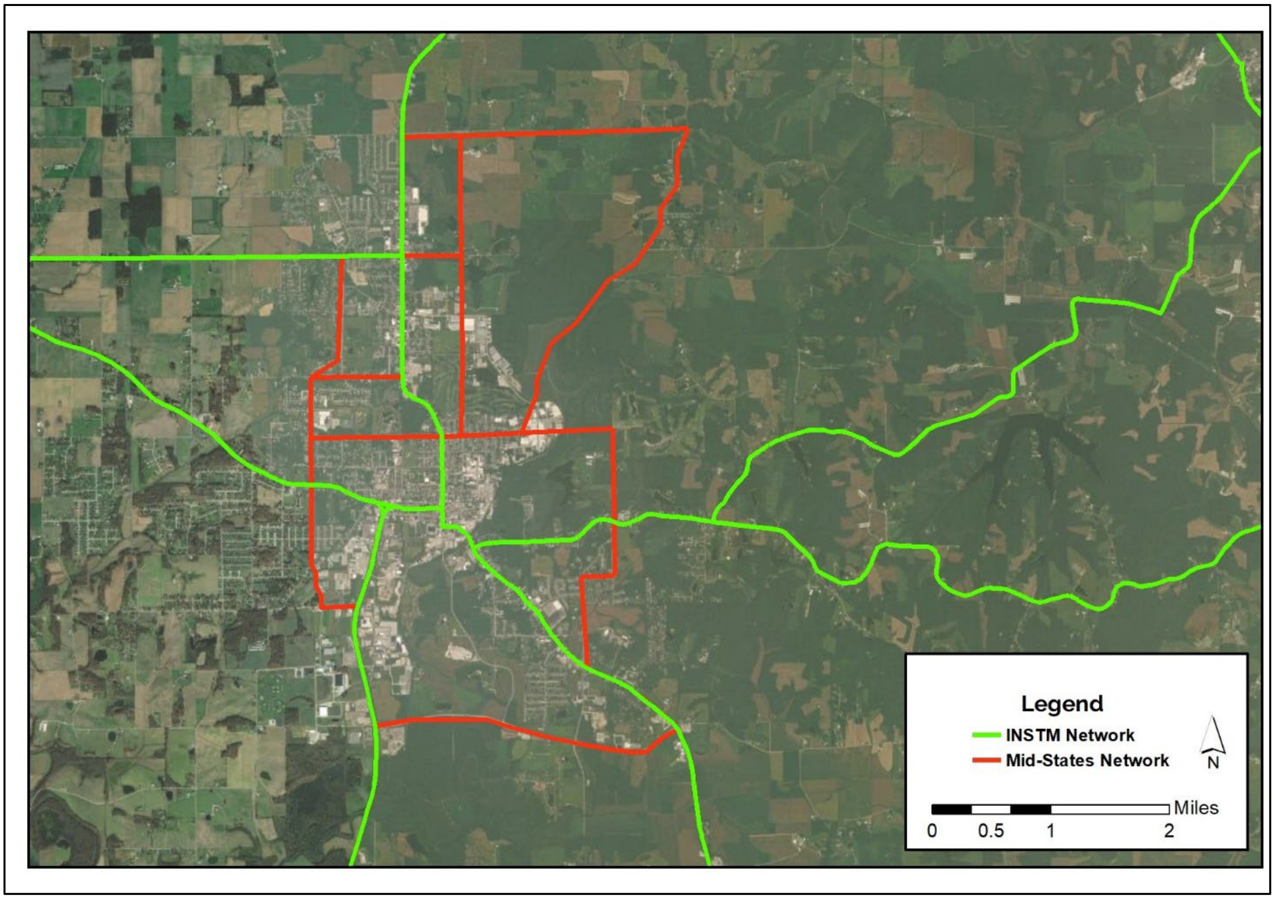


Figure 3-2: Added Base Year Network in Jasper

3.1.2 Rest of Indiana within the TDM Boundaries

Roadway network for Indiana outside the 12-county region were developed using the INSWM roadway network.

3.1.3 Portion of Kentucky within the TDM Boundaries

The Mid-State TDM boundaries include a significant area in Kentucky. Kentucky Transportation Cabinet's (KYTC) Statewide TDM roadway network was used to develop the roadway network for the model areas within Kentucky.

3.1.4 Portion of Tennessee within the TDM Boundaries

The Mid-State TDM includes some areas in Tennessee. Tennessee Department of Transportation's (TDOT) TDM roadway network was used to develop the roadway network for the model areas within Tennessee.

Roadway network from the three states were merged carefully to develop the combined Mid-State roadway network.

3.1.5 Centroid Connectors

Centroid connectors for the Mid-State TDM network were added to the roadway links by using automated TransCAD procedures. Centroid connector development included the following key considerations:

- Initial centroid connector generation process restricted centroid connectors to ensure that the connectors do not cross TAZ boundaries.
- Centroid connectors were not permitted to connect to freeways, ramps, or one-way links.
- Centroid connector lengths were restricted. The maximum length of a centroid connecter is 7.5 miles. Within Indiana, the longest centroid connector is 7.5 miles.
- Each TAZ was allowed to have a maximum of three centroid connectors.

Careful manual review of TransCAD's automated centroid connector generation process ensured that all TAZs have at least one centroid connector. Key attributes for the centroid connectors are shown in **Table 3.1-1.**

Table 3.1-1: Attributes of Centroid Connectors

Attribute	Description
FCLASS	99
THRU_LANES	20
AB_LANES	10
BA_LANES	10
SPD_LIMIT	45
MSFFS	45
MSHRCAP	2000

Figure 3-3 shows centroid connectors in the roadway network in and near Jasper, Indiana.

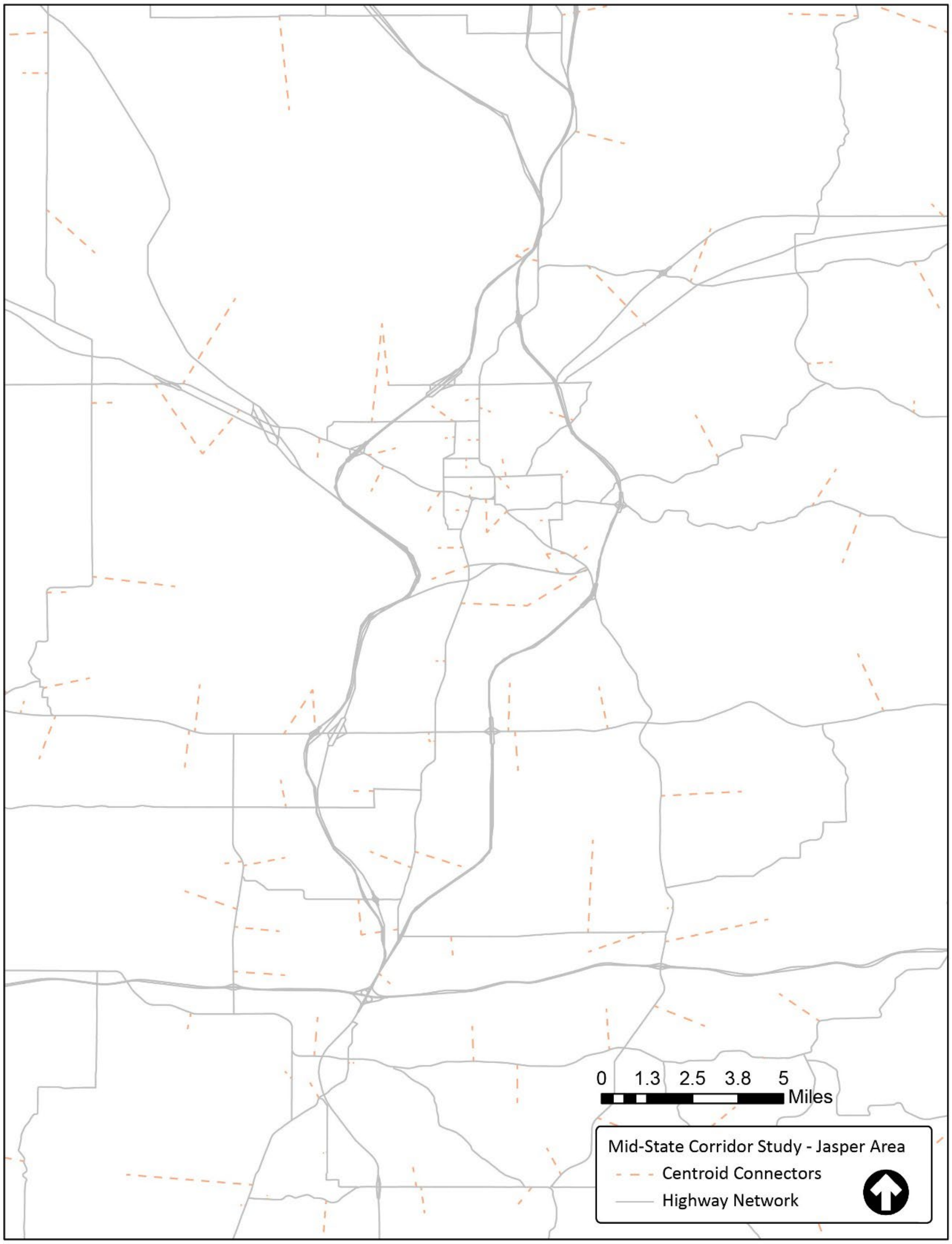


Figure 3-3: Centroid Connectors in Roadway Network in Jasper

3.1.6 Roadway Link Attributes

Roadway link attributes were developed for the Mid-State TDM using information in roadway networks of the Indiana, Kentucky and Tennessee statewide models. The INSWM roadway network contained roadway functional class information following the prior (pre-2010) Federal Highway Administration's (FHWA) functional classification system. All roadway links in Indiana were assigned appropriate functional classes following the current FHWA Roadway Functional Classification system. **Table 3.2** shows the new and old FHWA roadway functional classes.

Table 3.1-2: FHWA Functional Classes and Code Descriptions

Functional Class	New Code (post 2010)	Old Code (Pre 2010)
Rural Interstate	1	1
Rural Expressways	2	Didn't Exist
Rural Other Principal Arterial	3	2
Rural Minor Arterial	4	6
Rural Major Collector	5	7
Rural Minor Collector	6	8
Rural Local Access	7	9
Urban Interstate	1	11
Urban Expressways	2	12
Urban Other Principal Arterial	3	14
Urban Minor Arterial	4	16
Urban Major Collector	5	17
Urban Minor Collector	6	18
Urban Local Access	7	19

Under the post-2010 classification scheme, facilities are specified by a combination of facility types and area types. This information must be specified to identify default attributes (such as capacity) of each facility.

Brief descriptions of major roadway functional class mentioned in Table 3.1-2 include:

Interstates: These roadways are the highest classification of arterials with limited access and designed for high level of mobility by linking urban areas. Interstate highways are officially designated as "Interstates" by the US Secretary of Transportation.

Expressways: This roadway functional classification category is very similar to Interstates with strict access control and designed for long-distance travel. However, they do not have an official "Interstate" designation.

Other Principal Arterials: Major roadways serving major metropolitan areas and rural areas by providing high level of mobility. Abutting land uses may be served directly.

Minor Arterials: These roadways serve trips of moderate length and typically serve smaller geographic areas by offering connectivity to higher arterials.

Major and Minor Collectors: These roadways are crucial for providing connectivity between local roads and arterials. Typically, Major Collector routes have higher speed limits and lower access densities than their Minor Collector counterparts.

Local Roads: These roadways provide direct connections abutting land and are not designed for through traffic.

Table 3.1-3 shows the link attributes and brief descriptions for the Mid-State roadway network. Speed and capacity determinations are described in **Section 3.1.7**.

Table 3.1-3: Mid-State Roadway Network Attributes and Descriptors

Attribute	Description	Values
DIR	Direction	1 or -1 for One-way Links
		0 - Two -way links
FUNCCLASS	Functional Class	1- Interstate
		2- Freeway or Expressway
		3- Other Principal Arterial
		4- Minor Arterial
		5- Major Collector
		6- Minor Collector
		7- Local Road or Street
		8- Ramp (All Ramps)
SPD_LMT	Posted Speed Limit	Numeric Integer Value
THRU_LANES	Total Number of Lanes	Numeric Integer Value
REGION	Type of Area	Urban
		Rural
AB_LANES	Lanes in AB Direction	Numeric Integer Value
BA_LANES	Lanes in BA Direction	Numeric Integer Value
MSFFS	Base Free Flow Speed	Numeric Integer Value
MSHRCAP	Base Hourly Lane Capacity	Numeric Integer Value
STATE	Region in which the links are	Indiana
		Kentucky
		Tennessee
AADT	Field AADT from 2017	Numeric Integer Value
AADT_SINGL	AADT Single Axle Trucks 2017	Numeric Integer Value
AADT_COMBI	AADT Combined Axle Trucks 2017	Numeric Integer Value

Figure 3-4, found on the next page, shows roadway functional classes for the Mid-State network.

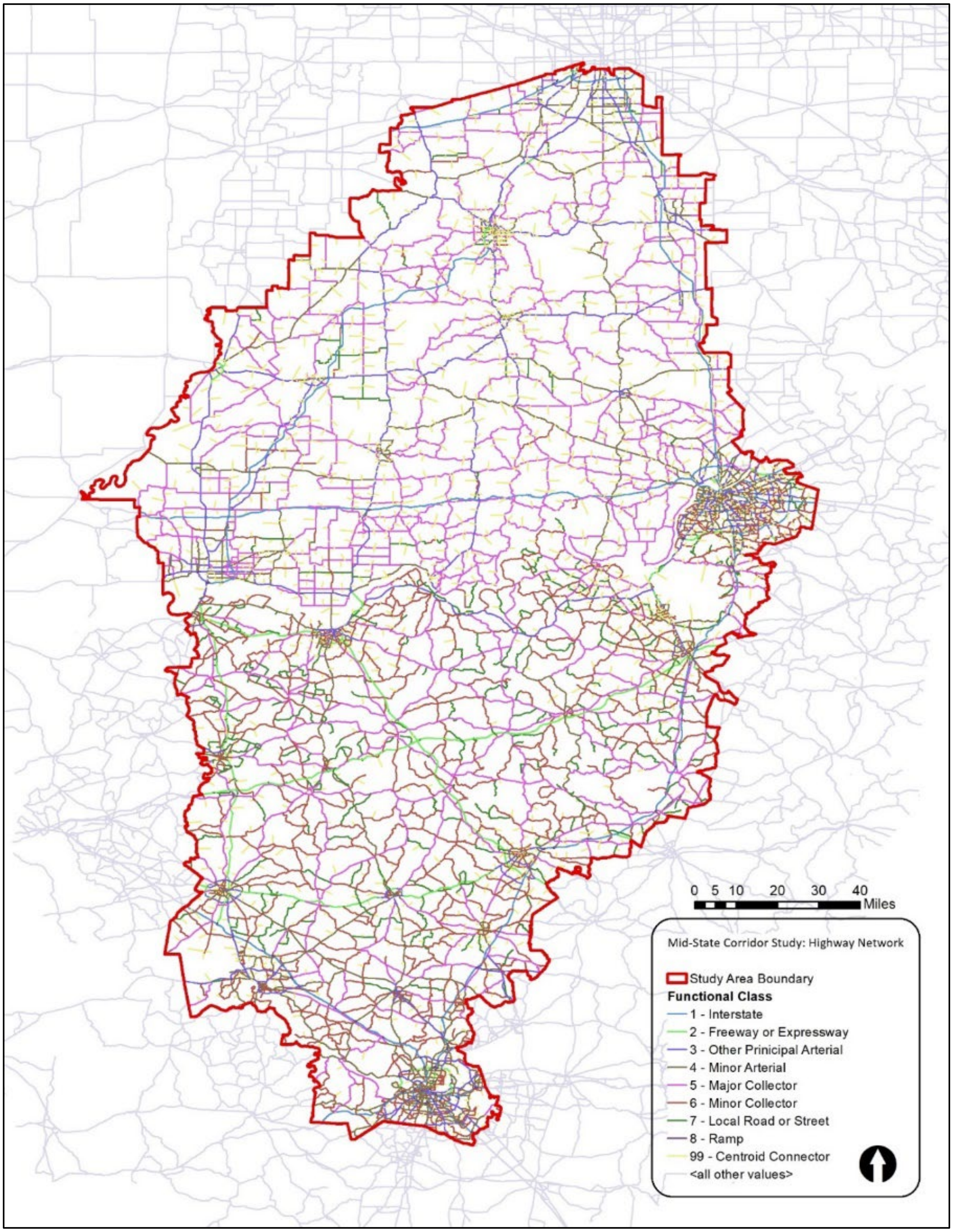


Figure 3-4: Mid-State Roadway Network Functional Classes

3.1.7 Speed-Capacity Estimation

The speed-capacity estimation for highway network links was based on detailed review of relevant research works and detailed understanding of the roadways within the model boundaries. Estimated free flow speed based on the Highway Capacity Manual (HCM) procedures typically underestimate free flow speed when field observed free flow speed is higher than 40 mph.[1] Field observed free flow speeds were typically found five percent to 15 percent higher than the posted speed limits for different roadways in both rural and urban areas.[2] Considering these facts and detailed evaluation of the roadway posted speed limits in the highway network, free flow speed for the highway network links were estimated as 10 percent higher than the posted speed limits.

Highway network link capacities (vehicle per hour per lane) were assigned based on roadway functional class and area type designations. **Table 3.1-4** shows link capacity (vehicle per hour per lane) for the roadways in the Mid-State network.

Table 3.1-4: Mid-State Roadway Network Link Capacity

Functional Class	Urban	Rural
1- Interstate	1800	1800
2 - Freeway or Expressway	1800	1800
3 - Other Principal Arterial	1500	1600
4- Minor Arterial	1300	1500
5- Major Collector	1100	1200
6- Minor Collector	400	500
7- Local Road or Street	500	600
8- Ramp (All Ramps)	1300	1300

3.1.8 Traffic Counts

Traffic count data for the Mid-State model area were obtained from the following sources:

- Indiana: Indiana Department of Transportation
- Kentucky: Kentucky Transportation Cabinet
- Tennessee: Tennessee Department of Transportation
- Direct Field Counts: Traffic count data was collected at 20 intersections and 15 roadway segments within the 12-County study area in late summer/early fall of 2019. The count locations were selected based on location of major population areas and roadways with higher regional importance (e.g., higher functional class). Field collected traffic count data supplemented INDOT counts.

[1] Evaluation of Free Flow Speeds on Interrupted Flow Facilities, Florida Department of Transportation, May 2013

[2] Development of Speed Models for Improving Travel Forecasting and Highway Performance Evaluation, Florida Department of Transportation, December 2013

INDOT's Statewide Model roadway network contained the traffic count Station ID information for specific roadways. INDOT's online MS2 Traffic Count Database System (TCDS) contains latest Annual Average Daily Traffic (AADT) volumes for the stations located throughout the state. AADT volumes for vehicles and trucks for the selected stations for 2016-2018 were obtained from the TCDS database.

KYTC's Highway Information System (HIS) database was used to obtain traffic counts for the roadway links with traffic count stations within Kentucky state boundary.

Traffic count data for the Tennessee portion of the roadway network were available in GIS shapefile format from TNDOT.

Traffic count data from these sources were in different format. AADT volumes for vehicles and trucks were carefully reviewed to create a combined traffic count file. Two-letter prefixes were added for each state (e.g., IN, KY, TN) with the traffic count station IDs to confusing station IDs from different states. Due to inconsistent data collection years and a lack of detailed long-term trend data across count stations, traffic counts from 2016, 2017, and 2018 were used, as available and without adjustment. **Table 3.1-5** shows sample traffic count data for different traffic count stations within the Mid-State model boundaries.

Table 3.1-5: Sample Traffic Count Data within Mid-State Model Boundaries

Station ID	AADT_16	AADT_17	AADT_18	SUTrk_16	SUTrck_17	SUTrck_18	MUTrck_16	MUTrck_17	MUTrck_18
IN100230	11,400	11,503	11,549	1,453	1,466	1,472	476	480	482
IN100292	7,045	7,108	7,136	542	547	549	69	70	70
IN100300	5,626	5,598	5,492	593	590	579	106	105	103
IN100301	6,321	6,289	6,170	520	517	507	44	44	43
KY001A43	0	0	13,546	0	0	0	0	0	0
KY001A46	0	12,034	0	0	513	0	0	241	0
KY001A70	5,973	0	0	430	0	0	194	0	0
KY001A74	1,270	0	0	0	0	0	0	0	0
KY001A75	1,375	0	0	0	0	0	0	0	0
KY001A76	1,448	0	0	0	0	0	0	0	0

3.2 TAZ and Socioeconomic Data Development

This section describes the data sources and methodology for the development of Traffic Analysis Zones (TAZs) and socioeconomic data for the Mid-State model area, including detailed quality reviews and checks. As described in **Section 3.1**, the areas of the Mid-State TDM geographic structure includes the following:

- Twelve county Study Area in Indiana
- Portions of Indiana within the TDM boundary
- Portions of Kentucky within the TDM boundary
- Portions of Tennessee within the TDM boundary

3.2.1 Twelve County Study Area in Indiana

The Mid-States Study Area includes the counties bounded by I-69 on the west and north, SR 37 on the east, and the Ohio River on the south. These were selected as the project study area because they may experience noteworthy changes in traffic patterns due to the project. **Table 3.2-1** shows the twelve counties along with their Federal Information Standard (FIPS[3]) codes.

Table 3.2-1: FIPS Codes for Twelve Mid-States Counties

#	FIPS Code	Name
1	18147	Spencer County
2	18125	Pike County
3	18123	Perry County
4	18117	Orange County
5	18173	Warrick County
6	18037	Dubois County
7	18055	Greene County
8	18025	Crawford County
9	18027	Daviess County
10	18101	Martin County
11	18105	Monroe County
12	18093	Lawrence County

[3] Federal Information Processing Series (FIPS) are numeric codes assigned by the National Institute of Standards and Technology (NIST). Typically, FIPS codes deal with US states and counties. US states are identified by a 2-digit number, while US counties are identified by a 3-digit number.

3.2.2 Other Portions of Indiana within the TDM Boundary

The Mid-States TDM boundaries also include all or parts of 21 additional counties in Indiana. **Table 3.2-2** lists Indiana counties within the TDM boundary but outside the 12 county Study Area.

Table 3.2-2: FIPS Codes for Indiana Counties within TDM Boundary, but Outside Mid-States Study Area

#	FIPS Code	Name	Scale*
1	18005	Bartholomew County	P
2	18013	Brown County	F
3	18019	Clark County	P
4	18043	Floyd County	F
5	18051	Gibson County	F
6	18061	Harrison County	F
7	18063	Hendricks County	P
8	18071	Jackson County	F
9	18079	Jennings County	P
10	18081	Johnson County	P
11	18083	Knox County	P
12	18097	Marion County	P
13	18109	Morgan County	F
14	18119	Owen County	P
15	18129	Posey County	P
16	18133	Putnam County	P
17	18143	Scott County	P
18	18145	Shelby County	P
19	18153	Sullivan County	P
20	18163	Vanderburgh County	F
21	18175	Washington County	F

***Counties are either fully (F) or partially (P) within the TDM boundaries**

3.2.3 Portions of Kentucky within the TDM Boundary

The Mid-States TDM area includes a significant portion of Kentucky. **Table 3.2-3** shows 26 Kentucky counties which are either fully or partially included in the TDM boundary.

Table 3.2-3: Kentucky Counties Fully or Partially within TDM Boundary

#	FIPS Code	Name	Scale*
1	21009	Barren County	P
2	21027	Breckinridge County	F
3	21029	Bullitt County	P
4	21031	Butler County	F
5	21047	Christian County	P
6	21059	Daviess County	F
7	21061	Edmonson County	F
8	21085	Grayson County	F
9	21091	Hancock County	F
10	21093	Hardin County	P
11	21099	Hart County	P
12	21101	Henderson County	P
13	21107	Hopkins County	P
14	21111	Jefferson County	P
15	21123	Larue County	P
16	21141	Logan County	F
17	21149	McLean County	F
18	21163	Meade County	F
19	21177	Muhlenberg County	F
20	21179	Nelson County	F
21	21183	Ohio County	F
22	21185	Oldham County	F
23	21213	Simpson County	F
24	21219	Todd County	F
25	21227	Warren County	P
26	21233	Webster County	P

***Counties are either fully (F) or partially (P) covered by the TDM boundaries**

3.2.4 Portions of Tennessee within the TDM Boundary

The Mid-State TDM includes portions of the State of Tennessee. **Table 3.2-4** shows the five Tennessee counties which are either fully or partially included in the TDM boundary.

Table 3.2-4: Tennessee Counties Fully or Partially within TDM Boundary

#	FIPS Code	Name	Scale
1	47021	Cheatham County	F
2	47037	Davidson County	F
3	47125	Montgomery County	P
4	47147	Robertson County	F
5	47165	Sumner County	P

***Counties are either fully (F) or partially (P) covered by the TDM boundaries**

Figure 3-5 shows the Mid-States TDM TAZ boundaries by area.

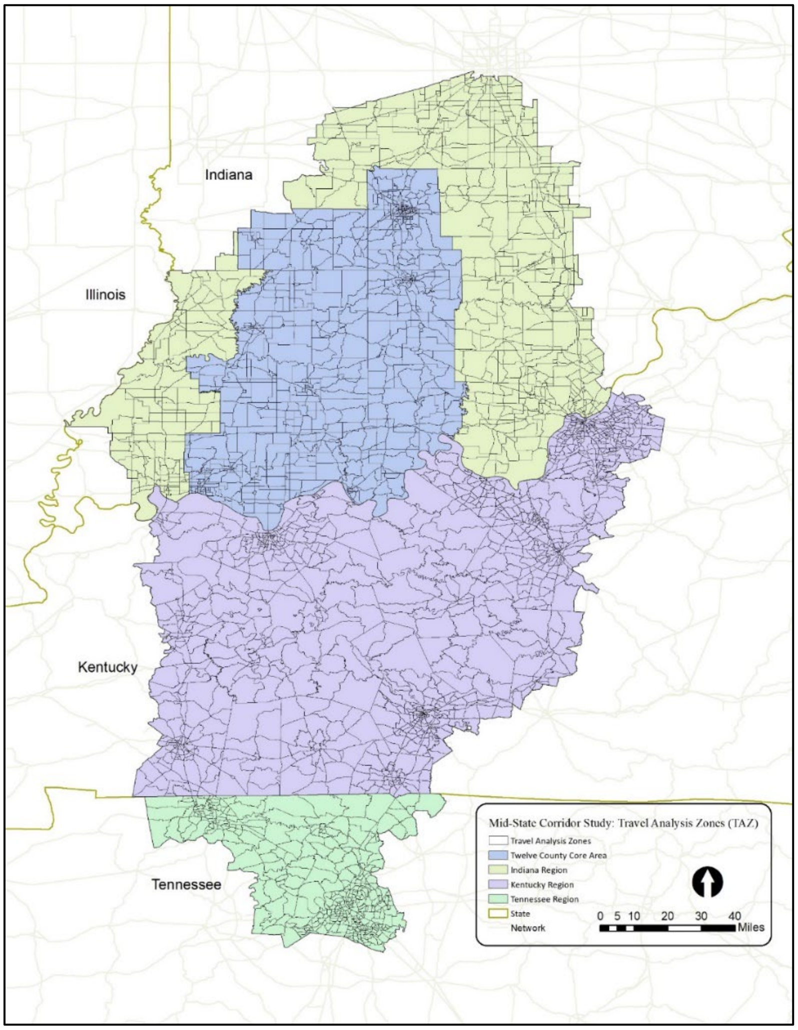


Figure 3-5: Mid-States TDM TAZ Boundaries by Area

3.2.5 Data Sources

Key data sources for developing Mid-States TDM TAZ boundaries include the following:

- US Census boundaries (Blocks, Block Groups, and Tracts)
- Current version of the INDOT's Statewide Model (INSWM) TAZ boundaries
- Current version of the KYTC's Statewide Model (KYSWM) TAZ boundaries
- Current version of the TNDOT's Statewide Model (TNSWM) TAZ boundaries

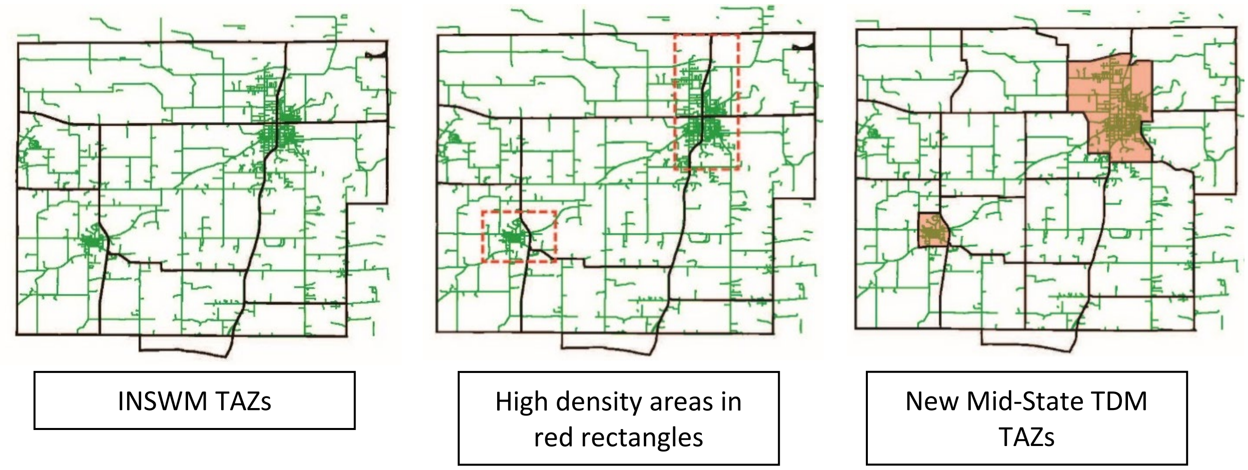
In addition to the above-mentioned sources, web-based imagery (Google Maps and ESRI) and TransCAD data GIS files (highways, water bodies) were used to review and modify TAZ boundaries.

3.2.6 TAZ Development Methodology

Mid-States TDM TAZ development in the twelve-county core area started with the INSWM. INSWM TAZ boundaries within the twelve-county core area were not detailed enough for the Mid-States TDM. INSWM TAZs within the twelve-county area were disaggregated through careful review of existing land-uses, presence of natural and human-made barriers, and existing political and planning boundaries. Web-based satellite imagery supplemented the process of separating town centers (densely populated areas) from other areas. Disaggregation process followed a strong protocol of preserving the overall boundaries of the INSWM TAZ boundaries.

INSWM had 460 TAZs in the Twelve-County area. Through this disaggregation process, an additional 366 TAZs were developed in the twelve-county Study Area. The Study Area has a total of 826 TAZs.

Figure 3-6 shows examples of disaggregating INSWM TAZs in the twelve-county core area by creating separate TAZs for the town centers or high-density land-uses from relatively low-density land-uses.



INSWM TAZs

High density areas in red rectangles

New Mid-State TDM TAZs

Figure 3-6: Examples of INSWM TAZs in Mid-States Study Area

TAZ boundaries for rest of Indiana within the Mid-States TDM area were the same as the INSWM TAZ boundaries with a few exceptions. Some INSWM TAZ boundaries were updated though careful review of natural and human-made barriers and political and planning boundaries.

Mid-State TDM TAZs in Kentucky followed the KYSWM TAZ boundaries with a few exceptions. Some KYSWM TAZ boundaries were updated though careful review of natural and human-made barriers and political and planning boundaries.

Mid-States TDM TAZ boundaries in Tennessee followed the TNSWM TAZ boundaries.

Table 3.2-5 shows the total number of TAZs in different geographies of the Mid-States TDM.

Table 3.2-5: Total Number of TAZs in Different Geographies of the Mid-States TDM

Geographies	# of TAZs
Twelve-County Core Area	826
Indiana	625
Kentucky	1,228
Tennessee	378
Total	**3,057**

3.2.7 Socioeconomic Data for Base Year 2017

Socioeconomic data is one of the key inputs for the Mid-State TDM. Due consideration was given to careful develop the socioeconomic attributes for each TAZs. Key socioeconomic attributes for each TAZs and their sources are shown in **Table 3.2-6**.

Table 3.2-6: TAZ Key Socioeconomic Attributes

Socioeconomic Attributes	Data Source	ACS Table
Total Population	2017 ACS Five Year Estimates	B01003
Household Population (HHPOP)	2017 ACS Five Year Estimates	TOTPOP - GQPOP
Group Quarter Population	2017 ACS Five Year Estimates	B26001
Total Households (HH)	2017 ACS Five Year Estimates	B11011
Average Household Size	2017 ACS Five Year Estimates	HHPOP / HH
Average Household Income	2017 ACS Five Year Estimates	S1901
K-12 School Enrollment (K12)	2017 ACS Five Year Estimates	S1401
Average Household Students	2017 ACS Five Year Estimates	K12 / HH
Average Household Workers	2017 ACS Five Year Estimates	B08202
Average Household Vehicles	2017 ACS Five Year Estimates	B08201
Household Seniors	2017 ACS Five Year Estimates	S1101
College/University Enrollment	2017 ACS Five Year Estimates	S1401

Socioeconomic data available from the different statewide models had different base years (e.g., Kentucky 2010, Indiana 2015 and Tennessee 2017). For the model's 2017 base year, demographic data used 2017 ACS five-year estimate data for all attributes by disaggregating county level data to the TAZ level. The first step was to assign 2010 census blocks to Mid-State TAZs and then calculating disaggregation factors as the ratio of block level households and population to their respective county control totals. The developed disaggregation factors were used to allocate base year (2017) county level data to each block within a TAZ.

The majority of census blocks were compatible with the TAZ boundaries. For those that were not, disaggregation at block level was done by calculating the ratio of the area of the block in TAZ to the total area of the block. This proportion was factored into block's households and population to get a uniform distribution of household and population data.

Block level data were aggregated at TAZ geography scale. The TAZ level demographic data was reviewed and checked for reasonableness. In some cases, Google images were used to check for areas with high and low concentration of households or population. Both **Figure 3-7** and **Figure 3-8** show base year population and household by TAZs in the Mid-States TDM respectively.

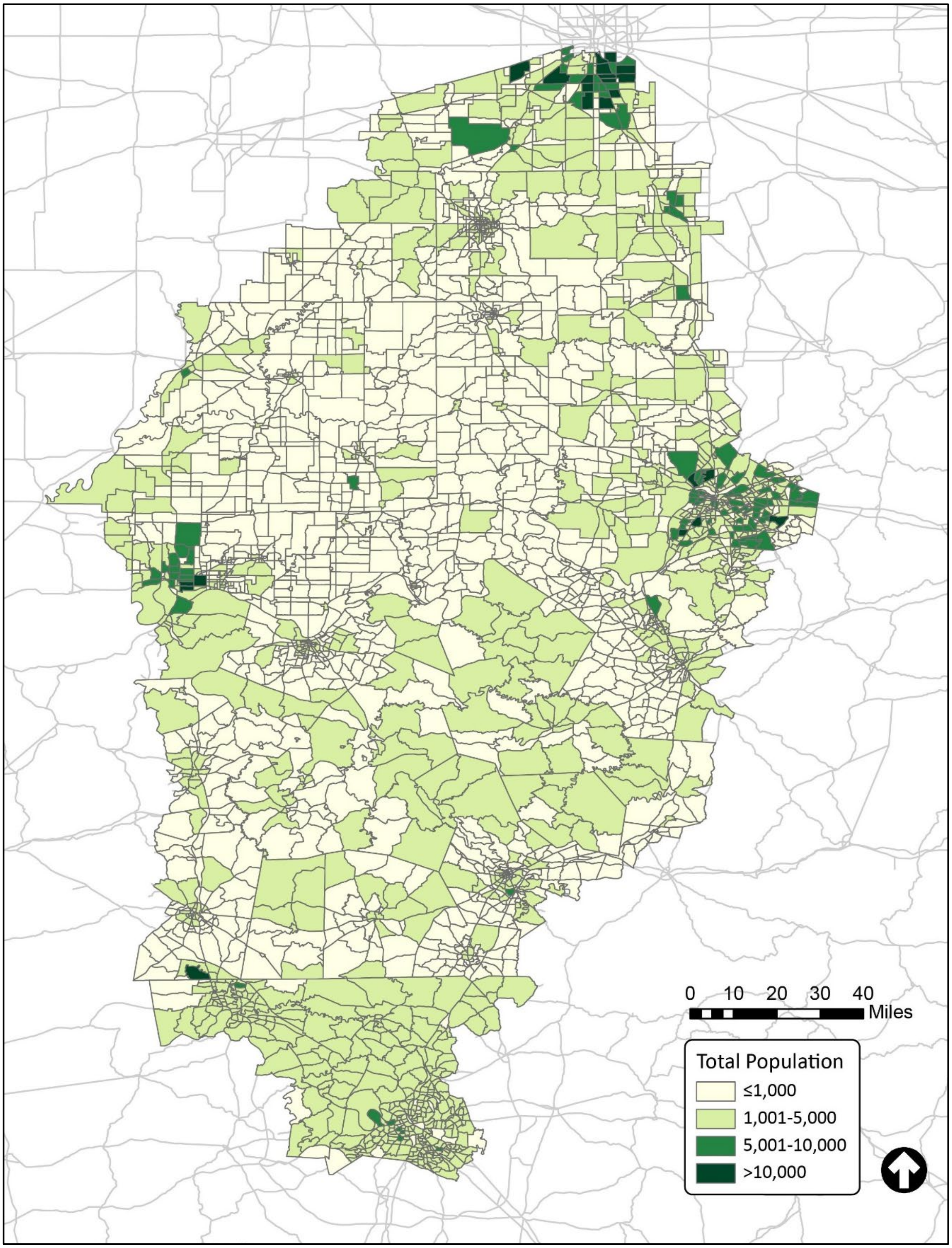


Figure 3-7: Base Year Population Totals by TAZs in the Mid-States TDM

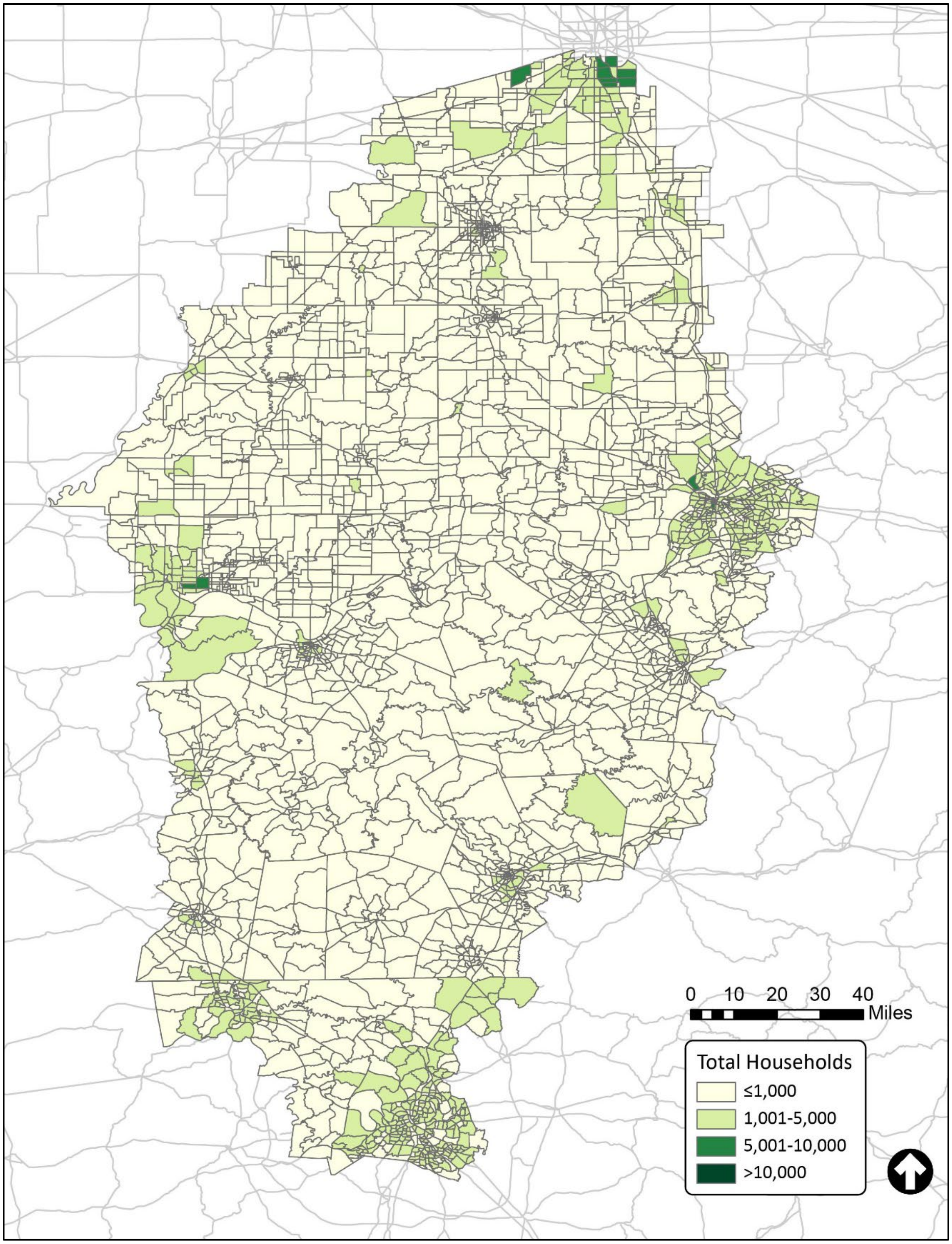


Figure 3-8: Base Year Household Totals by TAZs in the Mid-States TDM

The employment categories for the Mid-States TDM TAZs and the sources used for development of the employment data are shown in **Table 3.2-7**. The employment data was developed from Longitudinal Employer-Household Dynamics (LEHD) data with the most recent LEHD data available from 2017.

Table 3.2-7: Mid-States TDM TAZs Employment Categories

Employment Categories by North American Industry Classification System (NAICS) Codes	
Agriculture, Forestry, Fishing and Hunting	NAICS sector 11
Mining, Quarrying, and Oil and Gas Extraction	NAICS sector 21
Utilities	NAICS sector 22
Construction	NAICS sector 23
Manufacturing	NAICS sector 31-33
Wholesale Trade	NAICS sector 42
Retail Trade	NAICS sector 44-45
Transportation and Warehousing	NAICS sector 48-49
Information	NAICS sector 51
Finance and Insurance	NAICS sector 52
Real Estate and Rental and Leasing	NAICS sector 53
Professional, Scientific, and Technical Services	NAICS sector 54
Management of Companies and Enterprises	NAICS sector 55
Administrative and Support and Waste Management and Remediation Services	NAICS sector 56
Educational Services	NAICS sector 61
Health Care and Social Assistance	NAICS sector 62
Arts, Entertainment, and Recreation	NAICS sector 71
Accommodation and Food Services	NAICS sector 72
Other Services [except Public Administration]	NAICS sector 81
Public Administration	NAICS sector 92

Development of the employment data for the Mid-State TDM TAZs included the development of a geographic link between the TAZs and census blocks by allocating each census block to a zone. Once each block was allocated to the TAZ, the 2017 LEHD data for total employment and employment category by the North America Industry Classification System (NAICS) for each block was aggregated by TAZ. Specific quality checks of the employment data included:

- Ensuring the sum of all employment categories for each TAZ was equal to the total employment in each TAZ
- Developing thematic maps for total employment to visually check high and low employment zones using Google Earth images to make sure they were consistent with the actual land use.

Figure 3-9 shows base year total employment for each TAZs in the Mid-States TDM.

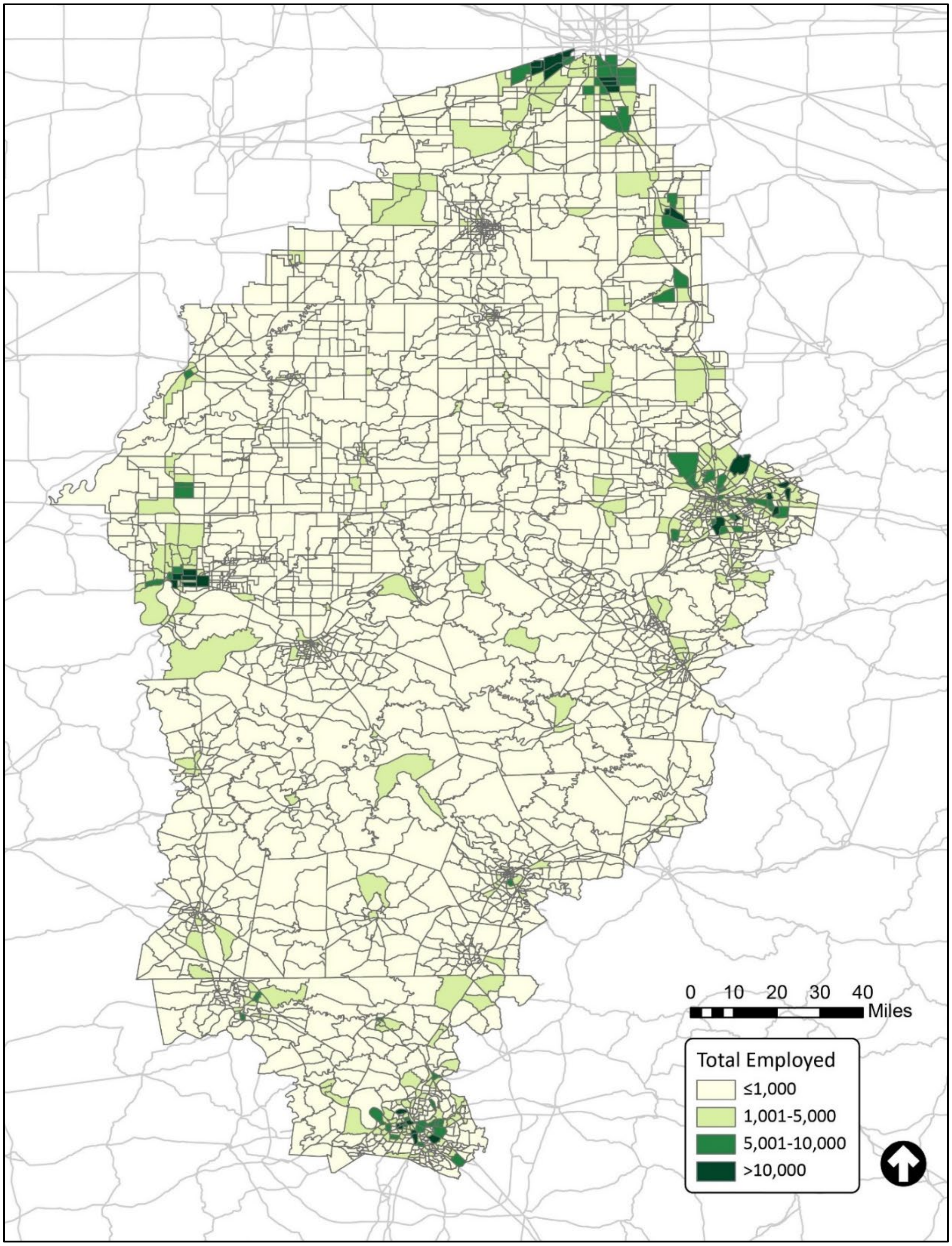


Figure 3-9: Base Year Employment Totals by TAZs in the Mid-States TDM

3.2.8 Socioeconomic Data for Forecast Year 2045

Horizon year (2045) socioeconomic data for the Mid-State TDM TAZs were obtained from the following sources:

- Indiana Statewide Model 2045 Socioeconomic projections
- Kentucky Statewide Model 2040 socioeconomic projections
- Tennessee Statewide Model 2045 socioeconomic projections
- Evansville Metropolitan Planning Organization (EMPO) 2045 socioeconomic projections
- Kentucky Regional Planning and Development Agency (KIPDA) 2045 socioeconomic projections
- Bloomington Monroe County Metropolitan Planning Organization (BMCMPO) 2045 socioeconomic projections

For the Mid-State TDM TAZs in Indiana, in general 2045 socioeconomic projections (total population in 2045 and growth in employment from 2017 to 2045) were obtained from the INSWM TAZs 2045 projections. For the Mid-State TDM TAZs which are within certain MPO boundaries (e.g., BMCMPO, EMPO, and KIPDA), 2045 INSWM projections were compared with the MPOs' socioeconomic projections and necessary adjustments were made were MPO projections were more reasonable. The MPO projections reflected additional local input not available to INSWM developers.

For Mid-State TDM TAZs in Kentucky, 2045 socioeconomic projections were obtained by using population and employment growth rates from 2010 to 2040 specified in the Kentucky Statewide Model. For the Mid-State TAZs within KIPDA's TDM boundaries, 2045 socioeconomic projections were obtained from the KIPDA's 2045 projections.

For the Mid-State TAZs in Tennessee, 2045 socioeconomic projections in the Tennessee Statewide Model were used.

4 PASSIVELY COLLECTED BIG DATA

Passively collected big data on the movement of persons and vehicles presents a valuable and powerful new resource for travel modeling and forecasting. Passive mobility data includes information from observations of millions of individual trips that can be harnessed for travel modeling and forecasting and simply understanding travel patterns in a region, how people circulate through a city on a daily basis.

Passively collected data for the Mid-States Corridor were provided and expanded by project team member RSG, a multidisciplinary consulting firm and national leader in developing and applying travel demand modeling and analytical techniques to predict travel behavior and transportation systems dynamics. RSG has worked with key mobile device data providers, and since 2019 has informed projects with its own mobility big data platform, rMerge.[4]

Sources and Types of Passive Data

For rMerge, RSG uses Location-Based Services (LBS) data, which can be generally understood as smartphone app data. LBS data is location data provided to applications on a smartphone (or other mobile device) by background programs on the device called location-based services. These LBS programs serve location data to apps based on the requirements of the app and the various information on location available to the device, primarily GPS and Wi-Fi beacons (but also in small amounts Bluetooth beacons and cell tower signaling).

While other passively-collected data sources exist, including cell tower signaling data and pure GPS feeds from in-vehicle navigation systems, several considerations support the choice of LBS data over these other sources for travel analysis:

- **Sample size** – Since roughly 2018 LBS datasets have provided larger datasets than cell tower signaling datasets. These datasets also provide information from more devices, and report on a greater proportion of total travel. The LBS data used by RSG typically contain observations on 10-15 percent of the population on any given day and close to half the population over the course of a month. In contrast, cellular datasets can capture a similar portion of the population but are inferior in locational precision, and quality in-vehicle navigation GPS data is only available from a very small portion of personal vehicles and obviously contains no observations of other modes.
- **Precision** – GPS data provides the greatest consistent level of precision in locations (to within 10 meters). The portion of LBS data that is from mobile device GPS signals is generally of similar precision to that of pure GPS streams from in-vehicle navigation systems and comprises a large portion of rMerge LBS data. A large portion of LBS data also comes from Wi-Fi beacon locations which are typically precise to within 30 to 50 meters (accurate enough for inferring location within a regional model zone system) and a small portion of LBS data comes from Bluetooth beacons, with precision generally similar to Wi-Fi. Under some circumstances, a small portion of LBS data can come from cell tower signaling information. Hence, while only a portion of LBS data provides the level of precision provided by in-vehicle navigation systems, all of it is as precise, and the vast majority of it more precise, than cellular data. All LBS data is filtered for device quality and can contribute to the overall inference of travel patterns within a region. While the consistently high level of location precision from in-vehicle navigation data make it superior for identifying travel speeds on network roadways, the larger breadth of LBS data make it superior for identifying travel patterns.
- **Frequency** – The temporal frequency of observations (sometimes called data density) is another key consideration in passive data as long gaps between observations lead to an incomplete (and systematically biased) picture of travel. Cell tower signaling typically has the worst data density, and GPS typically offers the best. Because LBS data comes from a variety of smartphone applications, each with its own use patterns and location reporting frequencies, LBS data density

[4] rMerge is a complex software and data system engineered to convert raw mobile device sightings into travel behavior datasets and insights. The tool has been in development for over four years and supports dozens of planning and infrastructure projects across the United States.

varies between devices in the dataset. As such, a portion of trips inferred from LBS datasets contain detailed trace observations along the route taken between origin and destination, while others include only observations at the origin and destination, or a few scattered points along the route. In all cases, rMerge trips are anchored at origin and destination with identified clusters of stationary device sightings.

- **Demographics –** Demographic biases are another important consideration in the selection of data sources. In-vehicle navigation GPS data is very heavily biased toward high income individuals who can afford to buy late model cars with navigation option packages. LBS data is substantially less demographically biased than in-vehicle navigation GPS data, since the vast majority of Americans now own smartphones.[5] [6]

In summary, as of this time, LBS data provides the best combination of desirable characteristics for travel analysis. It provides the largest sample sizes of any data source available, for a more diverse population than in-vehicle navigation data and with higher locational accuracy and data density than cell tower signaling data.

While presently LBS data provides the best combination of desirable characteristics for travel analysis, it is still important to consider its inherent limitations and issues. The focus of this document is largely to record and explain the steps taken to address these issues and limitations in order to provide the highest quality mobility data.

Advantages and Limitations

Passive data offers three key advantages compared to traditional data sources such as travel surveys:

- **Scale** – Passive data provides information on the movements of significant portions of the population. It may be possible to observe 10 – 15 percent of the population on a given day and up to half the population over the course of a month. Information is available on roughly 30 percent of heavy trucks on the road. This level of sampling supports types of analyses that simply are not possible with the sampling of traditional surveys (1 percent of the population or less). Passive mobility data provides estimates both of the distribution of trip lengths and the actual distribution of trips between locations.
- **Continuity** – Most passive data are collected continuously all day, every day. There are challenges and real costs associated with processing this level of information. But it now is possible to capture an entire month of data cost-effectively. It is possible to analyze data for multiple months or for an entire year or more. To ensure valid trend analysis or comparisons over time, it is critical to properly differentiate between changes in actual travel and changes in the data itself, such as to the percentage of people with devices providing data. However, the ability to monitor how travel changes over time, especially in response to stimuli such as new technologies or the construction of new infrastructure, is especially valuable in this era of rapid change.

[5] The Pew Research Center indicates 85% of Americans own a smartphone as of February, 2021. (https://www.pewresearch.org/internet/fact-sheet/mobile/)

[6] In 2016 Lochmueller Group and ETC Institute performed a systemwide origin-destination survey for the IndyGo, the Indianapolis, Indiana Transit System. It found that 77 percent of respondents owned smartphones. Over 57 percent of respondents had household incomes under $25,000.

- **Cost** – Since the only interaction required by people or companies observed passively is the brief and easy electronic submission of consent, large samples of data can be collected passively at relatively low cost compared to traditional methods of data collection. Traditional methods require either more extensive interaction with the subjects of study or the deployment, monitoring and maintenance of equipment. The cost of traditional data collection is largely a function of the amount of data collected. In contrast, the cost of passive data is largely a function of its quality; the largest portion of the cost is not data collection but rather its cleaning, processing and expansion.

However, there are limitations. The three key limitations of passive location or mobility data are:

- **Limited Scope** – All passive data available to date are basically “trace data”, which is simply a series of locations and timestamps associated with devices or vehicles. Sometimes there is ancillary information such as speed, heading, acceleration, device type or precision of the location estimate. There is no information on the person who operates the vehicle; to whom the device belongs; trip purpose or type of activity; who may be traveling with them or (for data from mobile devices as opposed to vehicles) the mode of travel. Data science allows inferring or imputing additional information (through the use of additional sources of data such as land use). However, there is always some error in these inferences. It is important to understand the methods by which these inferences are made.
- **Representativeness** – All existing commercially available passively collected datasets are based on incomplete sample frames. These datasets include only a select, non-random portion of vehicles or travelers with mobile devices. They exclude travelers without mobile devices or vehicles without navigation services. Moreover, short-distance trips or short-duration activities are typically under-represented in the data because capturing such trips requires more frequent observations of position. Travel to and from locations with poor coverage can also go un- or under-detected. Failure to account for such biases can produce erroneous representations and faulty predictions of trip lengths, trip flows between origins and destinations, and travel activity and traffic in general.
- **Privacy** – There is widespread agreement on the need to protect the privacy of individuals. However, protecting privacy inevitably involves some loss of information. There are differing approaches and perspectives on how best to protect privacy. All such protection requires some tradeoff between the loss of information and the level of certainty that privacy has been protected. Ideally, travel patterns cannot be attributed to specific individuals.

Passive mobility data complements traditional count and survey data. It provides information that surveys cannot (or can only with great cost). However, passive data will never fully replace data from surveys and counts. In the case of surveys, this is because passive data is—by its nature and the necessity of privacy protection—anonymous. It does not provide travelers’ characteristics and purposes, or the mode used. These are important for many types of forecasting (such as mode choice). There remains a significant need to understand how different kinds of people travel, how they travel differently for different purposes, how travel serves different kinds of activities or what mode of travel is used. Travel surveys are needed to provide these data. Such information cannot be observed passively, nor can there be confidence it is being correctly inferred without validation by survey data.

Passive mobility data can provide information on trucks and visitors, both of which are costly to collect in a survey. Passive mobility data can also collect much larger samples, which are important for less frequent phenomenon like longer distance trips, and for providing a detailed understanding of the spatial (OD) patterns of daily resident trips. While surveys capture many important details of daily resident trips (particularly regarding purpose and mode), no cost-constrained survey can itself provide an OD trip matrix at the level of zones or even moderately disaggregate districts. Traditional surveys typically contain observations for two percent or less of the cells in the OD matrix. In contrast, passive OD data typically provide observations a full order of magnitude greater. These data inform alternative data-driven model frameworks which can produce more accurate spatial results. They provide better understanding of spatial travel patterns. Moreover, since the prediction of mode and willingness-to-pay tolls depends critically upon where people are traveling to and from, passive data supports more accuracy in these dimensions as well.

4.1 Data Processing

With the ubiquity of connected devices and the location-based data they generate, actual origin-destination (OD) travel patterns can be observed for a significant portion of the general population and aggregated into OD travel matrices.

RSG's LBS data processor creates study-area specific OD trip lists by extracting raw sightings from a national dataset with over one trillion sightings per year (raw data obtained from the application data aggregation firm Veraset™), filtering for quality devices and processing to identify trips. In addition to trip origins and destinations, information on trip types, including whether either trip end is a "home", "work", or "other" location, is also inferred. **Figure 4-1** presents the overall workflow for generating observed trip tables.

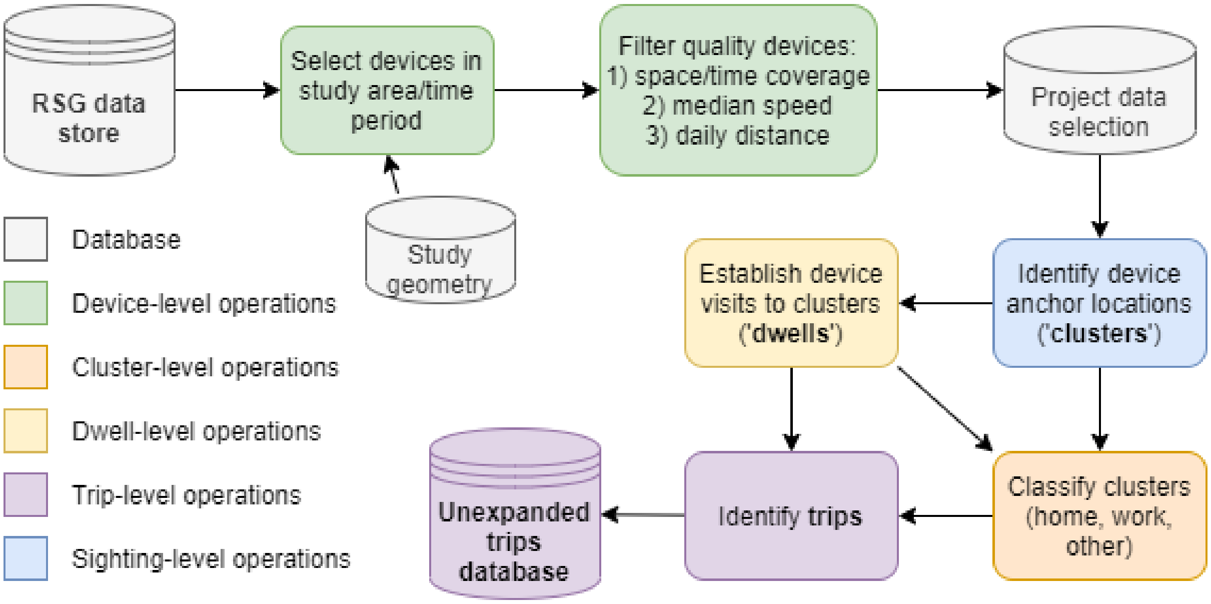


Figure 4-1: Passive Data Workflow

Additionally, heavy truck GPS data was acquired from the American Transportation Research Institute (ATRI), which is a not-for-profit research organization focused on the trucking industry. ATRI GPS data was processed similarly to the rMerge LBS data, resulting in an expansion process involving multiple vehicle classes. Given the accuracy of GPS data and the modeling application need of heavy truck trips, the processing of heavy truck trips is more straightforward than the passive passenger data processes outlined in this section.

4.1.1 Filtering

National LBS datasets include device sightings for well over a hundred million devices. However, not all devices are relevant to a given study area and many of these devices are observed only sporadically. Inferring trip patterns from sporadic data can lead to false inferences and generally poor-quality OD matrices. Therefore, RSG applies quality filters and uses only a subset of the total LBS data initially available. This provides a consistent high-quality data stream to reduce the chance of false trip inferences and provide high quality OD matrices.

In creating study area OD matrices, RSG first filters out low-quality devices. Based on extensive exploratory analysis used to understand noise in the data, filters to remove low-quality data consider factors like unreasonable speeds, long average gaps between sightings, geographic diversity in sightings, and frequency and temporal range of device sightings.

RSG LBS data from the month of October 2018 was processed for the Midstates model boundary, which includes portions of Indiana, Kentucky, and Tennessee, to generate observed trip lists.

4.1.2 Processing

Trips are identified from the raw coordinate-timestamp data using a two-step process. In the first step, individual sightings are classified as stopped or in motion based on the speed computed over a rolling time window. In the second step, a spatial clustering algorithm is applied on all stopped sightings to identify locations where devices have stopped (clusters)[7]. A smoothing algorithm classifies device movement status to filter out stops at traffic signals or stops due to congestion. Device home and workplace locations are inferred from clusters using several indicators, including overnighting and frequency of cluster visitation.

Table 4-1 summarizes processed data for the Mid-States study area. Within the month of October, 2018, a total of 1,602,896 total devices were seen. Of these, 294,323 devices (approximately 7.0 percent of the overall Mid-States 2015 population) were identified to have a home location within the model region. An additional 290,295 devices were identified to make trips within the Mid-States region but have a home location outside. The remaining 1,018,278 devices seen in the month were deemed to have insufficient data with which to accurately identify travel behavior, which aligns with similar LBS data extractions.

Table 4-1: LBS Data Summary

Items	Number
Resident Devices	294,323
Visitor Devices	290,295
Dropped Devices	1,018,278
Clusters	4,160,428

[7] To be identified as a cluster, a location requires “stopped” sightings in at least three unique five-minute time bins. Each cluster has an associated frequency metric, indicating the relative number of stops recorded.

Trips	14,438,435

A time-ordered diary of device dwells (i.e., specific times when a device is stationary at a stop cluster) is then assembled, and trips are constructed connecting all dwells in the diary. Trip ends are then tagged to study area custom geographies (TAZ polygons). Finally, long-distance trips are identified and intermediate stops (e.g., a quick stop at a service station on a longer trip) are identified and flagged to allow summaries of long-distance trips omitting them.

Figure 4-2 and **4-3** present national and regional views of study home, work, and other cluster locations.

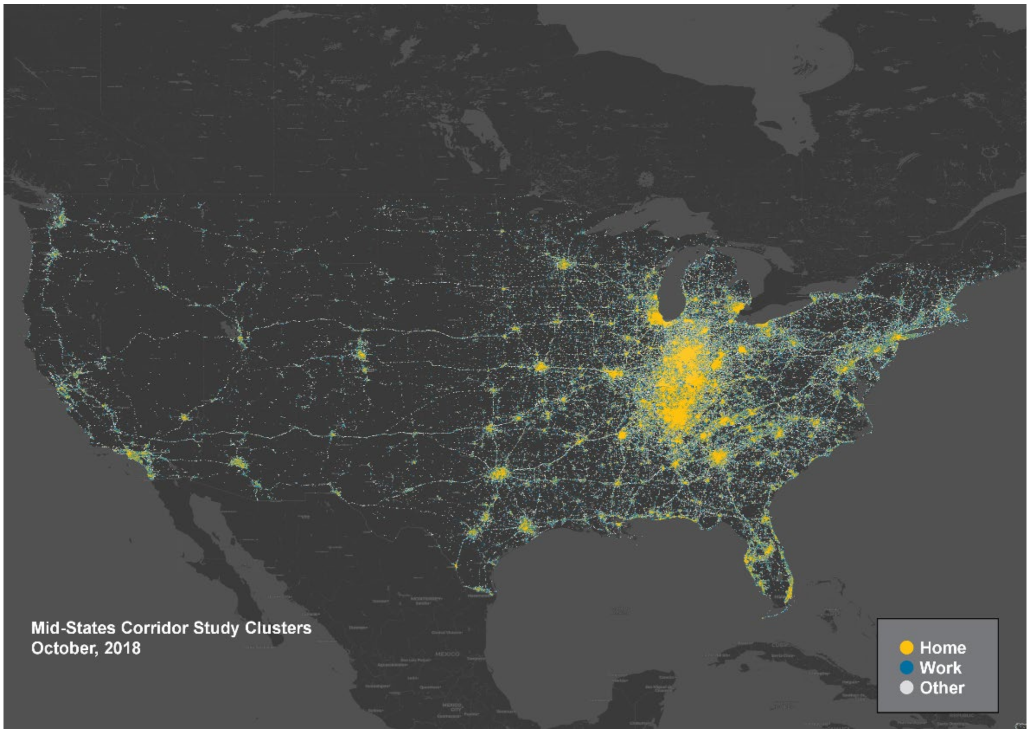


Figure 4-2: Cluster Locations - National Scale

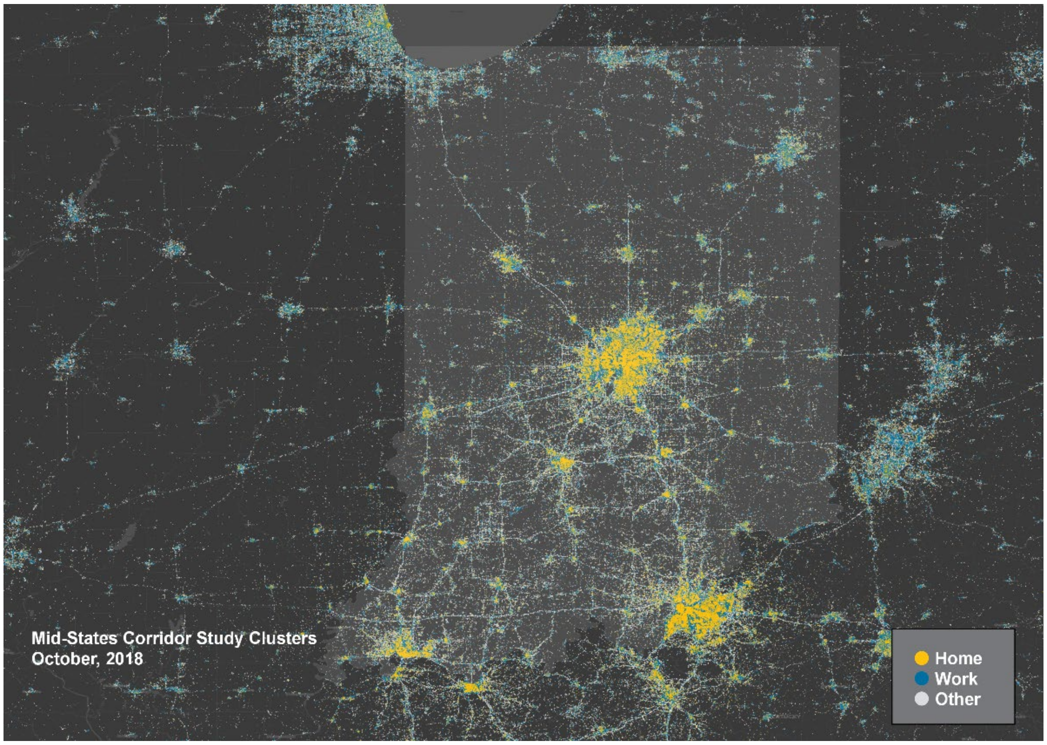


Figure 4-3: Cluster Locations - Model Scale

4.2 Data Expansion

While big data provides large amounts of data on millions of travelers and trips, it remains only a sample of all travelers and all trips. To produce estimates of total travel for the entire population, it is necessary to expand the data to represent all travel. The challenge for big data is that it is not a controlled random sample but rather a "convenience" sample. It is not necessarily a representative sample.

Big data is known to be a systematically biased sample in several ways, and failure to account for these biases can lead to erroneous representations and faulty predictions of trip lengths, trip flows between origins and destinations, present/future travel activity and traffic in general. For these reasons, it is critically important that expansion methods correct for biases and ensure the final expanded data product is representative of all travel.

4.2.1 Types of Bias

There are three main known types of bias present in big data. Each of these presents its own challenges and requires its own methods to address.

- **Demographic Bias** - All existing commercially available passively collected mobility data are based on incomplete sample frames. LBS datasets include only a select, non-random portion of travelers with mobile devices. They exclude travelers without mobile devices. Moreover, since LBS data is derived from data-supporting, location-aware apps, individuals with more apps are more likely to provide data. Given smartphone ownership and app usage trends, big data tends to under-represent seniors and low-income populations and over-represent young adults and affluent populations. These biases are decreasing over time and there are now significant numbers of seniors and low-income travelers in the data. These data can be expanded more readily to represent these groups. These under-represented groups must be expanded more, and over-represented groups expanded less to achieve a representative data expansion.

- **Duration Bias** - Short-distance trips or short-duration activities are often under-represented because capturing them requires more frequent location observations. This is a fundamental issue in all datasets where there is variability in the frequency of observations. Almost all sources of big data have this characteristic, albeit to varying degrees and for sometimes different underlying reasons. For LBS data, variations in the frequency of observations arise due to differing app needs for location data, different behavior of various LBS to different conditions (such as low battery charge or lack of Wi-Fi or 4G data service) and differences in user settings and permissions. Regardless of the underlying reason, when there is varying frequency in observations, the sampling probability of an event (e.g., a trip or an activity) is a function of its duration. The shorter event can more easily "slip through the cracks" between observations.
- **Geographic Bias** - Travel to and from locations with poor data coverage can also go un- or under-detected. Different devices and different LBS respond differently to differing levels of data service. In very rare circumstances, where there is no data service and no GPS line-of-sight (e.g., under a cliff or overhang in a remote area), there can be actual holes with no data. Typically, the problem is more subtle. There may be some areas where there is poor or no data service, but sampling rates in these areas are lower than in areas with good data service. Sightings in areas of poor data service often have to be expanded more to ensure representative patterns in the final data.

These three recognized forms of systematic bias are the main focus of our expansion processes. There may also be other, still as yet unrecognized forms of bias in passive data. For this reason, our expansion processes are also designed with some flexibility to capture and correct for subtle biases that may be measurable against control data, even when the underlying mechanism is not clearly understood.

4.2.2 Types of Control Data

Comparing two or more datasets allows for identification and correction of biases in a dataset. A dataset may be expanded by using one or more additional "control" data sets to measure and correct for biases in the data of interest. For instance, travel diary survey data is expanded or weighted primarily using Census Bureau datasets as control data. On-board transit surveys are expanded using estimates of route level ridership, such as on-off counts, as control data.

Three types of control data are available for expansion of passive mobility data:

- **Demographic and employment data** from the Census Bureau (and sometimes other private sources for employment data)
- **Travel counts**, primarily roadway traffic counts, although bicycle and pedestrian counts, transit ridership estimates, site visitation counts, ticket sales, etc., sometimes are used
- **Disaggregate trace data from smartphone travel surveys**, such as rMove. For these surveys, respondents provide details on trip characteristics and demographics to smartphone-based travel survey questions, while the survey app simultaneously collects GPS trace data for all trips.

RSG can and has used all of these types of control data to assess the representativeness of and expand passive datasets. The choice of which control dataset(s) are used for the expansion of a particular dataset depends on the type of data to be expanded and the control datasets available. For instance, a dataset for (1) understanding external travel to a region or a dataset for understanding the origins and destinations served by a particular roadway may be expanded differently than (2) a general dataset for understanding travel patterns in a region.

For a type (1) dataset, very detailed count data may be available for the entire study area, whether a roadway corridor or a region's external cordon. For a type (2) dataset, some count data may have a high sample rate (including counts by vehicle class and time of day). Other count data may have only a very sparse sample of total daily traffic counts. Similarly, some regions have recent travel survey data with smartphone GPS traces. Others have surveys but without trace data. Some other regions may have no recent survey data at all.

4.2.3 Ensemble Expansion

RSG customizes the expansion datasets, leveraging its expertise in travel data and behavior to make best use of all available data for each client. This includes using prudent judgment to address idiosyncrasies of both passive mobility data and local control data.

RSG's standard practice is to use an ensemble of expansion methods to expand passive data. In mathematical terms, final expansion factors are generally developed as a product of several component expansion factors. This recognizes that no single method that can address all three of the biases cited above. Within this flexible framework, methods may be added if some techniques prove inadequate. Alternatively, some techniques may be dropped if other methods suffice.

Consistent with general practice across the industry, RSG uses residence-based sample penetration to correct for demographic biases using census demographic data as the control. When smartphone survey data are available, RSG can use this as control data for correcting duration bias. Whenever counts are available, RSG uses them to address geographic biases, as well as duration bias when local smartphone survey data is not available.

4.3 Data Expansion

Given the relatively small amount of local smartphone survey data and traffic counts for the Mid-States Study Area, RSG used an ensemble of census and count-based methods to expand the Mid-States passive data. **Figure 4-4** illustrates the procedures used in expanding passive data for this project.

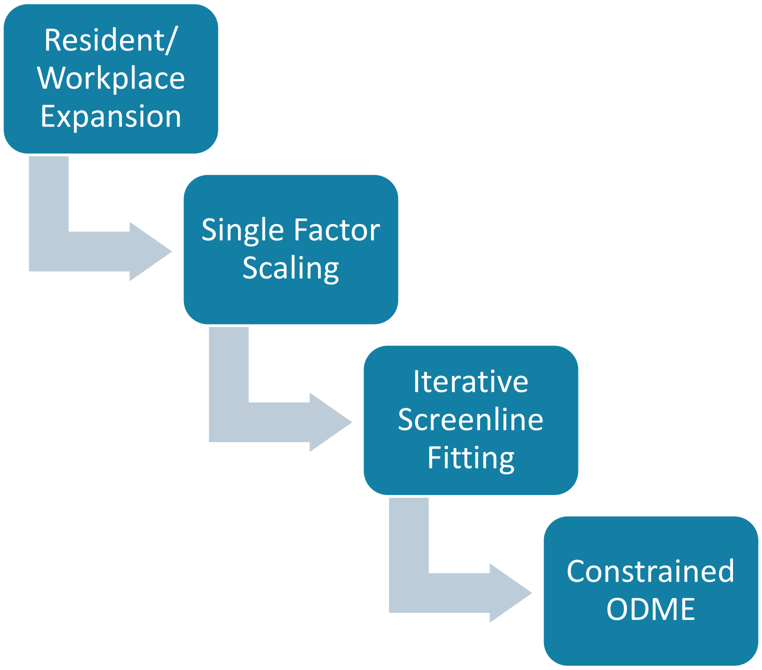


Figure 4-4: Applied Expansion Process

4.3.1 Residence Sample Penetration Expansion Methods

Demographic data from the Census Bureau or from travel-demand model land-use inputs can be used to measure sample rates and calculate expansion factors by residence zone. The number of observed devices residing in a zone or block group can be compared with the Census Bureau or travel demand model's estimate of the number of people residing in that same unit of geography. An expansion factor is simply the ratio of these two numbers. For Mid-States, US Census data, joined to model zones by Lochmueller Group, were used to calculate the demographic expansion factors by zone illustrated in **Figure 4-5**

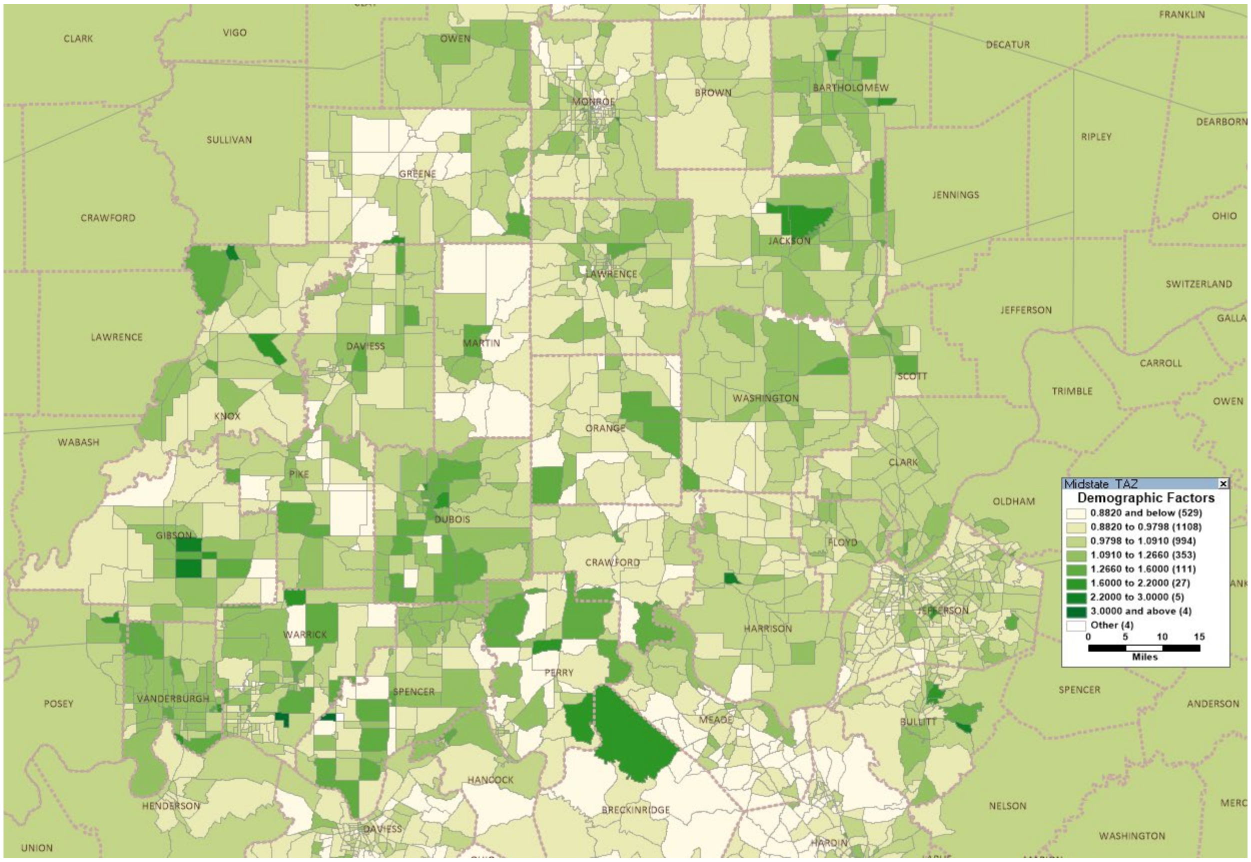


Figure 4-5: Normalized Demographic Expansion Factors

4.3.2 Single-Factor Scaling to Counts

Single-Factor scaling uses a single expansion factor based on a comparison of passive data to traffic counts at one or more locations. Traffic counts can be compared to LBS data by assigning the trips to roadway facilities using a network assignment model. While detailed scaling can be done through iterative examination of many individual count locations, single-factor scaling provides a single initial high-level adjustment, either based on the overall network-wide loading error (examining all counts together) or through comparison with an estimate of regional vehicle miles traveled (VMT) based on counts such as those found in FHWA's Highway Performance Monitoring System (HPMS).

The simplicity of this method makes it both easy to apply and easy to explain to nontechnical audiences. However, since only a single factor is used, it cannot correct for many issues including coverage variation within a region or trip-length bias. Given both its simplicity and limitations, it is commonly used in combination with other methods. For Mid-States, a single-factor scaling adjustment based on overall network loading error (ratio of the sum of all counts and the sum of all corresponding LBS routed flows) was applied following demographic expansion and prior to more detailed count-based expansions.

4.3.3 Iterative Screenline Fitting (Matrix Partitioning)

Iterative screenline fitting (ISF) or matrix partitioning is a special type of iterative proportional fitting to counts. It differs from the typical Fratar technique in that it relies on counts associated with multiple groupings of zones rather than individual zones. Also, zones may be part of more than one grouping.

The approach first identifies screenlines and/or cutlines (similar to those commonly used to validate travel models). A screenline partitions the study region into two subareas and aligns with the zone system used to define ODs. Traffic counts should be available or taken everywhere the roadway network crosses the screenline. (It is helpful to choose screenlines which follow natural/physical barriers such as rivers, freeways, and railroads which have limited roadway crossings.) The definition of cutlines is less restrictive, but the closer they are to having the characteristics of a true screenline the less the method relies on network pathfinding. **Figure 4-6** shows screenlines and cutlines used for the Mid-States LBS expansion.

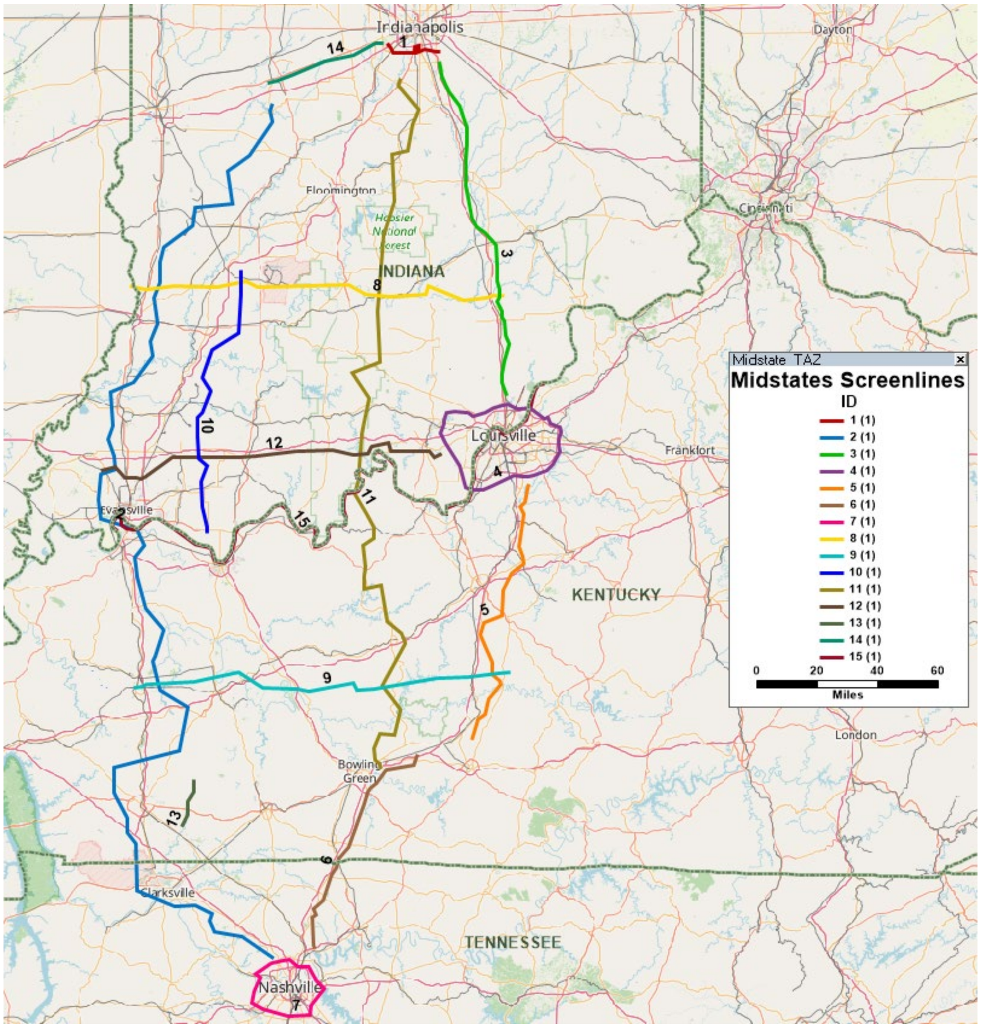


Figure 4-6: Screenlines for Passive Data Expansion

The sum of the traffic counts along each screenline and cutline is compared to the number of trips in the OD matrix which cross the screenline or cutline. For true screenlines, this comparison can be made without a network assignment model by partitioning or aggregating the OD matrix. Since each screenline partitions the region into two subareas, this partitions the OD matrix into four regions, two off-diagonal regions with trips between the subareas and two diagonal regions with trips within each subarea. Groups of OD trips can be compared against screenline counts without a network assignment model. A preliminary expansion factor is developed as the sum of the screenline counts divided by the sum of the off-diagonal regions of the matrix. When cutlines are used, the partitioning of the matrix is more complex and relies on a network assignment to identify groups of OD trips for cutline count comparisons. Factoring based on screenlines and cutlines results in OD trips matching the sum of counts for the current screenline or cutline. However, when trips between individual OD pairs cross multiple screenlines/cutlines, there is the potential for different expansion factors to be calculated for a single OD pair. For this reason, expansion factor calculations are iterated until expansion factors for individual OD pairs stabilize to values that minimize errors versus among all screenline/cutline counts. There may not be perfect agreement of the OD data with any individual screenline/cutline count.

RSG has successfully used this procedure to improve the expansion of several types of passive data in a number of regions around the country. The benefits of this approach increase with the number of screenlines which can be constructed. The construction of multiple screenlines and/or cutlines can be relatively easy or difficult, depending upon the amount and coverage of count data. If there is poor count coverage, it may not be possible to construct any complete screenlines or a sufficient number of cutlines for the method to be of much value. In regions with good count coverage and physical barriers to travel that create natural screenlines, the method can be of considerable value and significantly reduce the need to use assignment-based methods.

Fifteen screenlines/cutlines (shown in **Figure 4-6**) were used for the Mid-States LBS expansion, including a valuable screenline using the Ohio River to horizontally bisect the study region and provide an early control total for overall north/south travel. This screenline/cutline fitting before application of assignment-based methods reduces the likelihood of introducing expansion errors from network assignment models.

4.3.4 Constrained ODME

Origin-Destination Matrix Estimation (ODME) refers to a process whereby traffic data is used as an input to estimate the traffic volumes between each origin and destination in the form of a resultant O-D matrix. ODME is one of the most common approaches to expanding passive data (see for instance Zanjani et al., 2015; Han et al., 2016). A proper understanding of ODME is grounded in two important facts. First, counts do provide real information about underlying OD patterns, and second, counts alone cannot be used to identify OD patterns. The former is demonstrated in iterative screenline fitting. The latter is evident from the fact that the number of “known” traffic counts is always substantially smaller than the number of “unknown” OD flows. Statistically, this provides and under-determined problem. There is not a single, unique set of OD flows that correspond to a set of traffic counts on a network.

There are a variety of ODME algorithms which can produce significantly different results. ODME methods which use OD data only as a “seed” or starting point and produce an adjusted OD matrix purely by minimizing errors versus traffic counts can significantly distort the data, if unconstrained. Other methods are especially powerful as appropriate methods for data expansion. Some methods find a solution which minimizes errors versus counts and versus the original ODs. Other methods minimize error versus counts with appropriate constraints on adjustments to the original OD data. These methods are capable of correcting systematic biases related to trip lengths as well as coverage "holes" (provided there are at least some observations in the "holes" to expand). They also can avoid other errors which may be difficult to hypothesize a priori while ensuring that the data are not grossly distorted.

RSG typically uses constrained ODME algorithms that minimizes squared error of assigned volumes versus counts subject to constraints that the final trip values do not vary from the original by more than a given ratio or absolute amount. For the Mid-states project, RSG limited ODME adjustments at the cell level to between +200 percent and -50 percent; moreover, the ODME algorithm used does not allow for the introduction of trips in cells with no trips observed. Finally, RSG also performed data validation checks aimed at ensuring that ODME has not overfit to counts.

Using ODME in combination with and secondary to other expansion methods including demographic expansion, single-factor expansion and ISF allows imposition of tighter constraints on the ODME adjustments and greater confidence in the expansion while also allowing a tighter fit to traffic counts. This is RSG's standard practice and is the method followed for Mid-States.

Final results of the Mid-states data expansion, concluding with ODME, are presented in **Table 4-2**. The following set of statistics are presented for this post-expansion comparison of count data with routed OD flows:

- Observations: the total number of counts
- Avg. Count: average count volume
- Avg. Data: average routed OD flow through count location links
- t Statistic: inferential statistic for determining the significance of observed differences
- Error: overall percent loading error
- RMSE: root mean squared error
- MAPE: mean absolute percentage error
- r: correlation coefficient

Figure 4-7 and **Figure 4-8** present graphical plots comparing the assigned expanded OD flows and target count volumes. Note that the statistics presented in this table are not validations statistics which assess the ability of traffic assignments to replicate ground counts. Those statistics are provided in **Section 6** of this document.

Table 4-2: Detailed Expansion Results

	Observations	Avg. Count	Avg. Data	t Statistic	Error	RMSE	MAPE	r
All	1768	5940	5565	-1.3	-6.32	42.81	56.75	0.96
< 5,000 AADT	1144	1886	1937	0.8	2.70	61.61	77.21	0.71
5,000 to 10,000 AADT	331	7200	6459	-5.1	-10.29	28.97	19.00	0.53
10,000 to 20,000 AADT	180	13423	12721	-1.6	-5.23	33.90	22.51	0.46
20,000 to 30,000 AADT	73	23197	20627	-4.4	-11.08	17.39	13.63	0.63
> 30,000 AADT	40	46291	42225	-0.7	-8.78	20.30	16.58	0.94
Interstates	119	22443	22106	-0.2	-1.50	28.59	22.21	0.84
Expressways	76	7298	5722	-2.3	-21.60	48.05	20.52	0.81
Principal Arterials	256	10459	9368	-1.7	-10.43	24.54	16.61	0.95
Minor Arterials	276	6010	5570	-1.2	-7.33	38.01	47.80	0.87
Major Collectors	756	2028	1997	-0.3	-1.53	47.51	83.36	0.89
Minor Collectors	15	3250	3827	0.4	17.77	43.17	113.12	0.95
Local Roads	6	1064	1364	0.4	28.17	136.50	178.30	0.11
Ramps	224	2692	2416	-1.1	-10.26	64.28	50.08	0.80

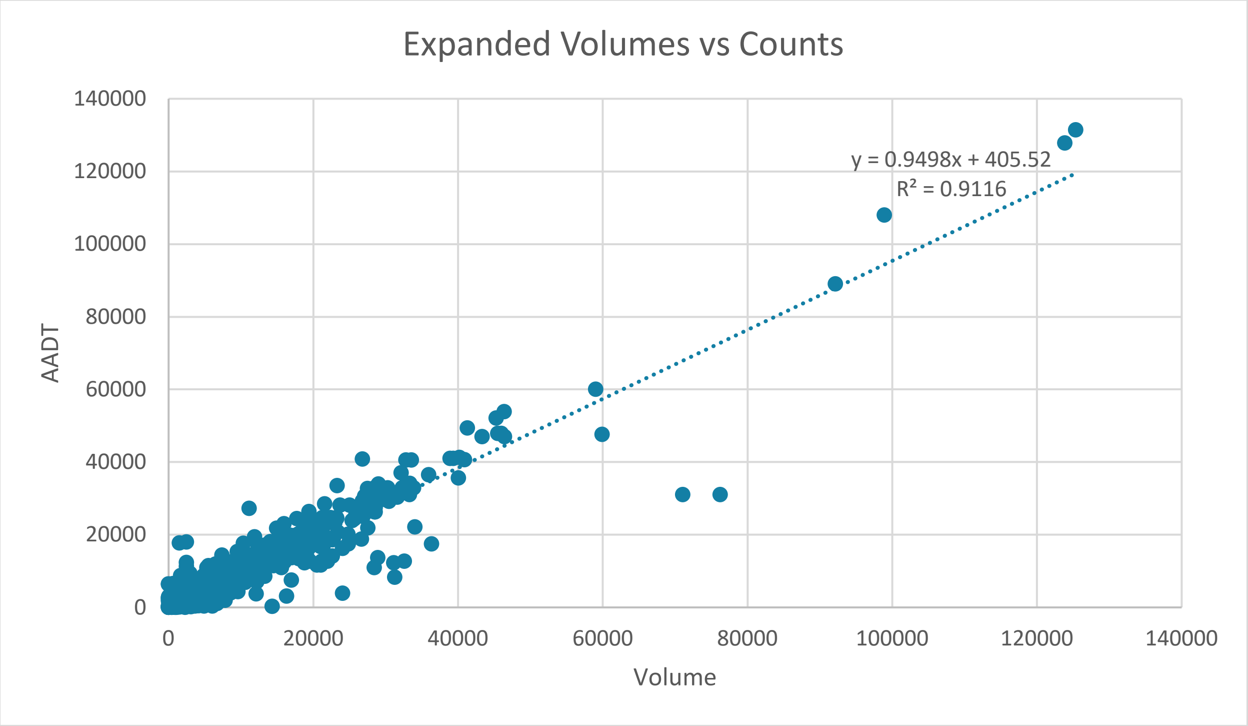


Figure 4-7: Expanded Flows vs. Counts

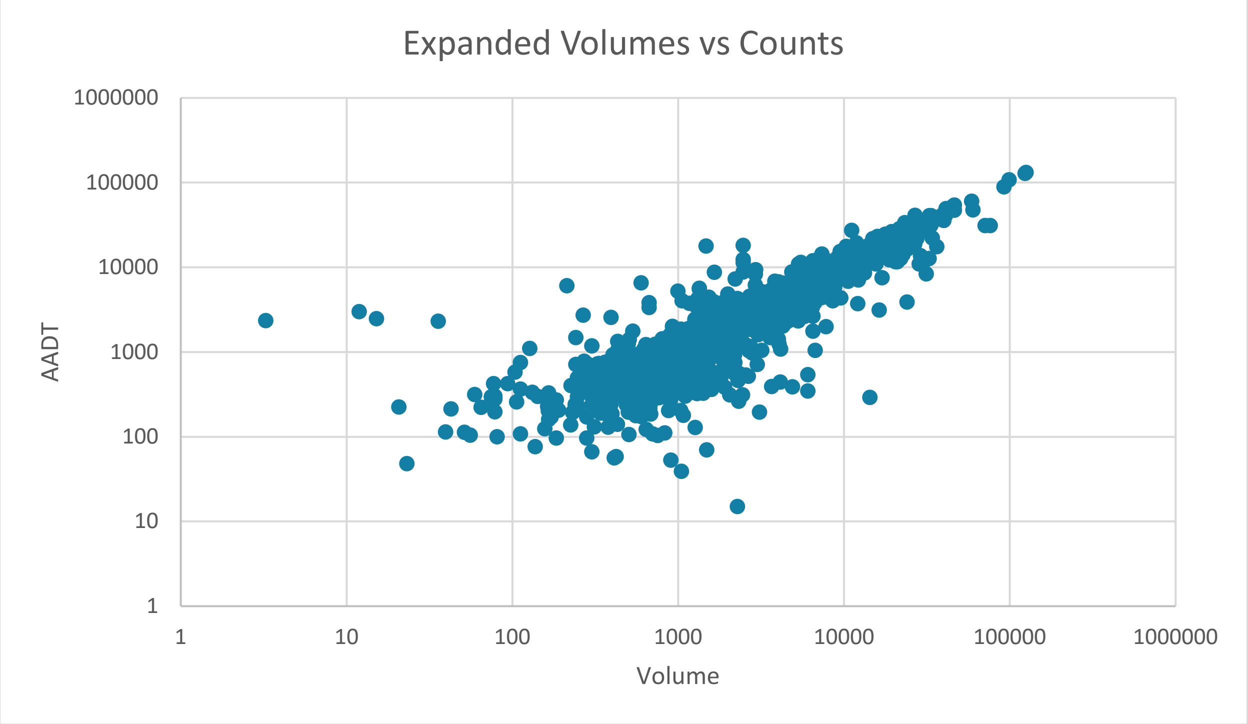


Figure 4-8: Expanded Flows vs. Counts (Log-Log)

4.4 Data Validation

Two rounds of data validation were performed, at the pre-expansion and post-expansion stages. A first round of pre-expansion validation checks was performed before any data expansion. These checks identify anomalies in particular datasets and potential errors in the data processing steps. This also provides baseline information on biases to guide expansion efforts to reduce bias. A second round of post-expansion validation checks assessed that the expansion removed as much bias as possible.

These checks ensure the integrity of the raw data and its processing, as well as the reasonableness and effectiveness of the expansion process. These checks demonstrate that the raw data provides valuable data on travel behavior and patterns; that the expansion corrects for known and observed biases; and that the resulting final dataset delivered is suitable for modeling and other analysis purposes.

4.4.1 Pre-Expansion Data Validation

To ensure the quality of the processed OD data output, RSG reviewed summary reports and performed a series of reasonability checks. These reports review both the underlying data and the identified trips and inferred attributes.

Input data are reviewed to ensure sufficient numbers of devices are consistently seen throughout the study month and that inferred trip making is consistent across the study month. Post processed data are examined to ensure the magnitude of inferred cluster types are reasonable, and that temporal patterns in trip making are reasonable for all trip type combinations (i.e., home to work trips versus work to home trips). Processed results are also mapped to ensure trip end densities align spatially with known land-use densities.

Figure 4-9 presents a summary of Mid-States LBS data trip counts for the seven origin/destination type combinations. **Figure 4-10** and **Figure 4-11** present the average weekday trip counts by time of day for these same trip types for residents and visitors, respectively. These plots show that inferred LBS trips are balanced between trip types (home-to-work vs work-to-home) and occur at expected times of day.

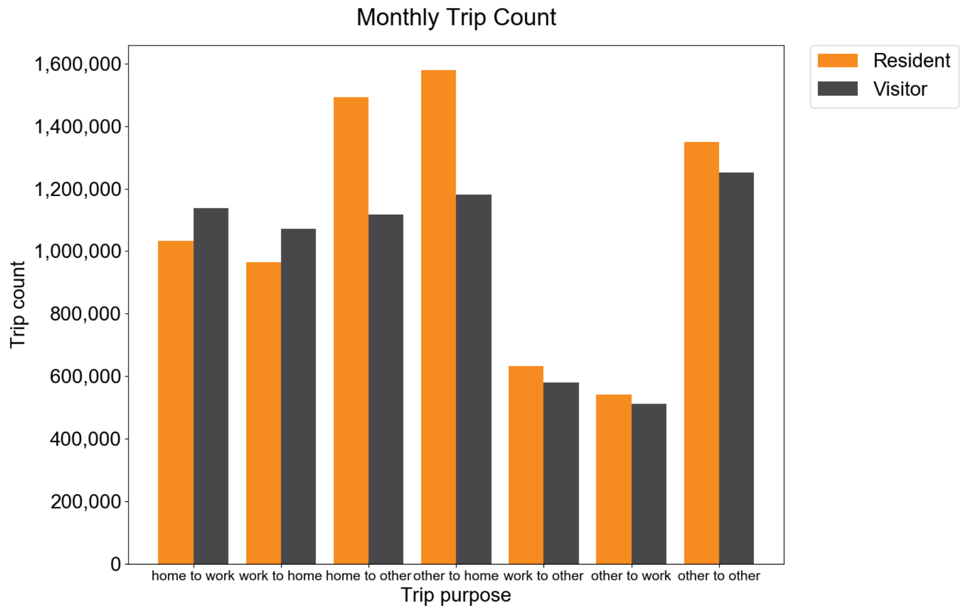


Figure 4-9: Passive Data Trip Summary

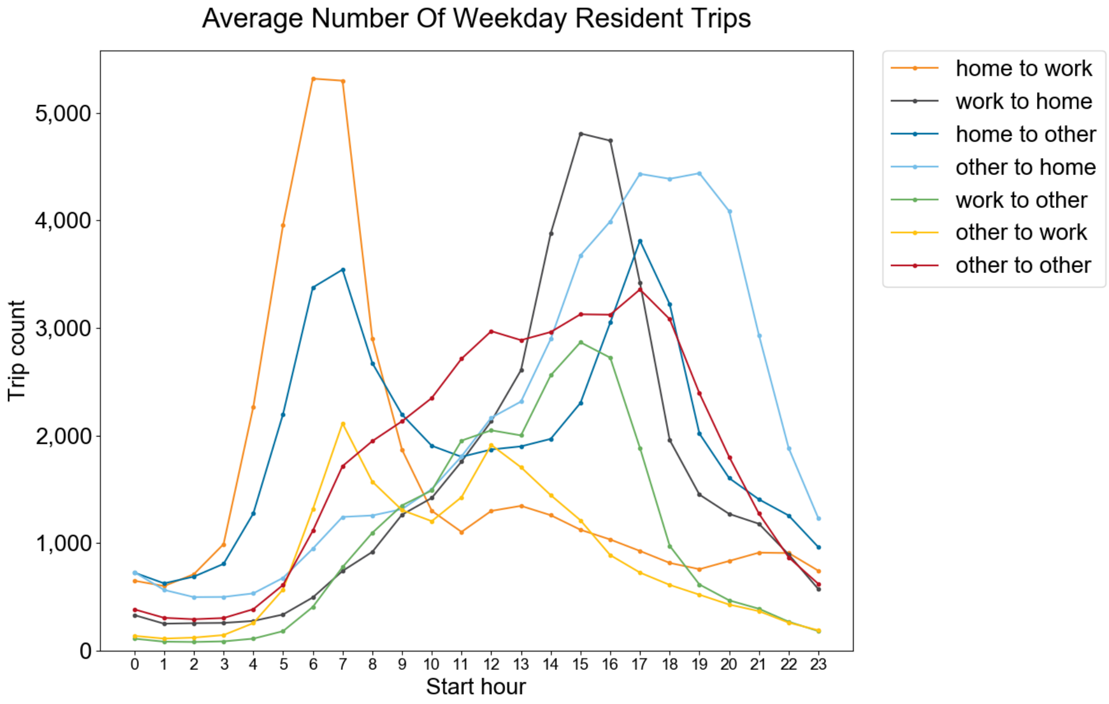


Figure 4-10: Average Weekday Hourly Trip Distribution - Residents

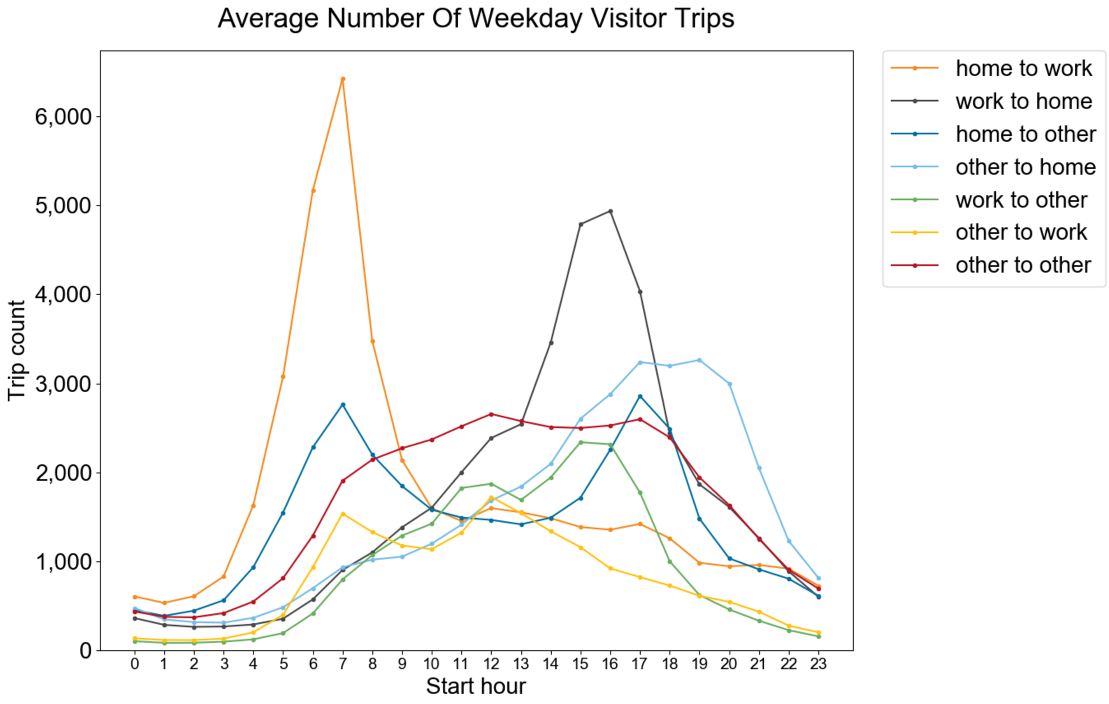


Figure 4-11: Average Weekday Hourly Trip Distribution - Visitors

4.4.2 Post-Expansion Data Validation

There are validation checks and expansion revisions throughout the expansion process. It remains critical to review the final results and factors applied at the completion of this process. RSG reviewed two main criteria comparing the expanded passive data to the initial pre-expansion data: an aggregate zonal comparison and a cell-level matrix comparison.

At the overall zonal level, the average expansion factor applied was 2.37. Given the wide ranges of the pre-expansion and post-expansion data, the expansion factors were normalized as part of the data validation. 96.1 percent of zones fell within a normalized factor of 0.0 and 2.0. Zones with expansion factors beyond 4.0, representing less than 0.2 percent of zones, were reviewed and deemed acceptable due to several factors. **Figure 4-12** presents the distribution of zones across the range of observed normalized expansion factors.

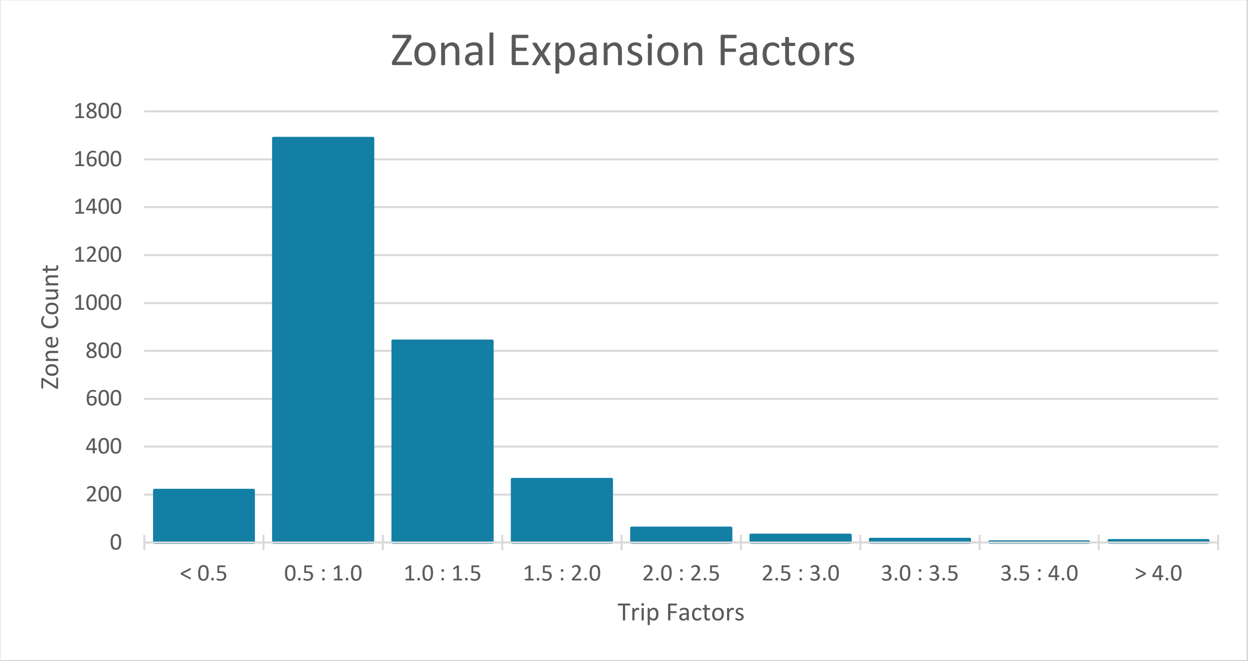


Figure 4-12: Zonal Expansion Factor Histogram

Examination of cell-level expansion factors for individual OD pairs confirms the expansion results in reasonable changes. **Figure 4-13** shows the similarity in the OD patterns and the bounds enforced on the expanded trips. The correlation coefficient from this comparison was 0.737 which demonstrates that the expansion process did not distort valuable information from the passive LBS data.

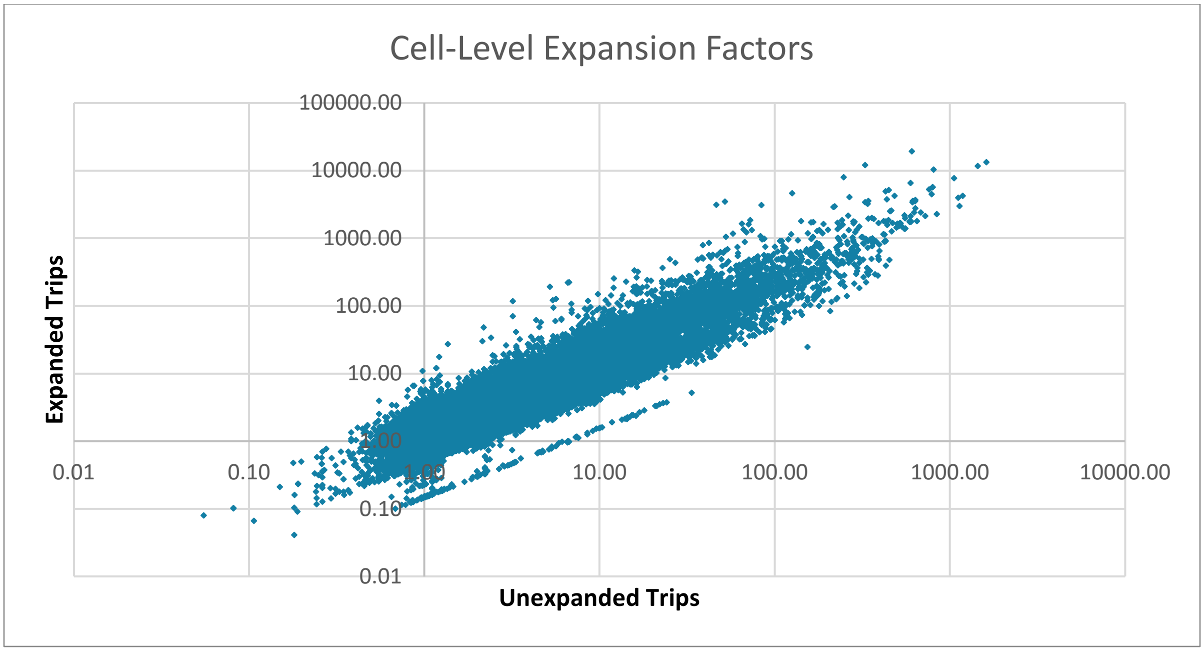


Figure 4-13: Cell-Level Expansion Factors

5 THREE-STEP TRAVEL DEMAND MODEL DEVELOPMENT

The Mid-States Corridor model development used an ensemble forecasting approach which combines traditional trip generation and distribution methods with machine learning and trend-based methods which use longitudinal passive data.

Ensemble modeling has been one of the most powerful techniques to emerge from machine learning and data science to date. The approach is increasingly relied upon across a wide range of industrial applications from meteorology to credit scoring to market research and targeted advertising. It is based on the fact that all models have error, but different models have different errors. Multiple models together can make more accurate forecasts, to the extent their errors offset. This approach also provides cross-validation of component models since differences in their forecasts should be plausible and logically related to the differences in assumptions and methodologies. The ensemble approach generally uses multiple simple models or one complex and several simple models. The alternative approach of activity-based modeling attempts to address model limitations with added complexity.

A data-driven three-step model, leveraging passive data OD matrices, was developed for this study. Trip generation, trip distribution, and assignment steps were performed along with a pivoting process using distribution-generated OD matrices and passive data OD matrices (see **Section 4**), which is applied prior to assignment. These processes are summarized in **Figure 5.1** and detailed in the following subsections.

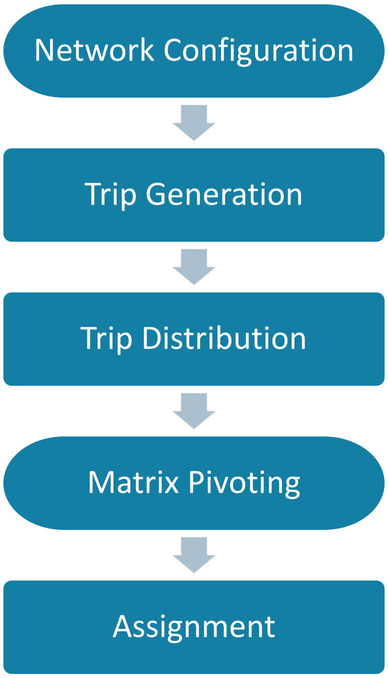


Figure 5-1: Model Flowchart

5.1 Trip Generation

Regression models using zone-level expanded passive data and zone level socio-economic data were developed to estimate the number of home-based trip productions by each zone based on the number of households and their attributes. Trip generation models were developed for the following five trip types:

- Passenger Trips:
  - Short-Distance Internal,
  - Long-Distance Internal,
  - External
- Heavy-Truck (MUT) Trips:
  - Internal
  - External

Passenger trips were segmented by distance and location. Short-distance and long-distance trips were defined as those whose trip lengths fell below or above 50 miles, respectively. Internal zones were identified as zones within the 12-county study area with available household data in the Mid-States TDM. Zones beyond the boundary of the internal zones were identified as external zones.

Table 5-1, found on the following page, presents the trip generation models. This table includes the model specifications, estimated coefficients and calibrated coefficients. A blank calibrated value indicates that the final specification did not change from the estimated coefficients.

Table 5-1: Trip Generation Parameters

Variable Name	Est. Value	Cal. Value	T-Stat
Short Distance Internal (R-squared: 0.8583)			
Food & Lodging Employment	5.932	4.379	9.364
Retail Employment	3.963	4.266	9.371
Service & Office Employment	1.598		19.741
Agriculture & Industrial Employment	0.629	0.489	6.109
Population/Workers Ratio	-0.042		16.561
Population	1.738		-17.380
Long Distance Internal (R-squared: 0.6817)			
Food & Lodging Employment	0.798	0.786	19.363
Retail Employment	0.055	0.108	1.719
Service Employment	0.500	0.544	5.834
Professional Employment	0.047	0.052	7.293
Agriculture & Industrial Employment	0.152	0.155	17.516
Population-Nonworkers Ratio	0.015		1.643
Population	0.016		17.388
Passenger External (R-squared: 0.1873)[8]			
Population-ExtDistance ratio	1463.518		0.589
PopDist * Employment	-0.109	0.00	-0.184
Heavy-Truck Internal (R-squared: 0.6421)			
Agriculture Employment	0.243		19.062
Industrial Employment	0.122		48.691
Food, Lodging, & Retail Employment	0.026		8.839
Heavy-Truck External (R-squared: 0.1723)			
Employment-ExtDistance Ratio	188.696		0.144
EmpDist * Population	-0.023	0.00	0.046

5.2 Trip Distribution

The functional form of the distribution model is shown in **Figure 5-2**. The model uses a doubly constrained gravity model for the three internal trip purposes (short-distance passenger, long-distance passenger, and MUT). This function is applied separately for each of the trip types.

[8] The initial external rates (for both passenger and heavy truck) appear low likely due to relative lack of data for external trip making. External trip forecasting was greatly improved by the pivoting process. See **Section 5.3**.

$$T_{ij} = P_i * \frac{A_j * f(d_{ij})}{\sum_{all\ zones\ z} A_z * f(d_{iz})} \quad \text{(Constrained to Productions)}$$

$$T_{ij} = A_j * \frac{P_i * f(d_{ij})}{\sum_{all\ zones\ z} P_z * f(d_{iz})} \quad \text{(Constrained to Attractions)}$$

Where: T_{ij} = the forecast flow produced by zone *i* and attracted to zone *j*
P_i = the forecast number of trips produced by zone *i*
A_j = the forecast number of trips attracted to zone *j*
d_{ij} = the impedance between zone *i* and zone *j*
$f(d_{ij})$ = the friction factor between zone *i* and zone *j*

Figure 5-2: Gravity Model Functional Form

The inputs to trip distribution include the pivoted outputs from trip generation: row “productions” and column “attractions” by TAZ, as well as a generalized cost impedance matrix. This matrix represents the cost of travel between each pair of TAZs. The impedance is used in the trip distribution model to estimate friction factors, which represent the impact of travel time on the likelihood of travel and are calibrated so that observed trip lengths and times are reasonable and match patterns from the observed passive data.

The friction factor equations take the form shown in **Figure 5-3**, with the estimated parameter values shown in **Table 5-2**.

$$f(d_{ij}) = \frac{a}{d_{ij}^{\ b} * e^{c(d_{ij})}}$$

Where: $f(d_{ij})$ = the friction factor between zone *i* and zone *j*
d_{ij} = the impedance between zone *i* and zone *j*
a,b,c = constants derived for each trip type to replicate survey data

Figure 5-3: Friction Factor Functional Form

Table 5-2: Estimated Friction Factor Parameters

Trip Type	A	B	C
Short Distance	5705.345	-1.275	-0.0571
Long Distance	1.0321	2.542	-0.0316
MUT	12252.37	-1.403	-0.0047

5.3 Pivoting & Traffic Assignment

The Mid-States travel model adopts a data-driven approach to traffic assignment. Future year trip matrices “pivot” off of observed base year trip matrices. Pivoting is based on the data expansion from the LBS passenger data and GPS truck data.

Results from the demand model base and future-year runs, along with the observed passive data demand matrices, are used as inputs to generate future-year pivoted matrices. When a future year demand model is run, the growth from the base year run is calculated and applied to the observed passive data from the base year.

Besides the cases where observed data is lacking, the pivoted estimation is generally defined as:

$$Pivoted = (Future_{model} - Base_{model}) + Base_{observed}$$

The use of this data-driven pivoting approach is increasingly common (e.g., the statewide models of Indiana, Illinois, Tennessee, North Carolina and Michigan) as a way to leverage the availability of big data. This process also allows the model to reproduce observed travel more faithfully because these pivoted forecasts are anchored to current real-world traffic conditions observed through passive data.

Trips from this pivoted matrix are assigned by period using multi-class equilibrium with generalized costs. The assignment uses tri-conjugate Frank-Wolfe algorithm, which is Caliper's Corporation's[9] implementation of Daneva and Lindberg (2003). The assignment includes two vehicle classes defined as Autos and Multi-Unit Trucks (MUT). Although "Autos" are a well-recognized vehicle class in modeling, within the passive data of the Midstates model, this "Auto" class encompasses other vehicle classes beyond passenger automobiles. These include buses, commercial vehicles and light trucks.

Generalized cost is segmented by two vehicle types – auto and multi-unit trucks. Generalized cost variables are used for network skimming and assignment and are determined from a linear combination of actual and weighted or perceived impedance variables. The actual impedance variables include free flow times, congested times and out-of-pocket travel costs. The perceived impedance variables include weighted delay and weighted distance by facility type.

The weighted delay captures the fact that people generally perceive a minute of travel in heavily congested conditions as more stressful than a minute of travel in free flow conditions. See **Table 5-3**. It is also a proxy for reliability. The facility-type distance weights can capture the fact that vehicles, especially trucks, are generally less likely to use minor facilities unless they offer an obvious travel time advantage. See **Table 5-4**.

These values were specified based on work in the Tennessee statewide model, which used data from previous studies and calibration through a genetic algorithm to obtain the values presented in **Table 5-3** and **Table 5-4**. Other inputs required to calculate generalized cost include travel time, posted speed, tolls, functional class and link length.

In addition to the actual measured travel times and costs, the assignment pathfinding includes a distance-based impedance term. The weights were estimated by running a genetic algorithm, which repeatedly runs the traffic assignment and attempts to find weights that reduce the loading error (percent RMSE). Major facility types generally have a lower weight than minor facility types. The facility type differences are most pronounced for trucks, which generally prefer using major facilities that have wider lanes.

[9] Caliper Corporation is the provider of TransCAD, the modeling software used for the Mid-States model.

Table 5-3: Variables Considered for Generalized Cost Estimation

Variables	Car Penalty	MUT Penalty
Value of Time ($)	$12.60	$55.10
Perceived Delay (minutes)	1.25	1.23

Table 5-4: Distance Weight by Vehicle and Facility Type

Facility Type	Cars	MUT
1 – Freeways and internal centroid connectors	0.45	0.86
2 – Expressways	0.6	1.04
3 – Ramps and Major roadways	0.7	1.15
4 – Minor roadways	0.8	1.26
5 – Minor roadways in urban areas	0.9	1.39
6 – All others	1	1.5

Validation of the base year model assignment is documented in **Section 6**.

6 DAILY TRAFFIC ASSIGNMENT AND VALIDATION

Daily Assignment Validation

Traffic assignment validation is a key step in travel demand modeling. The Mid-States TDM traffic assignment validation includes a detailed and robust validation for the 12-County Study Area and overall validation statistics for the whole model area. Traffic assignment validation statistics include daily percent loading error (%Error), Percent Root Mean Square Error (%RMSE), and Mean Absolute Percentage Error (MAPE) calculated based on field collected daily traffic volumes and model assigned daily traffic flows along key roadways.

Typically, multi-state TDMs encompassing vast rural areas such as the Mid-States TDM include a high concentration of low-volume highways, including interstates. The validation statistics for interstates in rural settings are typically higher than in urban TDMs, where interstate traffic volumes are significantly larger. Additionally, most statewide TDMs report their validation statistics by traffic volume range rather than by functional classification to avoid confusion with urban model guidelines.

Table 6-1 shows the assignment validation statistics for all vehicle classes within the 12-County Study Area. The table shows the daily percent loading error (%Error), percent root mean squared error (%RMSE), and mean absolute percentage error (MAPE) by volume group.

Table 6-1: Daily Assignment Validation Statistics for the 12-County Study Area

Roadway Daily Traffic Volume	% Error	%RMSE	MAPE
< 5,000	-4.14	37.1	55.1
5,000 to 10,000	-4.61	12.7	13.05
10,000 to 20,000	-4.25	11.6	8.6
20,000 to 30,000	0.38	5.93	0.31
> 30,000	0	0	0
Total	**-4.4**	**23**	**47.4**

As can be seen in **Table 6-1**, %RMSE values for roadways with more than 5,000 daily traffic volume were less than 15 percent. The 12-County Study Area achieves a percent error of -4.4 percent and an RMSE is 23 percent.

Figure 6-1 shows a scatterplot of the traffic counts versus overall model loading for the 12-County Study Area. The model volumes and traffic count generally show reasonable consistency. The r-squared value is 0.96, which displays high correlation between assigned volumes and counts.

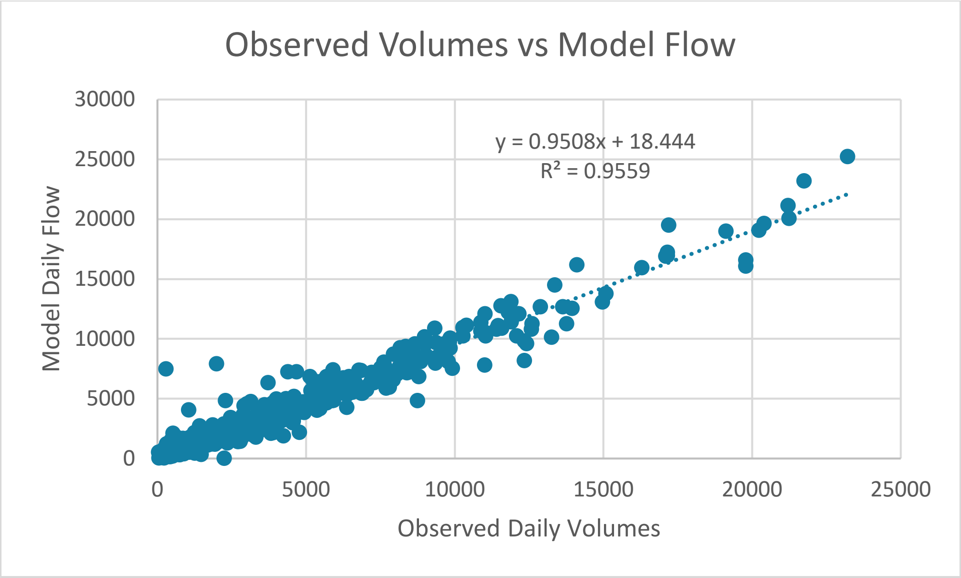


Figure 6-1: Daily Volume and Count Correlation

Table 6-2 shows traffic assignment validation statistics for the entire modeled area and compares these statistics with validation criteria from other states. As shown in Table 6-2, %RMSE values for the Mid-States TDM roadway segments by traffic volume ranges alongside the average %RMSEs for 10 statewide TDMs reported in NCHRP Report 836-B Task 91 (AASHTO, 2010). The 10 states in NCHRP Report 836-B

Task 91 are Alabama, Arizona, Florida, Indiana, Maryland, Michigan, Ohio, Oregon, Tennessee, Texas, Utah, and Wisconsin. This table also shows %RMSE acceptable limits from the Florida and Virginia Departments of Transportation. These %RMSE limits are widely used for statewide model validation in the United States. For all volume ranges, the Mid-States TDM %RMSE is either lower or almost equal to the average validation for the 10 statewide TDMs and are well within the validation criteria from Florida and Virginia.

Table 6-2: Daily Assignment Validation Statistics for the Entire Modeled Area

Roadway Daily Volume Range		% RMSE			
		Mid-States TDM	NCHRP 08-36[1]	Florida[2]	Virginia[3]
1	5,000	45.96	92.2	65	100
5,000	10,000	21.4	51.2	45	45
10,000	20,000	18.49	46.7	35	35
20,000	30,000	12.75	32.4	30	27
>30,000		23.6	22	25	25
All		**35**	**53**	**45**	**40**

[1] NCHRP Report 08-36B Task 91 (AASHTO, 2010)

[2] Urban Model Development Technical Report (Michigan DOT, 2019)

[3] Virginia DOT Travel Demand Modeling Policies and Procedures (VDOT, 2014)

Figure 6-2 graphically depicts model loading errors.

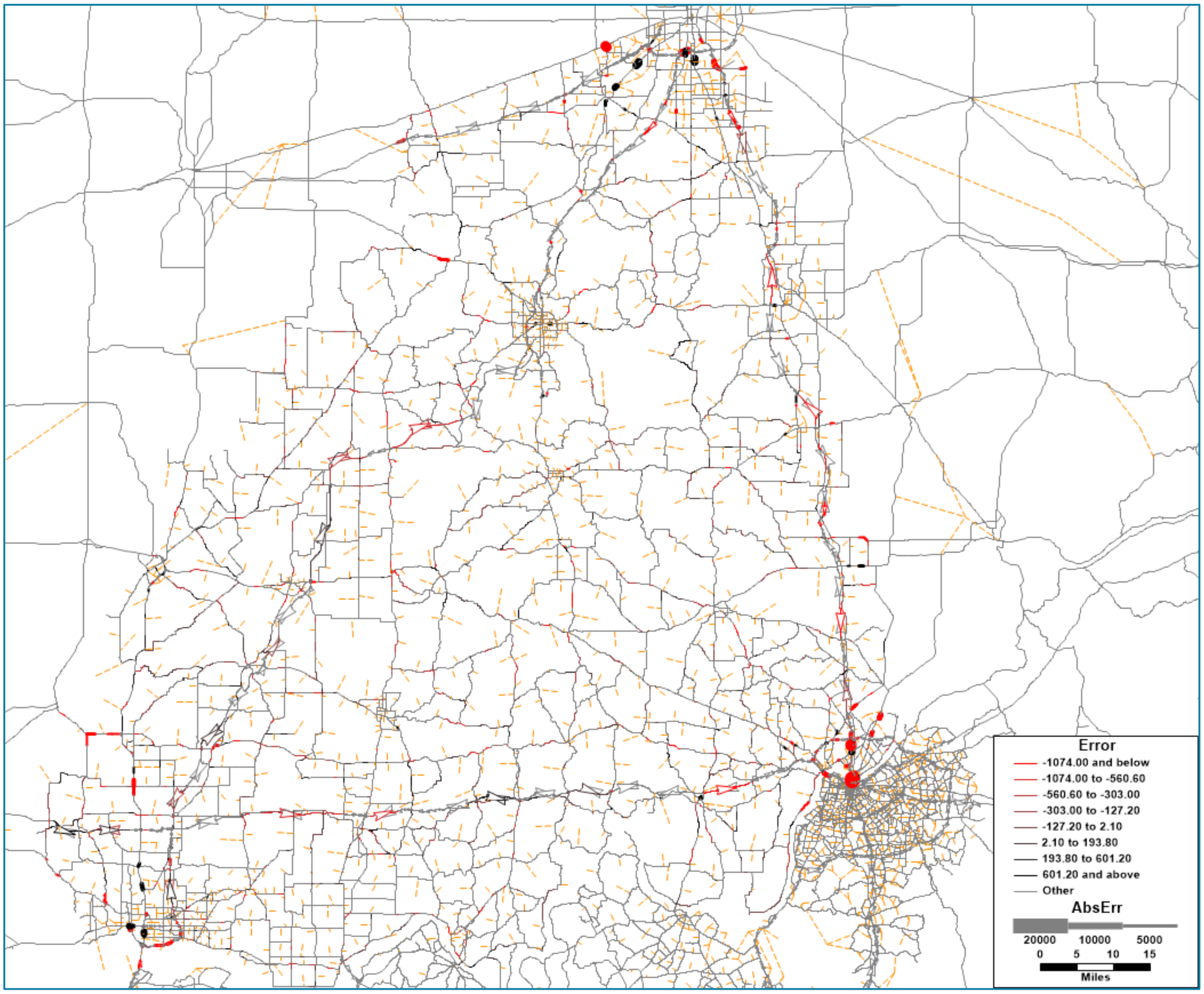


Figure 6-2: Model Loading Error

7 POST PROCESSING TOOLS

7.1 Post Processing Tools Description

Most performance measures were based upon calculations of vehicle miles traveled (VMT) and vehicle hours traveled (VHT). Depending upon the specification of the performance measures, VMT and VHT were calculated separately for autos and multi-unit trucks, as well as for specific origin-destination patterns. See **Appendix A** for details. VMT and VHT were calculated directly from assignment output files.

7.2 BCA Tool (RSG)

For the Mid-States model, RSG developed a benefit/cost analysis (BCA) post-processing tool. This tool estimates monetized benefits associated with multiple well-established factors including safety, travel time, and travel-time reliability.

A summary of the monetized benefits included in the BCA tool is summarized in **Table 7-1**.

Table 7-1: Benefit Factors

#	Benefit	Category	Type	Quantities
1	Safety	Safety	Link	Fatal, Injury, Property-Damage Only Crashes
2	Travel Time a) Passenger Vehicle b) Truck	Mobility	OD	Minutes of travel time saved by mode
3	Travel Time Reliability a) Passenger Vehicle b) Truck	Mobility	OD	Decrease in travel time variability (standard deviation of travel time)

The approach for incorporating each benefit factor is addressed below, including the methodology and monetization assumptions.

7.2.1 Safety

The safety benefit factor seeks to monetize the impact of changes in forecast motor vehicle crashes. The valuation of safety benefits has been a part of transportation BCA for decades; it was one of the original benefits in AASHTO's early Red Books in the 1960's and 1970's. Crash prediction methods used in BCA have varied, but AASHTO's Highway Safety Manual (HSM) is authoritative and widely used. RSG has used the HSM approach in various locations around the country and has found that it seems to work reasonably well, requiring only modest calibration coefficients.

The monetization of predicted crashes has also varied over time and place, but the release of the federal government's guidance on "value per statistical life" (VSL) has standardized the monetization of fatal crashes. Monetization of injury and PDO crashes still varies, but increasingly, injury crashes are valued in relation to the VSL along the lines used in the benefit calculator. Although there is uncertainty in any method, the methods for valuing safety benefits are well established and can be used with a relatively high degree of confidence.

Methods from the Highway Safety Manual (HSM) (AASHTO, 2010) are used for estimating safety benefits. The methods are implemented in the benefit calculator as they are implemented in FHWA's Interactive Highway Safety Design Model (IHSDM[10]) and documented in the Engineer's Manuals included in the tool's download with the help files. The Engineer's Manuals include all details for implementing the method, including estimated parameter values. In general, however, the method predicts the number of crashes using Safety Performance Functions (SPFs) for roadway segments (links) and intersections (nodes) together with Crash Modification Factors (CMFs).

The total annual number of crashes (N) are the sum of crashes along road segments (Nrs) and crashes at intersections (Nint).

$$N = N_{rs} + N_{int}$$

Both Nrs and Nint are predicted as the product of the number of crashes predicted by a SPF (NSPF), any relevant CMFs, and a calibration factor (Cr) that can be developed for particular jurisdictions or geographic areas to reproduce local observed crash rates or totals.

$$N_{rs} = C_r \times N_{SPFrs} \times CMF_1 \times \ldots \times CMF_n$$

$$N_{int} = C_r \times N_{SPFint} \times CMF_1 \times \ldots \times CMF_n$$

CMFs adjust baseline crash rates for specific conditions. CMFs can take various functional forms. Some CMFs are included as part of HSM/IHSDM, but additional CMFs can be found online.[11]

The SPFs predict the number of roadway segment or intersection crashes per year for nominal baseline conditions. Typically, for roadway segments, the SPFs take the form:

$$N_{SPFrs} = \alpha(\beta ADT)^{\gamma} \times Length\ (mi)$$

Where α, β, and $\gamma$are parameters for a given facility type and sometimes other specifics such as number of lanes. Typically, for intersections, the SPFs take the form:

$$N_{SPFint} = \alpha + \beta ADTonHighestVolumeApproach + \gamma ADTonLowestVolumeApproach$$

Where α, β, and $\gamma$are parameters for a given facility/area type. Some basic processing of the network data is required to compute *ADTonHighestVolumeApproach* and *ADTonLowestVolumeApproach* by joining highway network node and link data and these calculations are built into the benefit calculator.

CMFs which use information available for the whole model network, such as number of lanes and truck percentage are included in the benefit calculator. The method predicts total crashes, which are split into crash severity categories (fatal, injury, and property-damage-only (PDO)).

The monetary value of fatality and injury collisions was calculated based on the US DOT "Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation

[10] http://www.fhwa.dot.gov/research/tfhrc/projects/safety/comprehensive/ihsdm
[11] http://www.cmfclearinghouse.org

Analyses – 2015 Adjustment", which gives the valuation of fatality collisions at $9.4 million. The valuation is based on empirical studies and is defined as the additional cost that individuals would be willing to bear for improvements in safety (that is, reductions in risks) that, in the aggregate, reduce the expected number of fatalities by one. For injury collisions, the US DOT approach monetizes by applying a factor to the VSL based on severity (AIS1-5). Using the AIS2 (moderate) factor of 0.047 results in a valuation of injury collisions at approximately $441,800. For property-damage-only (PDO) collisions, we selected a value of approximately $1,522. **Table 7-2** presents the parameters used for crash monetization.

Table 7-2: Final Crash Monetization Parameters

Parameter	Value
Fatality Cost	$9,400,000
Injury Cost	$441,800
Property-Damage-Only Cost	$1,522

The motor vehicle crashes safety benefit calculation is implemented as a series of link calculations. A link-based approach is preferred to a zone-based approach since it avoids issues related to aggregation bias caused by using an area-based unit of analysis for network-level phenomena. The generic SPF illustrated in the Method subsection above is implemented separately for each functional class – freeways, rural two-lane highways, rural multi-lane highways, and urban/suburban arterials. The CMFs corresponding to each functional class SPF are estimated using various network attributes such as lane width, shoulder width, grade, presence of median/barrier, lighting, etc. Default or average values of certain network attributes are used when not available. Specifically, average lane widths and shoulder widths assumptions are based on FHWA's Highway Functional Classification Concepts, Criteria and Procedures (Table 3-5)[12]. In cases where the appropriate network attributes are not available, the CMF value is set to 1.0.

Since the benefits calculator operates at the link level, node level attributes for intersection SPFs are computed via link level calculations. This involves identification of intersections (non-centroids and nodes connected to more than two links) by their control type and computation of the minimum and maximum volume at the intersection. Since all the calculations in the benefits calculator are implemented at the link level, each intersection will appear in the processor as many times as the number of links connected to it. Therefore, to avoid double counting, the SPF for each intersection is divided by the number of links it is connected to.

For the Mid-States application, the CMFs for intersections are set to 1.0 since the required operational details (lighting, angle, etc.) are not available in the network. For pedestrian crashes, a medium pedestrian activity is assumed. Finally, a multiplicative calibration factor is added to each SPF which is set to 1.0 for the Mid-States application.

The monetization rates for crashes vary by crash types – fatal, injury and PDO crashes – as described in the Monetization subsection above. Distribution factors are used for splitting total crashes into different crash categories. Crash distribution factors were determined using regional data and are summarized in **Table 7-3**.

[12] https://www.fhwa.dot.gov/planning/processes/statewide/related/highway_functional_classifications/fcauab.pdf

Table 7-3: Crash Distribution Factors by County

County	Fatality	Injury	PDO
CRAWFORD	0.007	0.146	0.847
DAVIESS	0.023	0.319	0.658
DUBOIS	0.005	0.206	0.789
GREENE	0.017	0.192	0.791
LAWRENCE	0.011	0.220	0.769
MARTIN	0.021	0.289	0.691
MONROE	0.002	0.220	0.777
ORANGE	0.007	0.210	0.782
PERRY	0.003	0.185	0.812
PIKE	0.009	0.309	0.682
SPENCER	0.013	0.227	0.759
WARRICK	0.005	0.214	0.781

7.2.2 Travel Time

Travel time savings are a significant benefit factor for most transportation projects, plans, and policies.

For existing trips, travel time savings are simply the decrease in travel time for that trip. However, when trips are induced or suppressed, the benefit is calculated based on consumer surplus theory (as shown in **Figure 7-1**). The basic idea for induced demand is that although the traveler was unwilling to make the trip given the original travel time (cost), as the cost decreases, at some point the traveler would choose to make the trip. The travel time savings for an induced trip should be measured as any further decrease in travel time beyond the point at which the trip is induced.

In the absence of other information, the "Rule of Half" (ROH) assumes that trips induced between a baseline cost and an alternative scenario cost would, on average, be induced at the average of these costs and hence should accrue half of the travel time savings as existing trip-makers. In economic terms, this benefit for induced demand is the change in consumer surplus, and the ROH amounts to linearization of the travel demand function. This method has been applied to user costs and travel time savings in particular in the context of transportation BCA for many years and is established good practice.[13]

[13] See Abelson, P. and D. Hensher. "Induced Travel and User Benefits: Clarifying Definitions and Measurement for Urban Road Infrastructure." In Handbook of Transport Systems and Traffic Control edited by Kenneth J. Button and David A. Hensher. Pergamon, 2001.

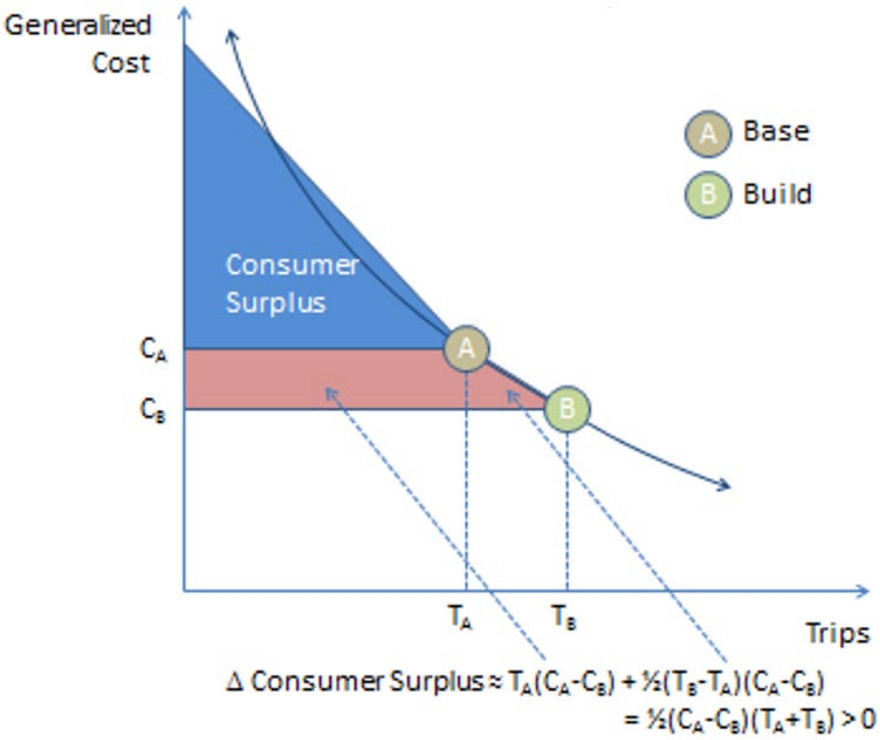

Generalized Cost
A Base
B Build
Consumer Surplus
C_A
C_B
A
B
T_A
T_B
Trips
Δ Consumer Surplus $\approx T_A(C_A-C_B) + \frac{1}{2}(T_B-T_A)(C_A-C_B)$
$= \frac{1}{2}(C_A-C_B)(T_A+T_B) > 0$

Figure 7-1: Consumer Surplus

The valuation of travel time is a challenging topic, much debated in literature. Many different studies and methodologies have produced a range of estimates of the value of travel time savings. These studies help establish a reasonable range of values but can make assignment of a particular value difficult. For these reasons, the value of time (VOT) settings are user configurable in the benefit calculator so that the default VOT assumptions can be changed.

US DOT offers guidance on VOT in "Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis" [14]. This guidance recommends against using separate VOT for different travelers, in part to ensure economic analyses do not favor projects or policies beneficial to higher income citizens over those beneficial to lower income citizens. While the guidance acknowledges evidence that VOT vary by purpose in general, it only recommends differentiating between on-the-clock travel for business from all other personal travel (including commuting) if an analysis can support this distinction. The benefit calculator, therefore, allows for separate VOT by trip purpose, but we have populated all trip purposes with a single VOT of $12.63, derived by adjusting the national VOT for all purposes ($14.10/hr) to reflect average incomes in the 12-county study area.

US DOT guidance acknowledges that the total value of truck time also involves the opportunity cost of the time the goods transported are in transit, which varies considerably depending on factors associated with the freight being carried, including the rate at which the goods will become obsolete, whether the goods are perishable, and dependency on timely delivery of the goods for downstream production

[14] 2016 Revised Value of Travel Time Guidance.pdf (transportation.gov)

processes. As a result, there is not a set recommendation on the total value of truck time. For this study a default value of time of $45.00/hr was used for trucks.

The benefit calculator includes matrix-based calculation of the rule-of-half (ROH) using demand matrices and network skims from the travel model to estimate travel time savings including changes in consumer surplus. The final ROH based travel time savings are converted to monetary values by multiplying by VOT. The temporal discount rate and annualization factor are applied to estimate annual travel time savings.

7.2.3 Travel Time Reliability

Travel time reliability is a measure of unexpected delay. As defined by FHWA, travel time reliability is the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day.[15] The SHRP 2 program has conducted considerable research on travel time reliability in recent years, and several methods for estimating travel time reliability and its value have now been demonstrated. RSG conducted a meta-analysis of the literature, including the AASHTO Redbook[16], SHRP 2 L03[17,] L04[18,] L05[19], L11[20], and C04[21]. From these various approaches a consensus / ensemble predictor was produced for the buffer time index as a function of volume to capacity (v/c) ratio. The various published functions are shown together with the ensemble function produced by the meta-analysis in **Figure 7-2**.

[15] Travel Time Reliability: Making It There On Time, All The Time. 2006. FHWA, http://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm

[16] American Association of State Highway and Transportation Officials (AASHTO), 2010. *User and Non-User Benefit Analysis for Highways*.

[17] National Academies of Sciences, Engineering, and Medicine, 2012. *Analytical Procedures for Determining the Impacts of Reliability Mitigation Strategies*. Washington, DC: The National Academies Press.

[18] National Academies of Sciences, Engineering, and Medicine, 2014. Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools. Washington, DC: The National Academies Press.

[19] National Academies of Sciences, Engineering, and Medicine, 2013. Incorporating Reliability Performance Measures into the Transportation Planning and Programming Processes. Washington, DC.

[20] National Academies of Sciences, Engineering, and Medicine, 2013. Evaluating Alternative Operations Strategies to Improve Travel Time Reliability. Washington, DC: The National Academies Press.

[21] National Academies of Sciences, Engineering, and Medicine, 2012. Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand. Washington, DC: The National Academies Press.

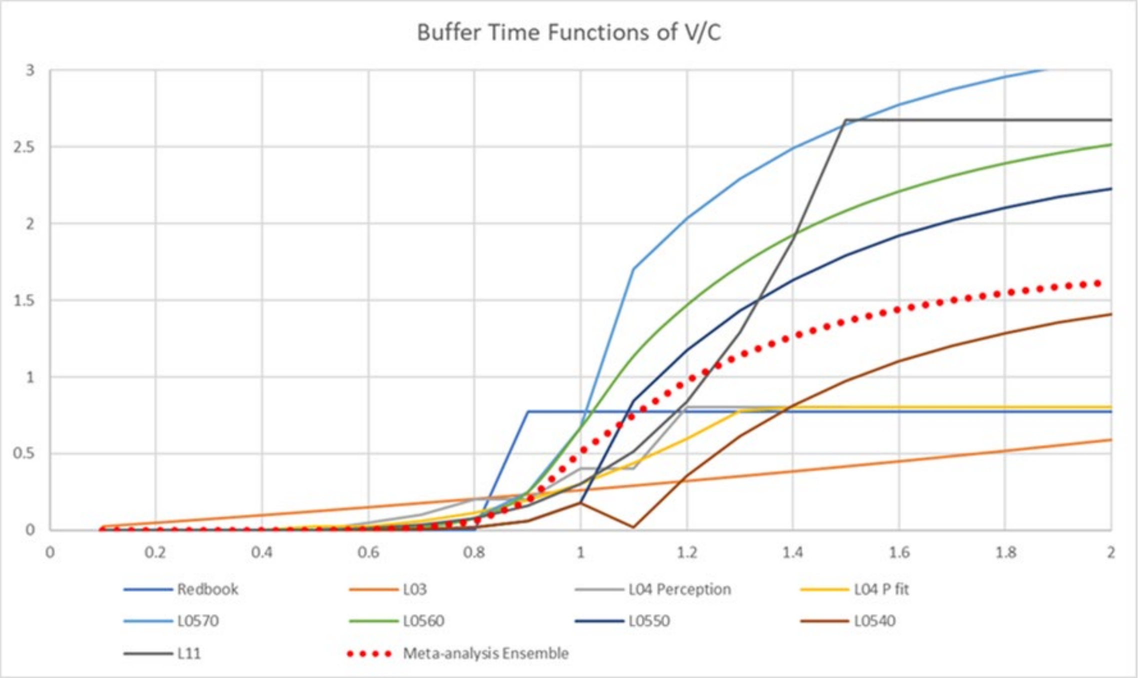


Figure 7-2: Ensemble Predictor of Buffer Time From Meta-Analysis

The ensemble buffer time function produced by the meta-analysis is given below:

$$Buffer\ Time = 3.67\ \times average\ congested\ time\ \times \ln\left(\frac{55}{D}\right)$$

where

$$D = \begin{cases} if\, \frac{v}{c} < 1 & \frac{55}{1 + 0.15 \times (v/c)^{10}} \\ if\, \frac{v}{c} \geq 1 & 33.5 + 15 \times (v/c)^{-3} \end{cases}$$

The reliability measure and skim are applied as a model post-processor for the estimation of reliability benefits for the calculator. The benefit calculator then estimates the reliability benefits using the same ROH OD matrix calculation that is used for calculating travel time savings. The reliability benefits are converted to a monetary value by multiplying with the VOT parameters described in the Travel Time section, above. Separate calculations are performed for passenger vehicle and truck reliability benefits. The temporal discount rate and annualization factor is applied to the total costs to obtain annual costs.

8 CONCLUSION

Advanced travel forecasting tools were developed for the Mid-State Corridor Study. These tools represent the state of the practice with regards to Big Data, ensemble forecasting and travel time benefits analysis. These tools accurately represent travel behaviors observed over a large multi-state region. The travel forecasting model limits extend from Indianapolis into Northern Tennessee to capture the full extent of north-south travel patterns that could be attracted to the Mid-State Corridor.

The travel forecasting model was developed using passive data collected from cell phones and GPS devices as the foundation. An extensive process was employed to eliminate inherent biases and format the data for model application. The use of Big Data provides unparalleled insights into existing travel patterns that were simply not available just years earlier. The model's RMSE of 23.17 confirms a highly accurate model validation that exceeds standard of practice validation targets.

A three-step model was developed to forecast travel in the future. However, this standard approach was bolstered by ensemble techniques, which combine traditional trip generation and distribution methods with machine learning and trend-based methods using longitudinal passive data. This approach recognizes that all models have errors but utilizing multiple different models cancels some of the errors, resulting in a more refined forecast.

Assumptions regarding future growth and socioeconomic indicators were based on forecasts published by the applicable state and metropolitan planning organizations. The travel forecasting model's spatial structure (TAZs) were fully nested within that of the Indiana Statewide Travel Demand Model to promote consistency with the state's forecasting tool. This consistency was reinforced by regular coordination with the Indiana Office of Planning & Programming that occurred during development of the travel forecasting model.

A range of post-processing tools were used to quantify and compare project benefits. These benefits are detailed in **Appendix A – Transportation Performance Measures**. These including analyses of traffic assignments to measure performance on core goals. These core goal performance included improved accessibility between key travel pairs, improved labor force access to employment centers, efficiencies in truck freight movements and improved access to major intermodal centers. The travel model also was used to measure performance on secondary goals which reflected other desirable outcomes. These included crash reductions at key locations and reduced congestion within Dubois County.

9 BIBLIOGRAPHY

AirSage Data Expansion. Charlotte DOT, 2017.

Alexander, L., S. Jiang, M. Murga and M. Gonzalez. Origin-Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 240-250.

American Association of State Highway and Transportation Officials (AASHTO). Highway Safety Manual, 1st Edition. 2010.

Bernardin, V. and L. Amar. Using Large Sample GPS Data to Develop an Improved Truck Trip Table for the Indiana Statewide Model. Presented at the 4th TRB Conference on Innovations in Travel Modeling, Tampa, Florida, May 2012.

Bernardin, V., S. Trevino, and J. Short. Expanding Truck GPS-based Passive Origin-Destination Data in Iowa and Tennessee. Presented at the 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

Bernardin, V., J. Chen, S. Trevino, and Y. Lee. Incorporating Big Data in an Activity-based Model for Chattanooga. Presented at the 16th National TRB Transportation Planning Applications Conference, Raleigh, NC, 2017.

Bernardin, V., N. Ferdous, H. Sadrsadat, S. Trevino, and C. Chen. "Integration of the National Long Distance Passenger Travel Demand Model with the Tennessee Statewide Model and Calibration to Big Data." *Transportation Research Record: Journal of the Transportation Research Board*. No 2653, 2017, pp. 75-81.

Bernardin, V., and H. Sadrsadat. Review of Methods for Data Validation and Expansion of Passively Collected Origin-Destination Data. Presented at the 97th Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.

Big Data Analytics for the Northeast Indiana Region. Draft Report. Northeastern Indiana Regional Coordinating Council, 2017.

Bindra, S., B. Grady and J. Deshaies. Using Cellphone Origin-Destination Data for Regional Travel Model Validation. Presented at the 15th National TRB Transportation Planning Applications Conference, Atlantic City, NJ, 2015.

Calabrese, F., M. Colonna, P. Lovisolo, D. Parata and C. Ratti. Real-time Urban Monitoring using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems*. Vol. 12. No. 1, 2011, pp. 141-151.

Calabrese, F., G. Di Lorenzo, L. Liu and C. Ratti. Estimating Origin-Destination Flows using Mobile Phone Location Data. *IEEE Pervasive Computing*. Vol. 10, No. 4, 2011, pp. 36-44.

CDM Smith, Lochmueller Group Inc. Illinois Statewide Travel Demand Model Network and TAZ Development Technical Documentation. 2019.

CDM Smith. South Carolina Statewide Model Documentation. 2018.

Corradino Group, WSP and Stantec. Model Update, Kentucky Statewide Traffic Model. 2012.

Donnelly, R. and J. Kressner. Forecasting with Data-Driven Models. Presented at the 16th National TRB Transportation Planning Applications Conference, Raleigh, NC, 2017.

ETC Institute and Lochmueller Group. IndyGo On-Board Transit Survey – Final Report. May 15, 2017

Gur, Y. J., S. Bekhor, C. Solomon and L. Kheifits. Intercity Person Trip Tables for Nationwide Transportation Planning in Israel Obtained from Massive Cell Phone Data. *Transportation Research Record: Journal of the Transportation Research Board*. No 2121, 2009, pp. 145-151.

Han, Y., K. Kaltenbach, S. Thomson, J. Balaji and D. Hulker. Innovative Analysis Methods of Mobile Phone Data in the Best Travel Demand Modeling Practice in Kentucky. Presented at the 15th National Tools of the Trade Conference, Charleston, SC, September, 2016.

Huntsinger, L. F. and R. Donnelly. Reconciliation of Regional Travel Model and Passive Devise Tracking Data. Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

Iqbal, M. S., C. F. Choudhury, P. Wang, M. Gonzalez. Development of Origin-Destination Matrices using Mobile Phone Call Data. *Transportation Research Part C: Emerging Technologies*, Vol. 40, 2014, pp. 63-74.

Kressner, J. and L. Garrow. Using Third-Party Data for Travel Demand Modeling: Comparison of Targeted Marketing, Census, and Household Travel Survey Data. *Transportation Research Record: Journal of the Transportation Research Board*. No 2442, 2014, pp. 8-19.

Lee, R. J., I. N. Sener and J. A. Mullins. An Evaluation of Emerging Data Collection Technologies for Travel Demand Modelling: from Research to Practice. *Transportation Letters: The International Journal of Transportation Research*. Vol. 8, No. 4, 2016, pp. 181-193.

Ma, J., F. Yuan, C. Joshi, H. Li and T. Bauer. A New Framework for Development of Time-Varying O-D Matrices based on Cellular Phone Data. Presented at the 4th TRB Innovations in Travel Modeling Conference, Tampa, FL, 2012.

McAtee, S. Validating Trip Distribution in Southeast Michigan using GPS Data. Presented at the 16th National TRB Transportation Planning Applications Conference, Raleigh, NC, 2017.

Milone, R. Preliminary Evaluation of Cellular Origin-Destination Data as a Basis for Forecasting Non-Resident Travel. Presented at the 15th National TRB Transportation Planning Applications Conference, Atlantic City, NJ, 2015.

Rojas, M., E. Sadeghvaziri and X. Jin. Comprehensive Review of Travel Behavior and Mobility Pattern Studies That Used Mobile Phone Data. *Transportation Research Record: Journal of the Transportation Research Board*. No 2563, 2016, pp. 71-79.

Toole, J. L., S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, M.C. Gonzalez. The Path Most Traveled: Travel Demand Estimation using Big Data Resources. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 162-177.

Wang, P., T. Hunter, A. M. Bayen, K. Schechtner and M. C. Gonzalez. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports*. Vol. 2, No. 1001, 2012.

Wang, J., D. Wei, K. He, H. Gong and P. Wang. Encapsulating Urban Traffic Rhythms into Road Networks. *Scientific Reports*. Vol. 4, No. 4141, 2014.

WSP. North Carolina State Travel Model, Model Development Technical Documentation. 2019.

Zandbergen, P. Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. Transactions in GIS. Vol. 13, No. 1, 2009, pp. 5-25.

Zanjani, A., A. Pinjari, M. Kamali, A. Thakur, J. Short, V. Misore, and S. Tabatabaee. Estimation of Statewide Origin-Destination Truck Flows from Large Streams of GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*. No 2494, 2015, pp. 87-96.

Zhang, W., A. Kuppam, V. Livshits and B. King. Evaluation of Cellular-based Travel Data – Experience from Phoenix Metropolitan Region. Presented at the 15th National TRB Transportation Planning Applications Conference, Atlantic City, NJ, 2015.