North Carolina Model Users Group
November 1, 2018

Incorporating Big Data in Model Development

Stantec

Overview

- Key Topics
  - Big Data Usage
  - Calibration Applications
- Four Projects
  1. Greenville Model
  2. Mid-Currituck Bridge
  3. Lake Pontchartrain Causeway
  4. Central Texas Turnpike System
- Data Types
  1. Speed Data
  2. Trip Pattern Data

Key Topics

1. Big Data Usage:
   - Prior Data had Limitations
   - Data & Tools are improving
   - Data Evaluation is Critical

2. Model Calibration Objectives:
   - Trips by Vehicle Type
   - Speeds
   - Trip Patterns

Big Data Evolving and Improving

- Passive O-D Data Samples have expanded in recent years
  - Samples now much greater than obtained from household surveys
- Increased use of GPS-enabled vehicles and Smart Phones
  - Replaces less accurate approximation of cell phones via triangulation
  - Enables physical tracing of vehicles within network

Location-Based Services Data

Location Based Services data is provided from smartphone apps that track the locations of phones and other devices to provide specific services, such as weather forecasts, shopping options and restaurant reviews as well as other services. There are data available from hundreds of these apps and number of apps continues to expand.

O-D Sample Size Limitations with Various Methods

- Typical HH Survey Data < 1%
- Prior Versions of Streetlight <2%

Larger Sample Size Yield Better Understanding of Travel Patterns

- As an example, Streetlight Data is currently at 23% sample
- Representative disaggregate samples

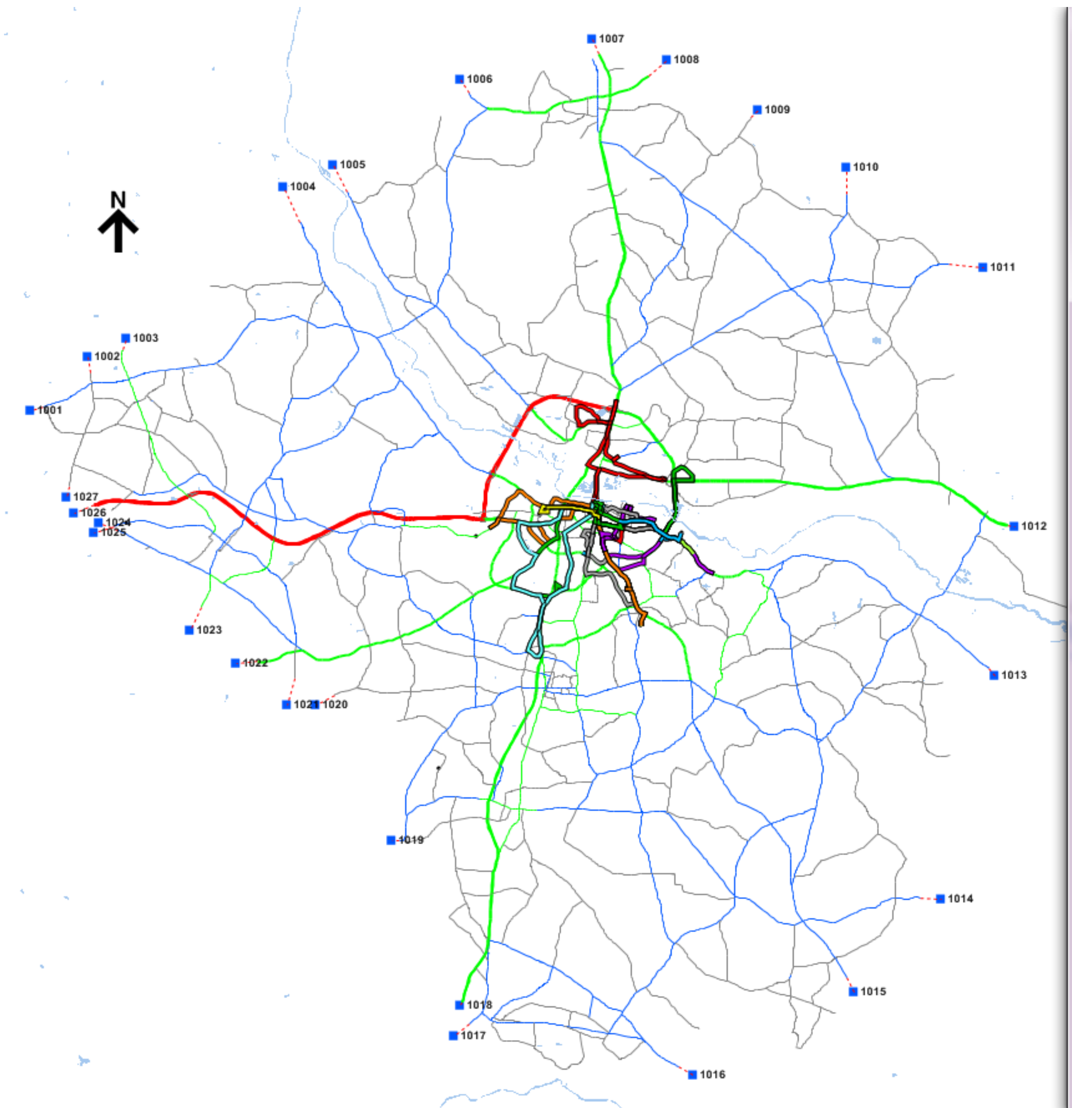
Passive Data Limitations

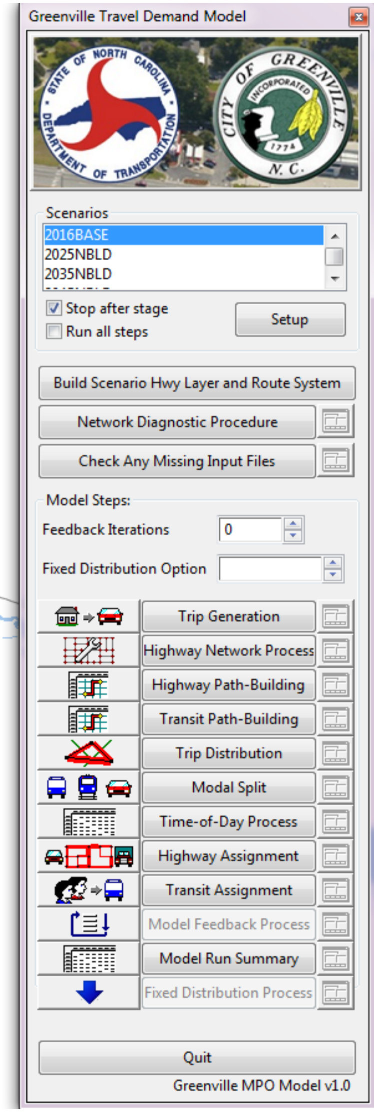
- Spatial Precision
  - LBS data:
    - 25 meter spatial precision
    - Pings are sent as devices are moving
  - Cellular data:
    - 100-300 meter spatial precision
    - Pings are sent less frequently
- Person Trip Characteristics
  - Traveler Information
    - inferred from home zone
  - Purpose Characteristics
    - inferred from frequency and duration
    - Aggregation into generic purposes (HBW, HBO, NHB)
- Truck Samples are still relatively small

Passive Data Limitations

- Device Activation
  - LBS data relies on users proactively opting in at apps that track location. Battery power consumption may restrict some usage.
  - Cellular data does not require proactive opting in
- Research has Identified Observed Biases
  - LBS data may be under-estimating short district trips.
- Passive Data should be Evaluated
  - O-D Patterns
  - Speeds

Greenville Model Development Project



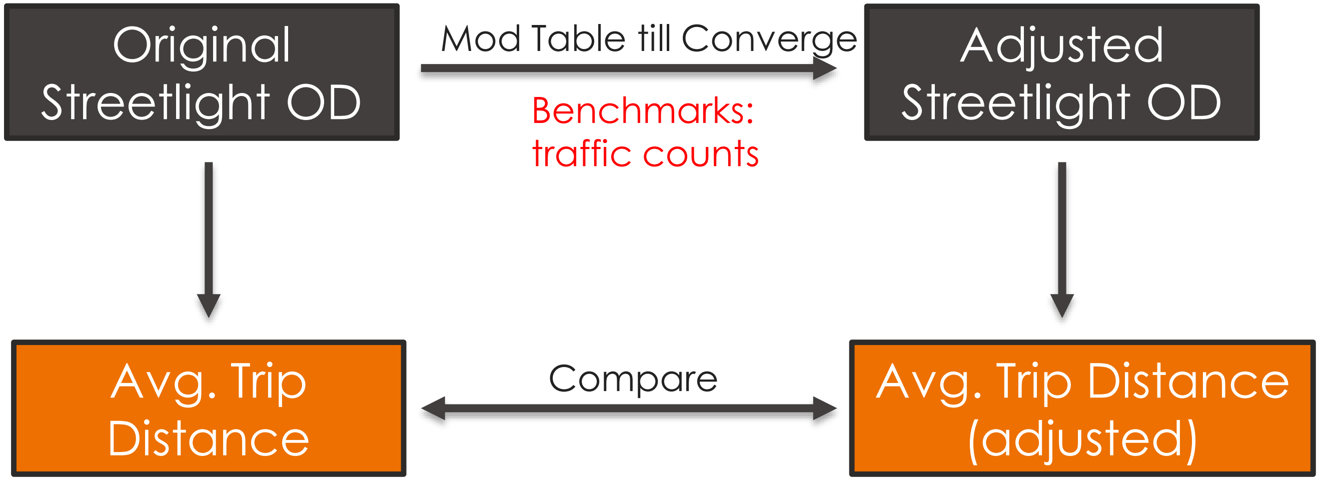


Greenville Model - Passive Data Evaluation

- Origin-Destination Patterns
  - If O-D is Under-Representing Short Distance trips, Capture that Difference via ODME Techniques
  - Effectively creates Band of Variation by Impedance Interval
- Speed Data
  - Verify HERE data with Independent Source (Google)

Methodology

- Use Origin-Destination Matrix Estimation (ODME) to Identify Differences by Impedance Intervals

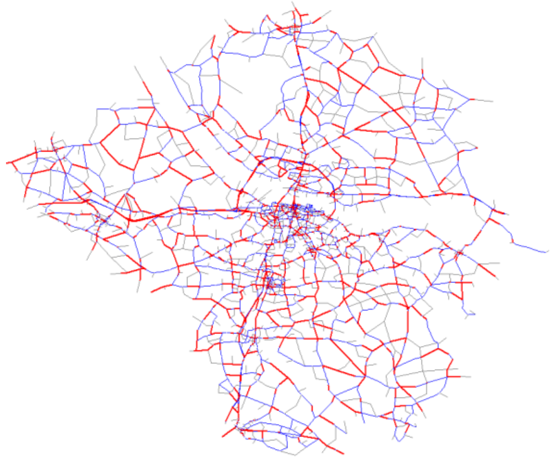


Traffic Count Coverage – Pitt County

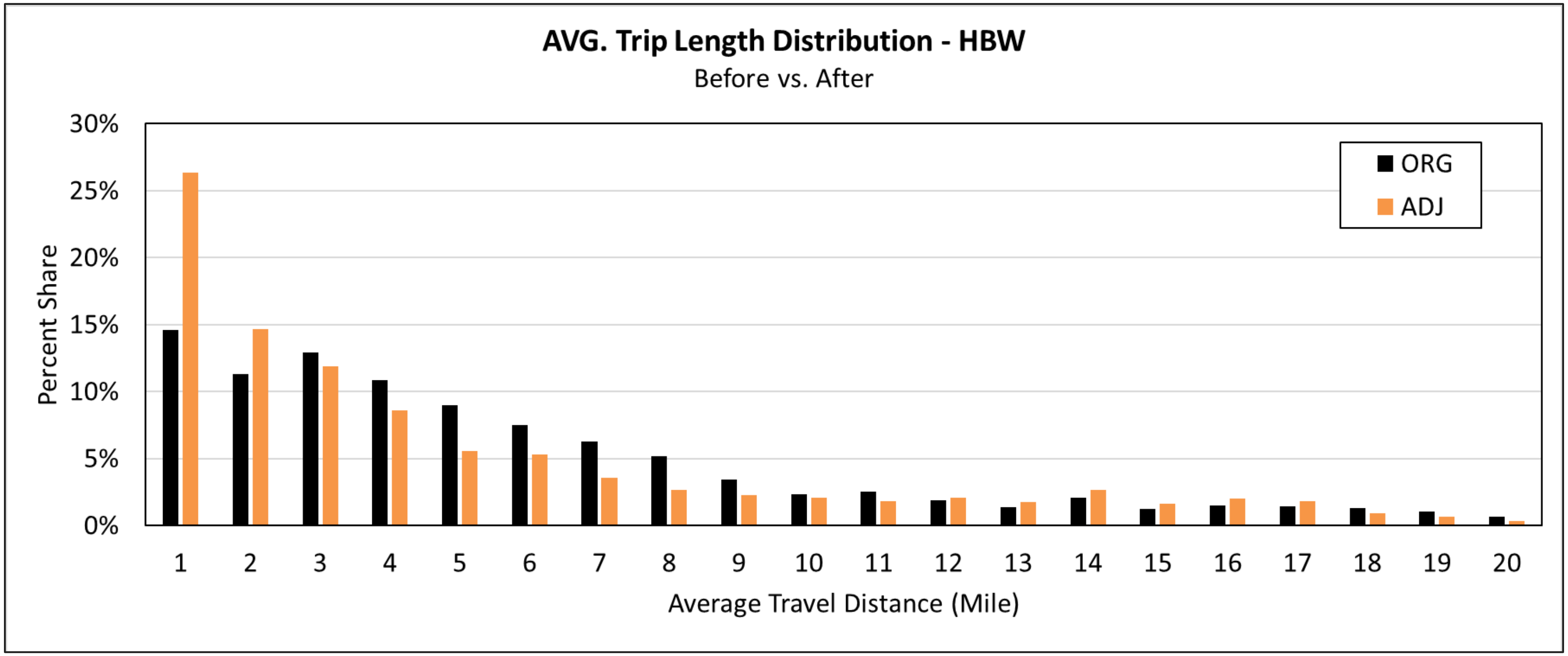
TOTAL COUNTS

FACILITY TYPE	AREA TYPE		
	Urban	Rural	TOTAL
Freeway	--	30	30
Principal Arterial	213	72	285
Minor Arterial	563	112	675
Major Collector	294	446	740
Minor Collector	--	152	152
Local Road	188	530	718
Low-speed Ramp	1	3	4
High-speed Ramp	4	18	22
TOTAL	**1,263**	**1,363**	**2,626**

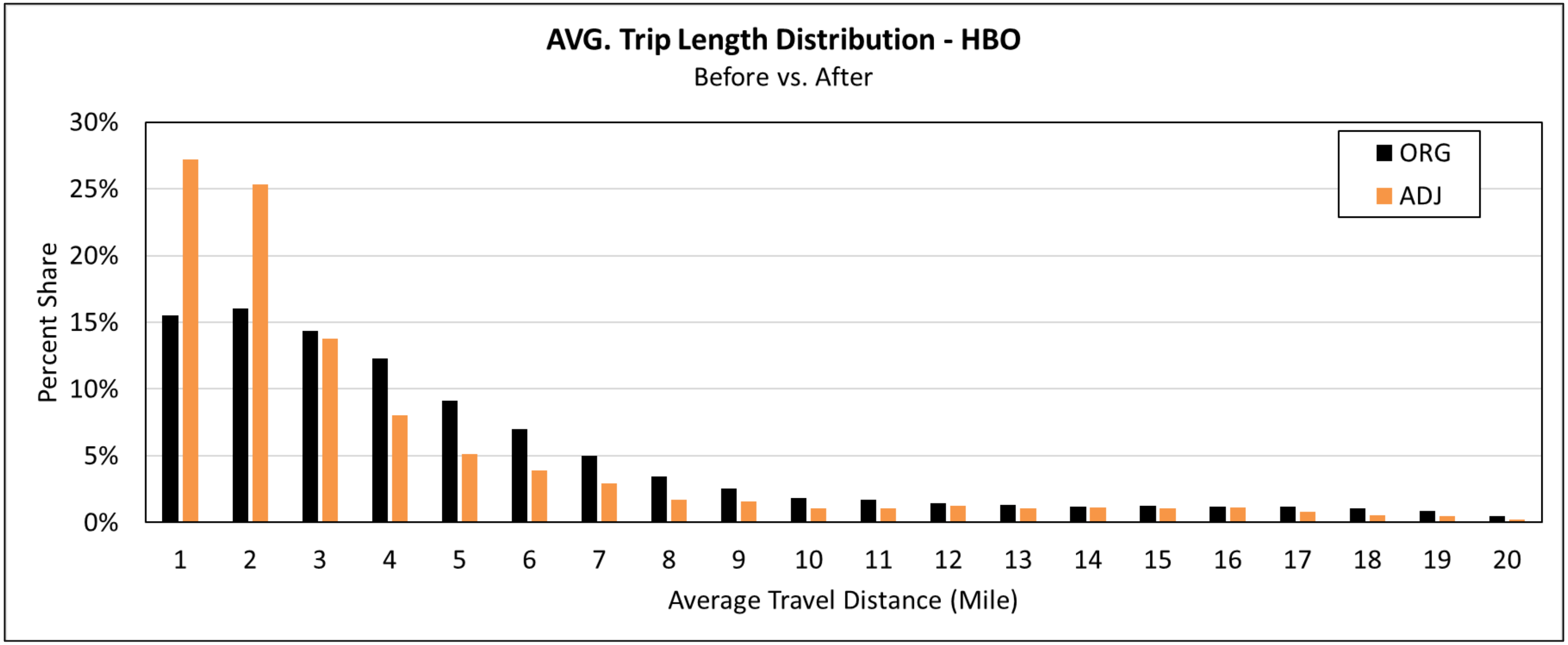
32.3% data coverage



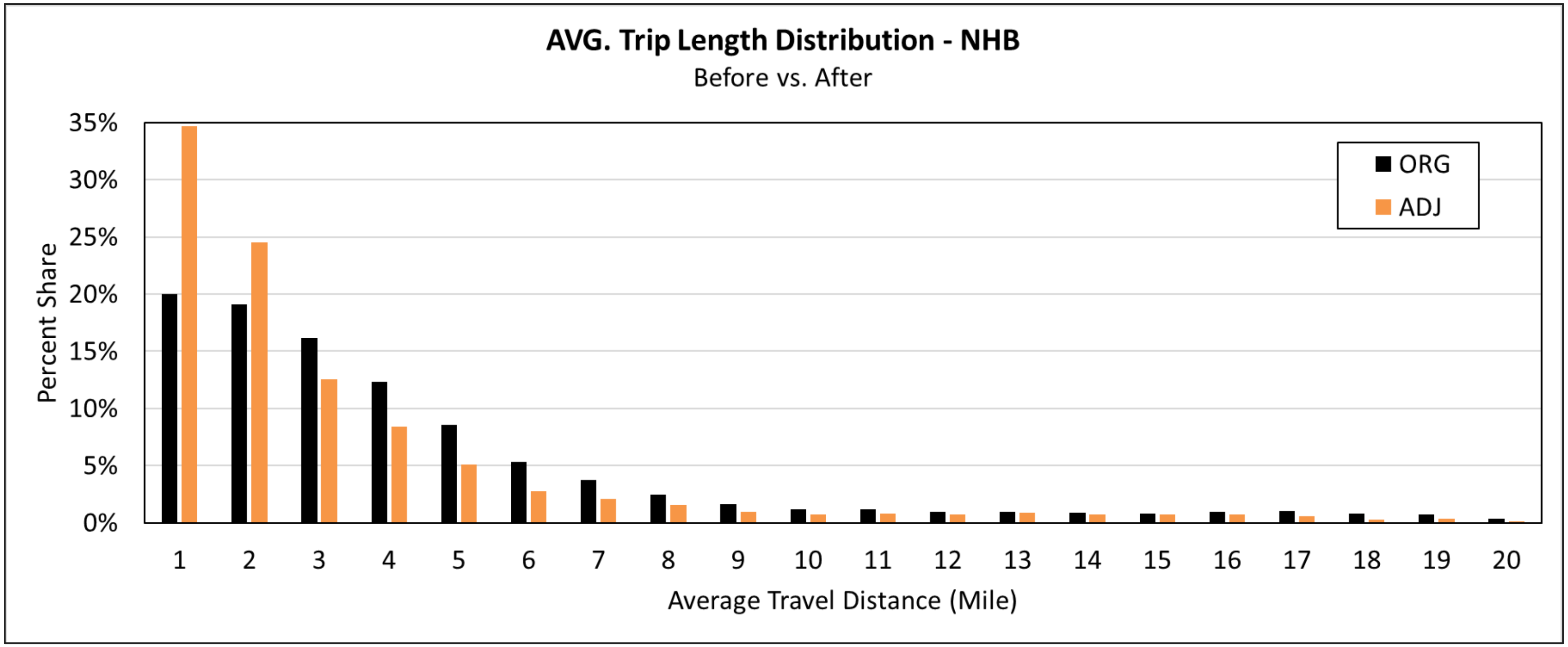
Average Trip Length Distribution - HBW



Average Trip Length Distribution - HBO



Average Trip Length Distribution - NHB



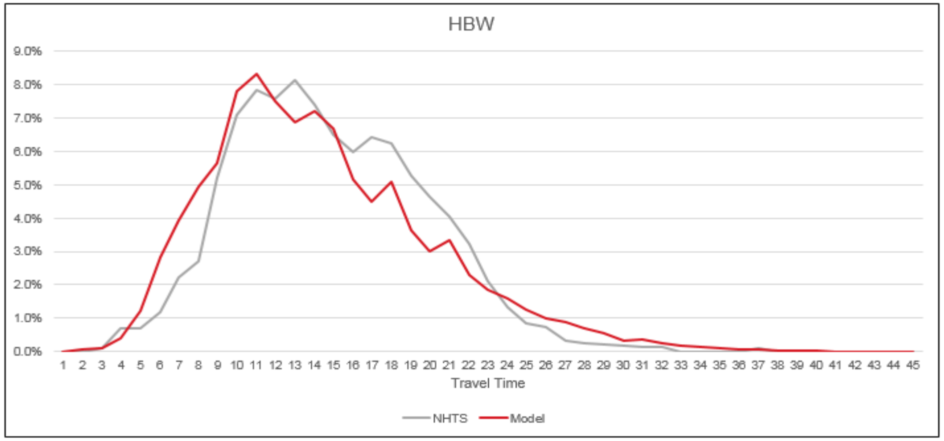
Adjusted Average Trip Length

PURPOSE	Avg. Distance			Avg. Travel Time		
	ORG	ADJ	%DIFF	ORG	ADJ	%DIFF
HBW	5.8	5.5	**-5%**	11.6	11.2	**-4%**
HBO	5.2	4.4	**-16%**	10.6	9.6	**-10%**
NHB	4.3	3.7	**-15%**	9.1	8.5	**-7%**

Greenville Model – Distribution Calibration

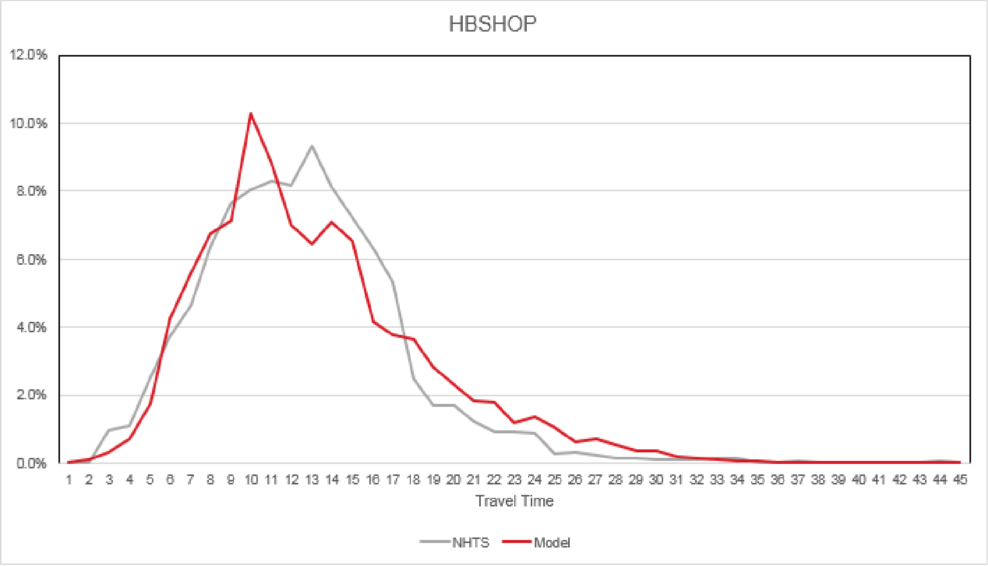
- Person Trips
  - Calibrate Individual Purposes using NHTS Data
    - HBW
    - HBSH
    - HBO
    - NHBW
    - NHBO
  - For HBO and NHB, Aggregate to Streetlight Purposes
  - E-I Purposes use Streetlight Aggregate Trip Purposes
- Truck Trips
  - Use Streetlight Patterns by Truck Type
  - Separate Internal and E-I Distributions
- E-E Trips
  - Use Streetlight Patterns by Vehicle Type

Calibration using 2017 NHTS
(HBW & HBSH)



Source	Time
NHTS	14.33
Model	13.92

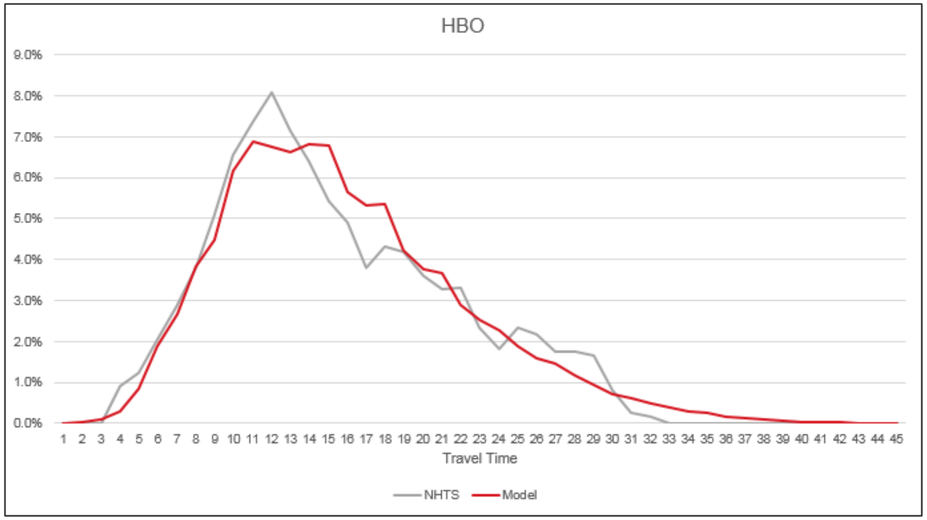
Source	Intra Zonal
NHTS	2.1%
Model	1.3%



Source	Time
NHTS	12.07
Model	12.65

Source	Intra Zonal
NHTS	3.2%
Model	2.1%

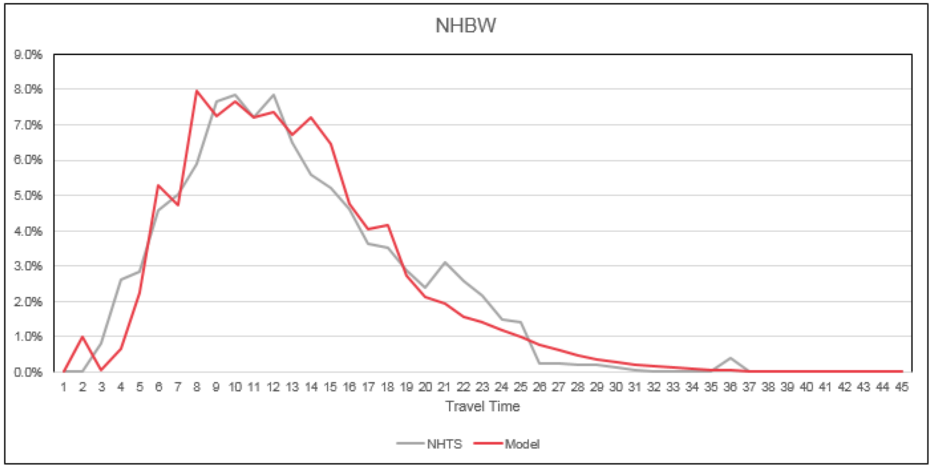
Calibration using 2017 NHTS
(HBO)



Source	Time
NHTS	14.89
Model	15.37

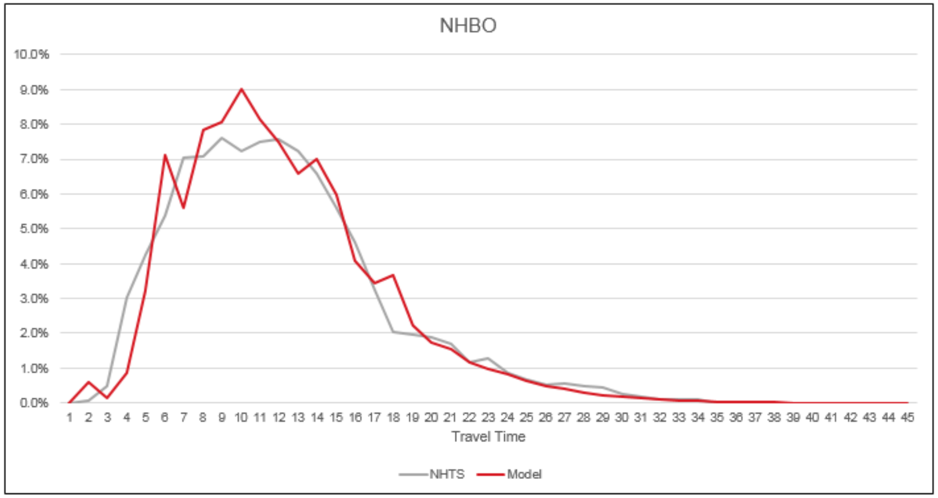
Source	Intra Zonal
NHTS	0.5%
Model	1.2%

Calibration using 2017 NHTS
(NHBW & NHBO)



Source	Time
NHTS	12.58
Model	12.61

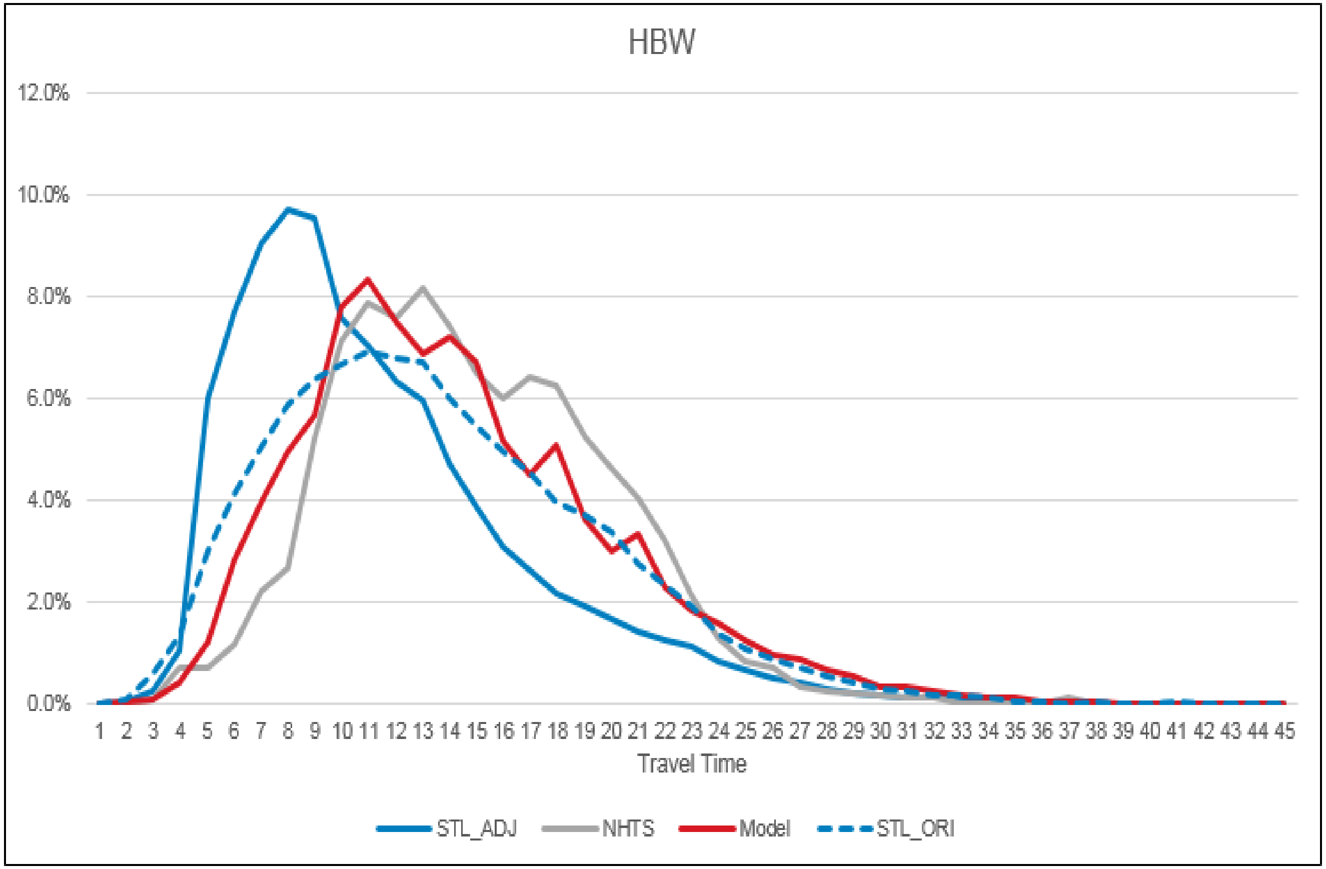
Source	Intra Zonal
NHTS	6.6%
Model	4.6%



Source	Time
NHTS	11.59
Model	11.75

Source	Intra Zonal
NHTS	5.8%
Model	6.6%

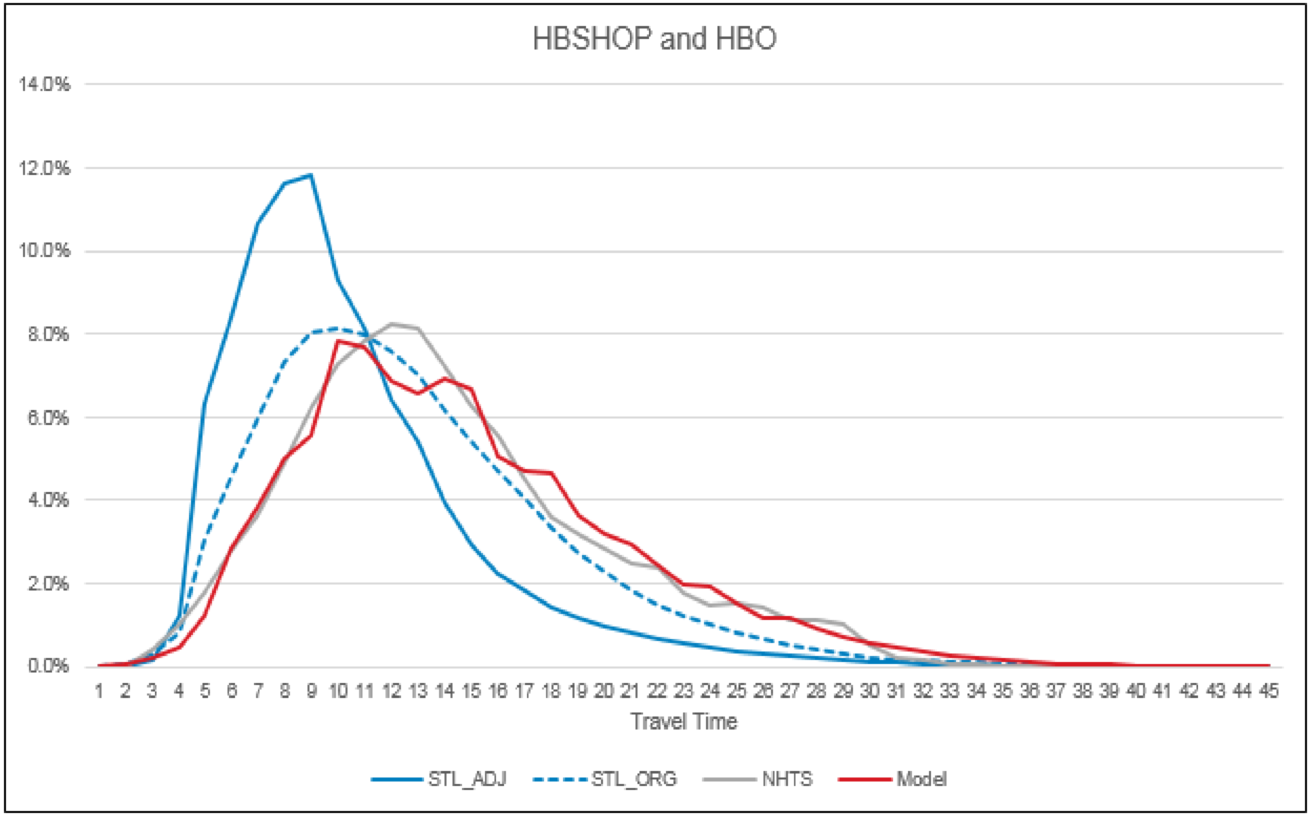
HBW Distribution Validation



Source	Time
STL_ADJ	10.79
STL_ORG	13.11
NHTS	14.33
Modeled	13.92

Source	Intra Zonal
STL_ADJ	1.4%
STL_ORG	3.2%
NHTS	2.1%
Modeled	1.3%

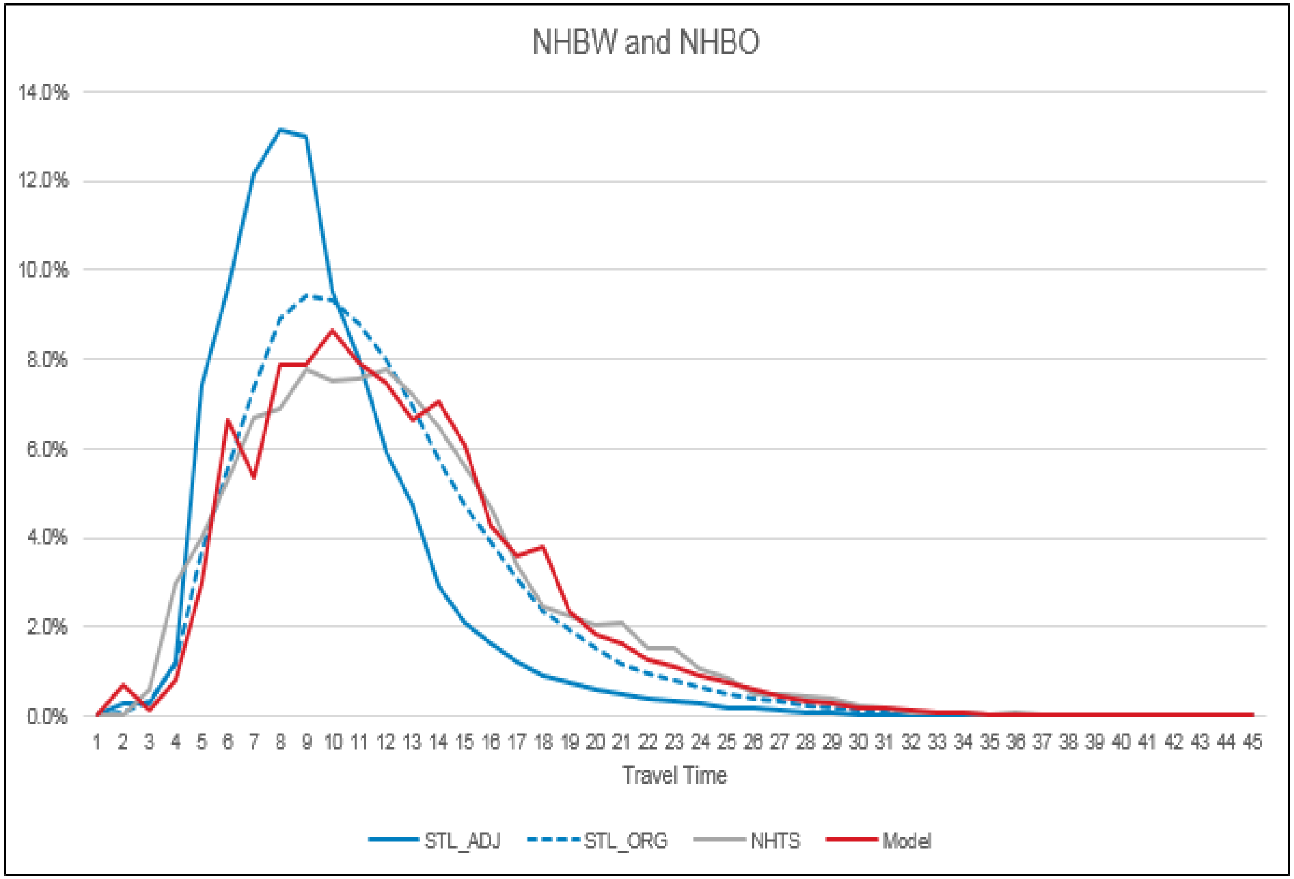
Aggregate HBO Distribution Validation



Source	Time
STL_ADJ	9.81
STL_ORG	12.25
NHTS	13.73
Modeled	14.27

Source	Intra Zonal
STL_ADJ	0.7%
STL_ORG	2.3%
NHTS	1.6%
Modeled	1.5%

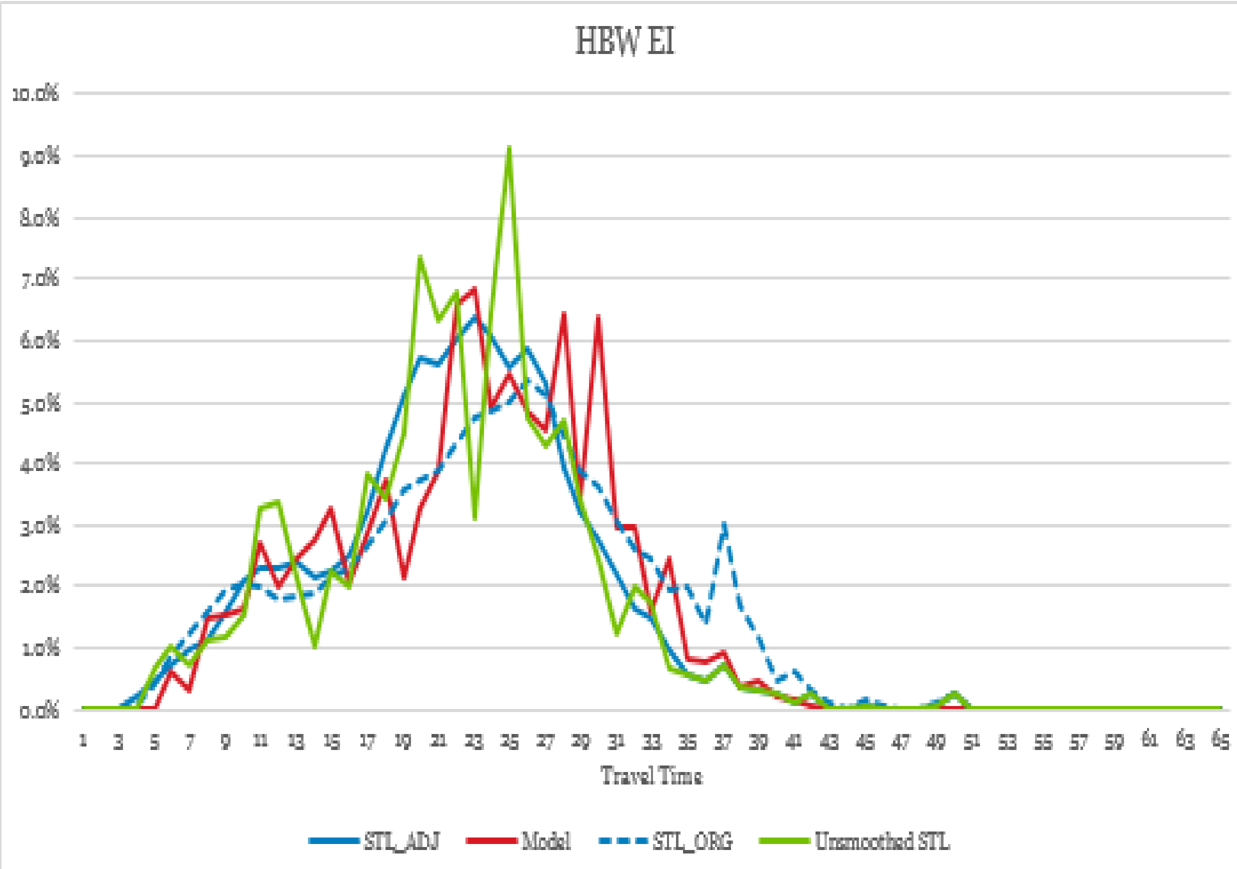
Aggregate NHB Distribution Validation



Source	Time
STL_ADJ	8.97
STL_ORG	11.14
NHTS	11.83
Modeled	11.97

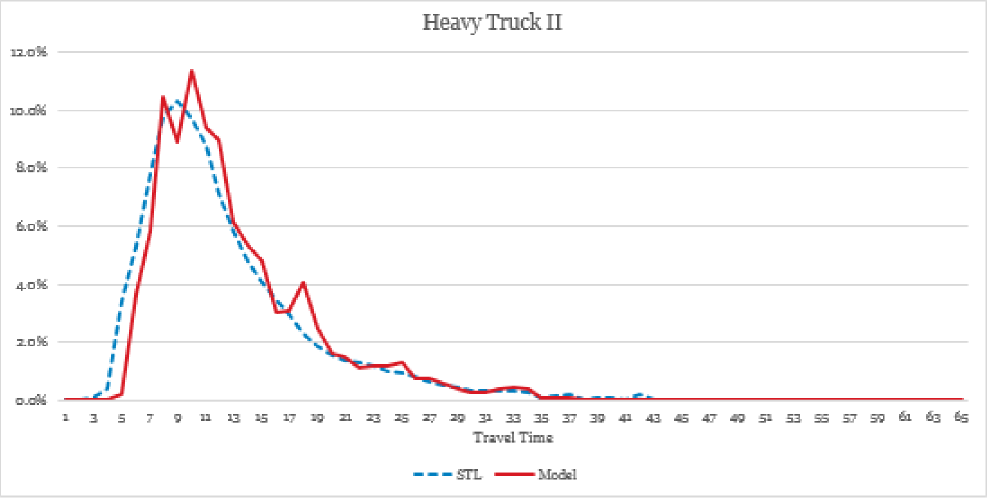
Source	Intra Zonal
STL_ADJ	0.9%
STL_ORG	3.1%
NHTS	6.0%
Modeled	6.1%

Example E-I Auto Distribution - HBW

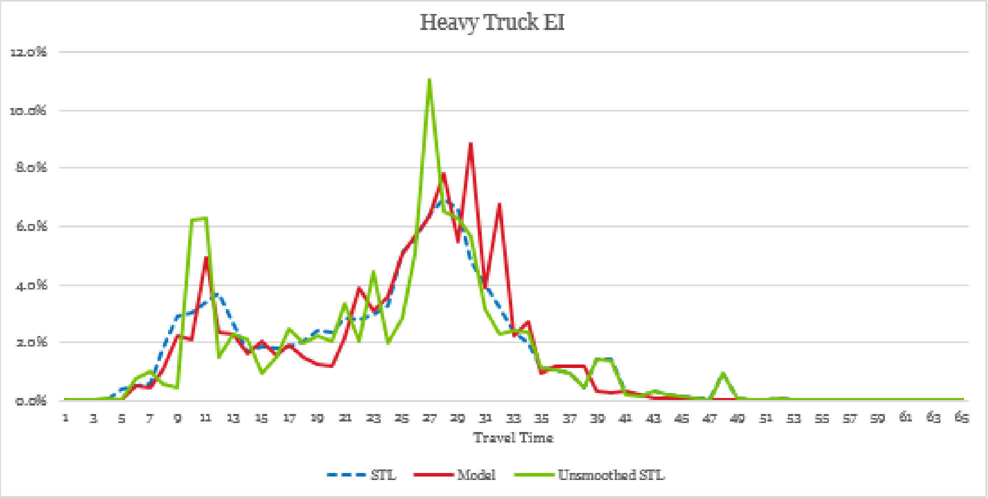


Source	Time
STL-ADJ	21.43
STL_ORG	23.38
Model	22.52

Example Truck Distribution - Heavy



Source	Time
STL	11.91
Modeled	12.59



Source	Time
STL	23.34
Modeled	23.96

Mid-Currituck Bridge Traffic & Revenue Study

Mid-Currituck Bridge Traffic & Revenue Study

Outer Banks, North Carolina

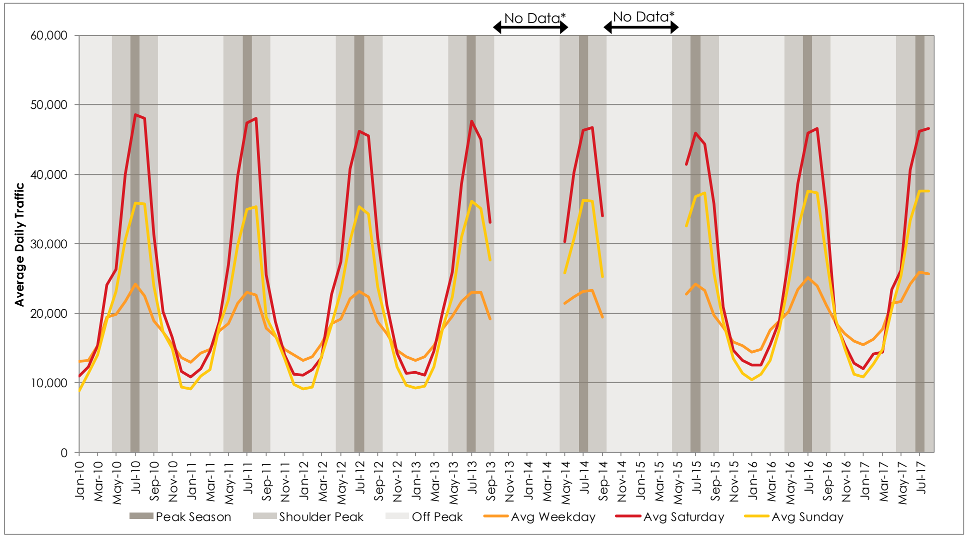
- Traffic variation highly seasonal
- One entry point at Wright Memorial Bridge
- 2+ hours of delay

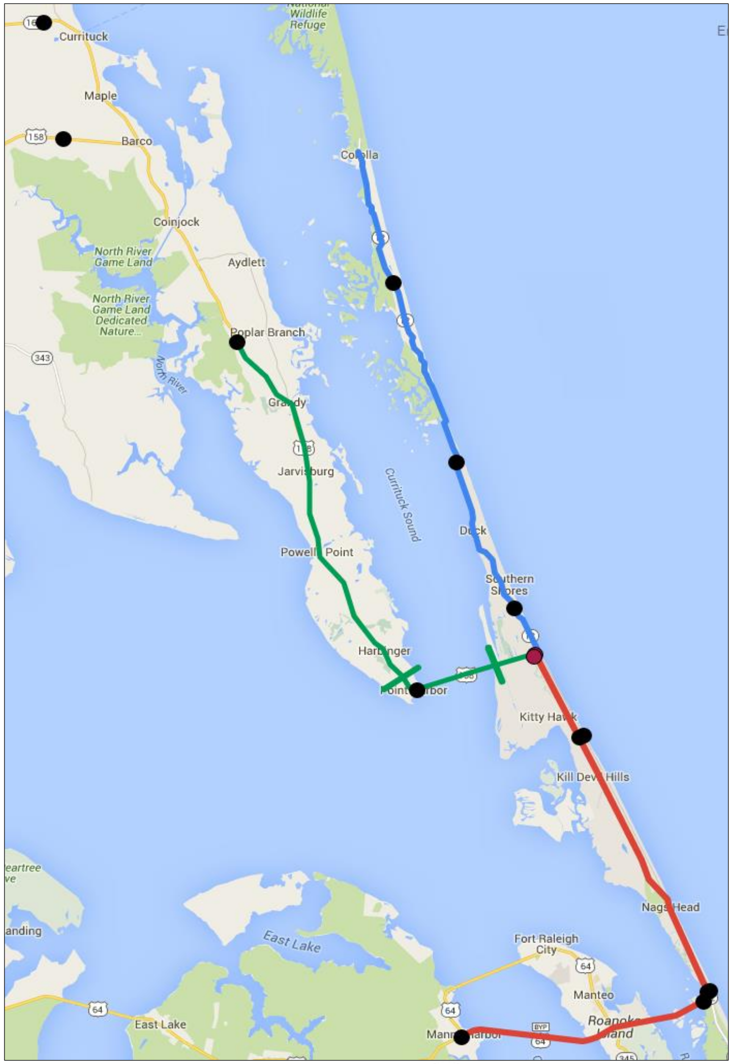
Stantec

Mid-Currituck Bridge Travel Times

Three Distinct Seasons:

Off-Peak	Winter	32 weeks
Shoulder Peak	Spring, Fall	12 weeks
Peak	Summer	8 weeks

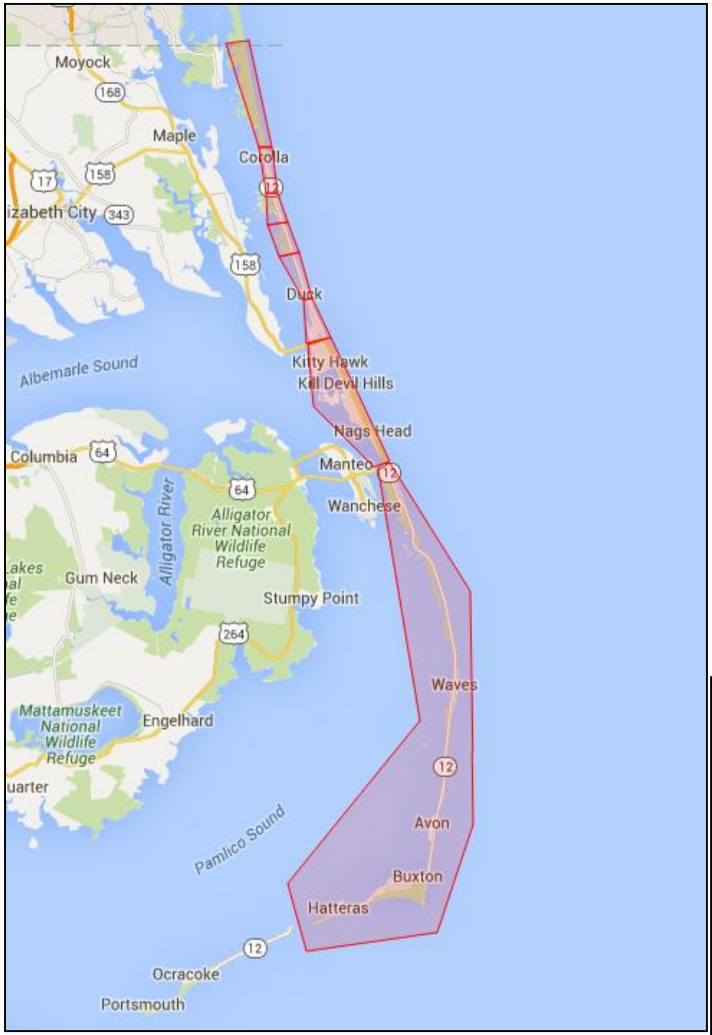


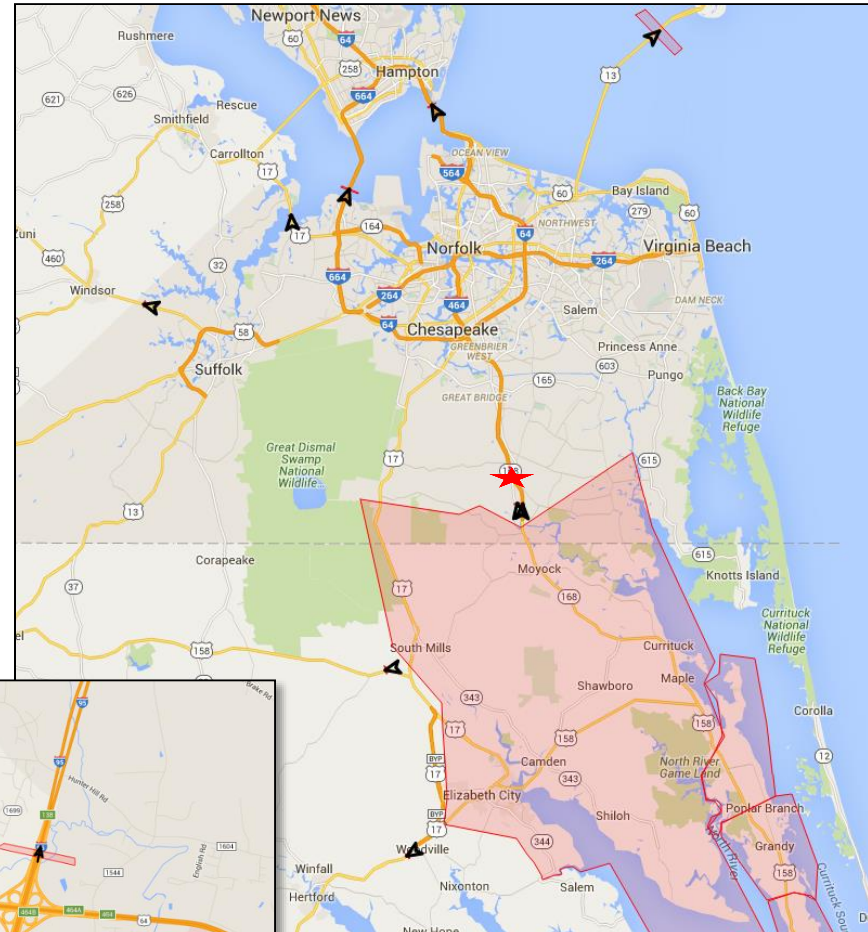


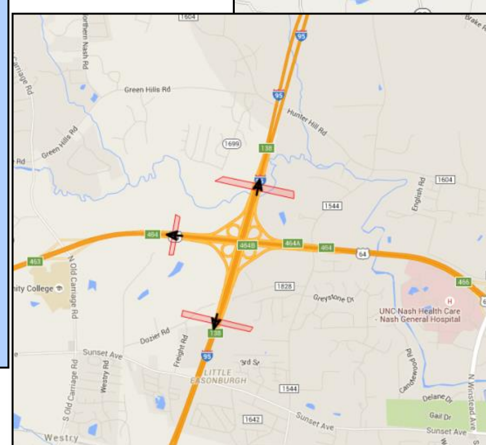
Speed Data Collection:

HERE data & independent travel time runs

StreetLight Origin-Destination Data







★ Chesapeake Expressway

Stantec

Early Version of Data

- Data Source
  - INRIX – 1% Sample
- User-Defined Zones and Selected Links
  - For selected links, some issues with capturing traffic
- Obtained Data by Season
  - Wednesday
  - Saturday
  - Sunday

Expansion Process

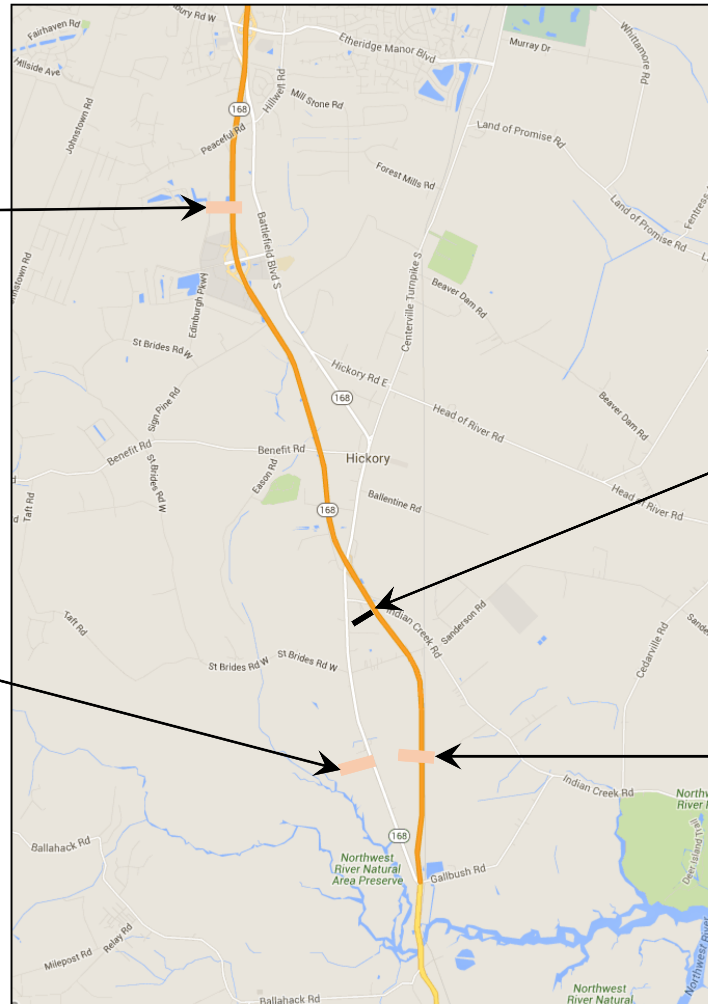
DESTINATION → / ORIGIN ↓		O1/O2 Corolla & North	O3/O4 Between Corolla & Duck	O5 Duck	O6 Southern Shores	O7 Kitty Hawk, Nags Head	O8 South of Baum Bridge	D1 Wright Bridge	D100 Roanoke Island & West	TOTAL
AM PEAK										
21, 22, 23, 24, 31, 32	NC north of Albemarle Sound	8%	8%	6%	5%	54%	3%	1%	3%	**88%**
3, 4, 33	Norfolk, Chesapeake, VA Beach*	0%	0%	0%	1%	6%	1%	--	1%	**10%**
6, 7, 11, 15, 16, 34	Long Distance Trips north of VA	0%	--	--	0%	1%	0%	--	0%	**2%**
Total		**8%**	**9%**	**6%**	**6%**	**61%**	**4%**	**1%**	**5%**	**100%**
MIDDAY										
21, 22, 23, 24, 31, 32	NC north of Albemarle Sound	3%	4%	4%	10%	57%	3%	1%	2%	**84%**
3, 4, 33	Norfolk, Chesapeake, VA Beach*	1%	0%	0%	0%	9%	1%	0%	1%	**13%**
6, 7, 11, 15, 16, 34	Long Distance Trips north of VA	--	--	--	0%	3%	0%	0%	--	**3%**
Total		**4%**	**4%**	**4%**	**10%**	**68%**	**4%**	**1%**	**3%**	**100%**
PM PEAK										
21, 22, 23, 24, 31, 32	NC north of Albemarle Sound	5%	4%	2%	2%	60%	1%	3%	5%	**82%**
3, 4, 33	Norfolk, Chesapeake, VA Beach*	1%	--	--	--	9%	2%	1%	2%	**16%**
6, 7, 11, 15, 16, 34	Long Distance Trips north of VA	--	--	--	--	2%	--	--	--	**2%**
Total		**5%**	**4%**	**2%**	**2%**	**72%**	**3%**	**4%**	**7%**	**100%**

Applied to traffic counts to produce ‘observed’ data used in model

Observed Toll Diversion Data

- Capturing Diversion Shares for Long Distance Travelers

Last SB exit before toll



Single mainline gantry

Un-tolled route

Toll route

OD Data Site

Stantec

Streetlight Data (Early Version)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Vehicle Type	Origin Zone Name	Origin Zone Is Pass-Through	Origin Zone Direction (degrees)	Destination Zone Name	Destination Zone Is Pass-Through	Destination Zone Direction (degrees)	Day Type	Day Part	O-D Traffic (StL Index)	Origin Zone Traffic (StL Index)	Destination Zone Traffic (StL Index)	Avg Trip Duration (sec)	Dest code	% By Origin	% By Destination
2	Personal	O1	no	N/A	O1	no	N/A	0: Average Day (M-Su)	0: All Day (12am-12am)	132	175	263	673	O1	75%	50%
3	Personal	O1	no	N/A	O1	no	N/A	0: Average Day (M-Su)	1: Early AM (12am-6am)	1	1	1	931	O1	100%	100%
4	Personal	O1	no	N/A	O1	no	N/A	0: Average Day (M-Su)	2: Peak AM (6am-10am)	12	21	20	600	O1	57%	60%
5	Personal	O1	no	N/A	O1	no	N/A	0: Average Day (M-Su)	3: Mid-Day (10am-3pm)	69	86	127	736	O1	80%	54%
6	Personal	O1	no	N/A	O1	no	N/A	0: Average Day (M-Su)	4: Peak PM (3pm-7pm)	40	53	79	610	O1	75%	51%
7	Personal	O1	no	N/A	O1	no	N/A	0: Average Day (M-Su)	5: Late PM (7pm-12am)	11	14	35	544	O1	79%	31%
8	Personal	O1	no	N/A	O1	no	N/A	1: Average Weekday (M-Th)	0: All Day (12am-12am)	129	176	236	633	O1	73%	55%
9	Personal	O1	no	N/A	O1	no	N/A	1: Average Weekday (M-Th)	1: Early AM (12am-6am)	2	2	2	931	O1	100%	100%
10	Personal	O1	no	N/A	O1	no	N/A	1: Average Weekday (M-Th)	2: Peak AM (6am-10am)	9	15	17	661	O1	60%	53%
11	Personal	O1	no	N/A	O1	no	N/A	1: Average Weekday (M-Th)	3: Mid-Day (10am-3pm)	69	92	114	622	O1	75%	61%

Issues & Limitations

- Early version
  - Required tedious & intricate analysis
  - Vehicle type but no inferred trip purpose
- Cannot obtain origins of very long trips
- '5 meters in 5 minutes' rule used to define trips
- Sample size created issues with expansion

Stantec

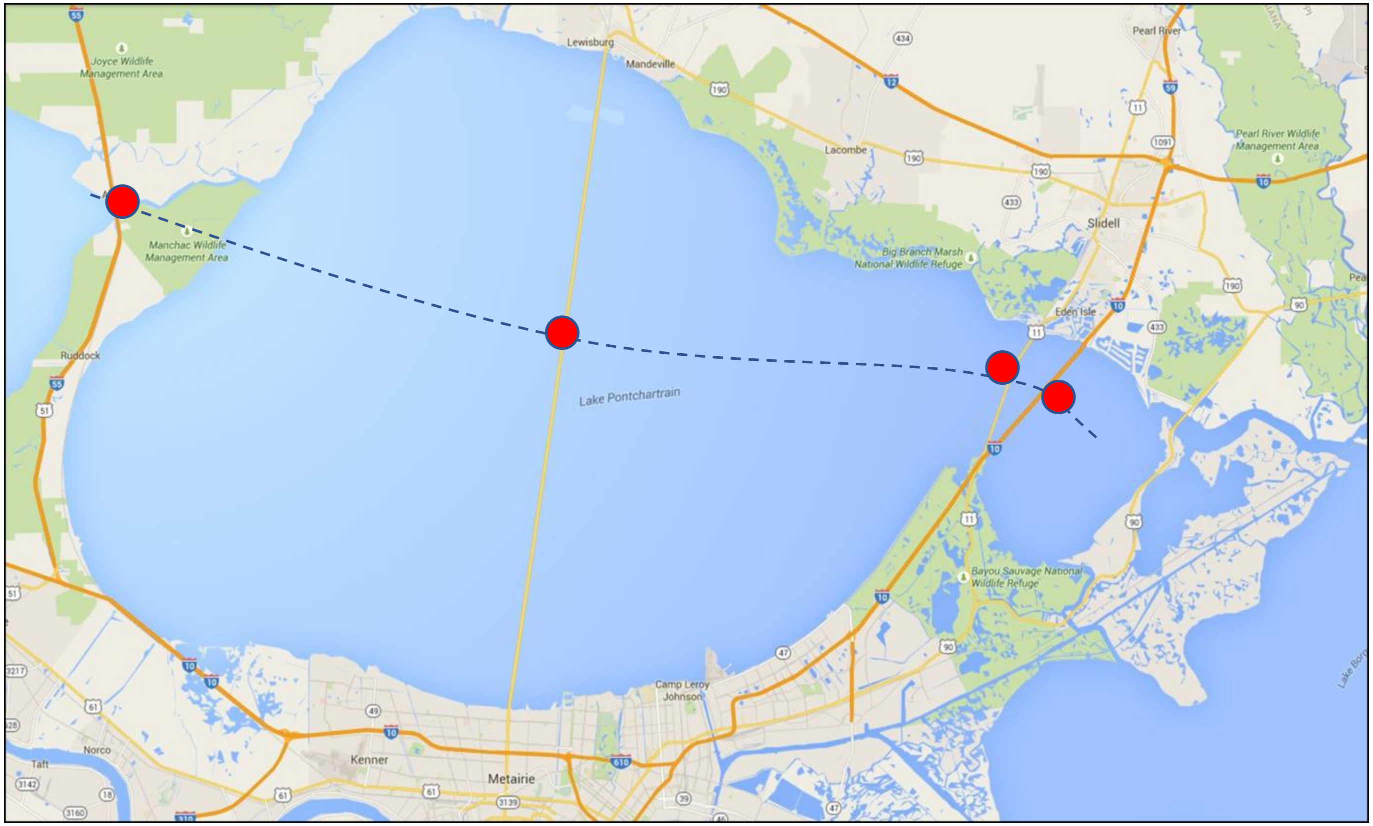
Lessons Learned

- Trip definition rules (5 meters in 5 minutes) can greatly affect data for longer trips encountering severe congestion, where delays may appear as separate trips
- Zone boundaries precise to avoid double counting
- Compare season progression first
  - Had to make many adjustments during post-processing

Lake Pontchartrain Causeway Traffic & Revenue Study

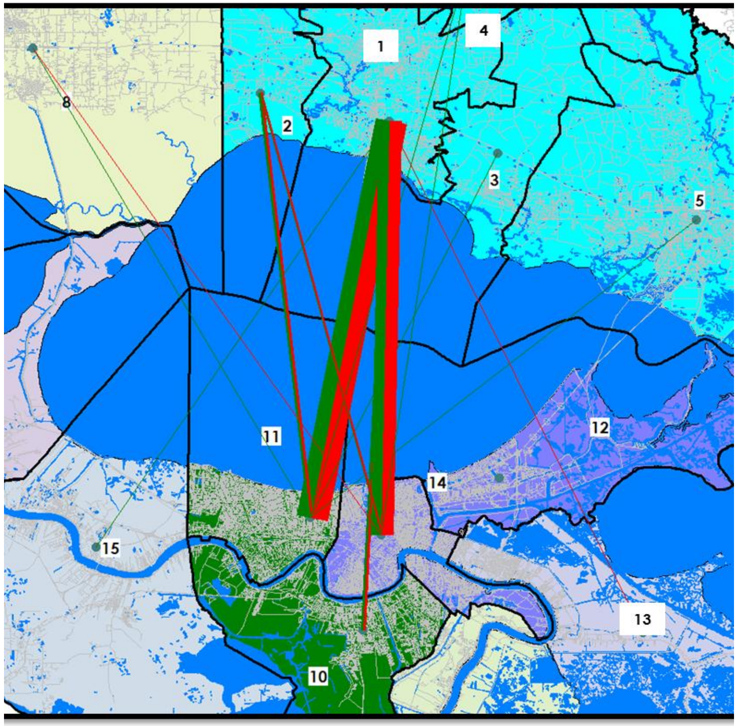
AirSage Data

- Origin-Destination Data for Crossing Links
  - Causeway & Competing Roadways (I-55, I-10, and US 90)
  - One Month of Data (April 2015)

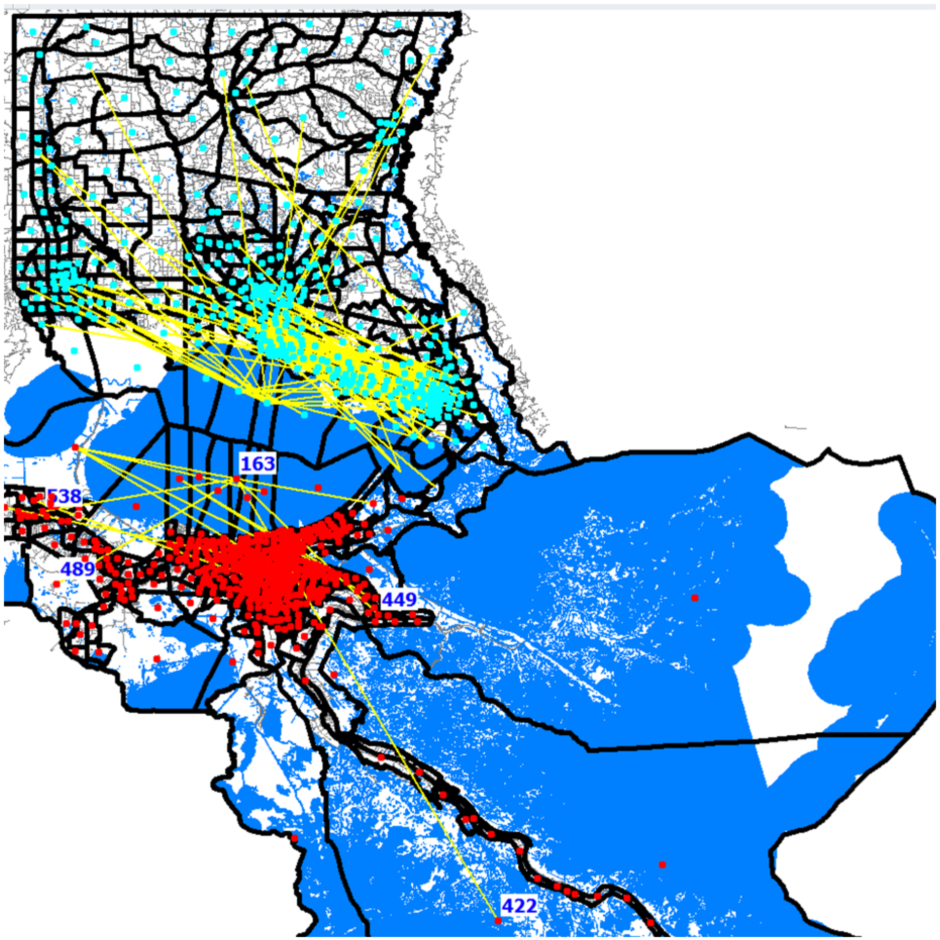


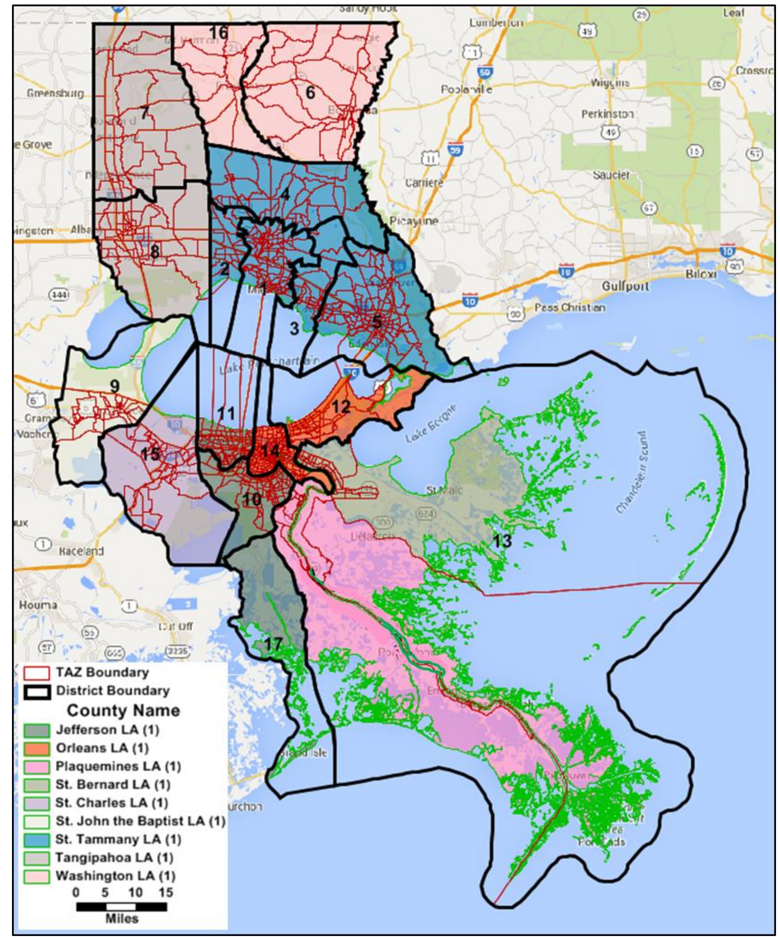
Data Characteristics

- Data Source
  - Cell Phone Signaling Data
- Type of Day
  - Average Weekday
  - Average Weekend Day
- Time Periods
  - AM Peak Period (6AM – 10 AM)
  - PM Peak Period (3PM – 7PM)
  - Daily
- Inferred Trip Purposes
  - HBW
  - HBO
  - NHB
- Data expanded with Traffic Counts

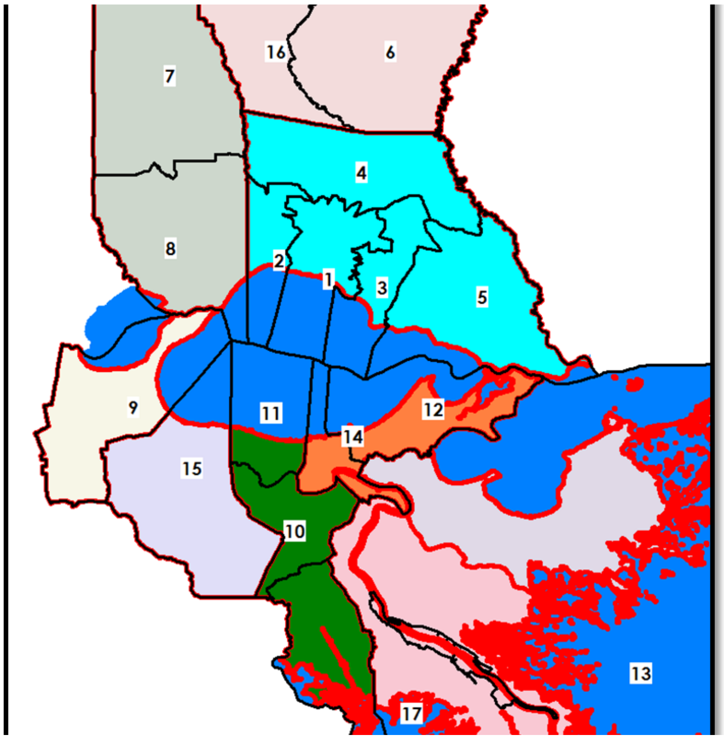


Data assessment revealed some unreasonable trip patterns





Data Usage – Verification of Market & Share



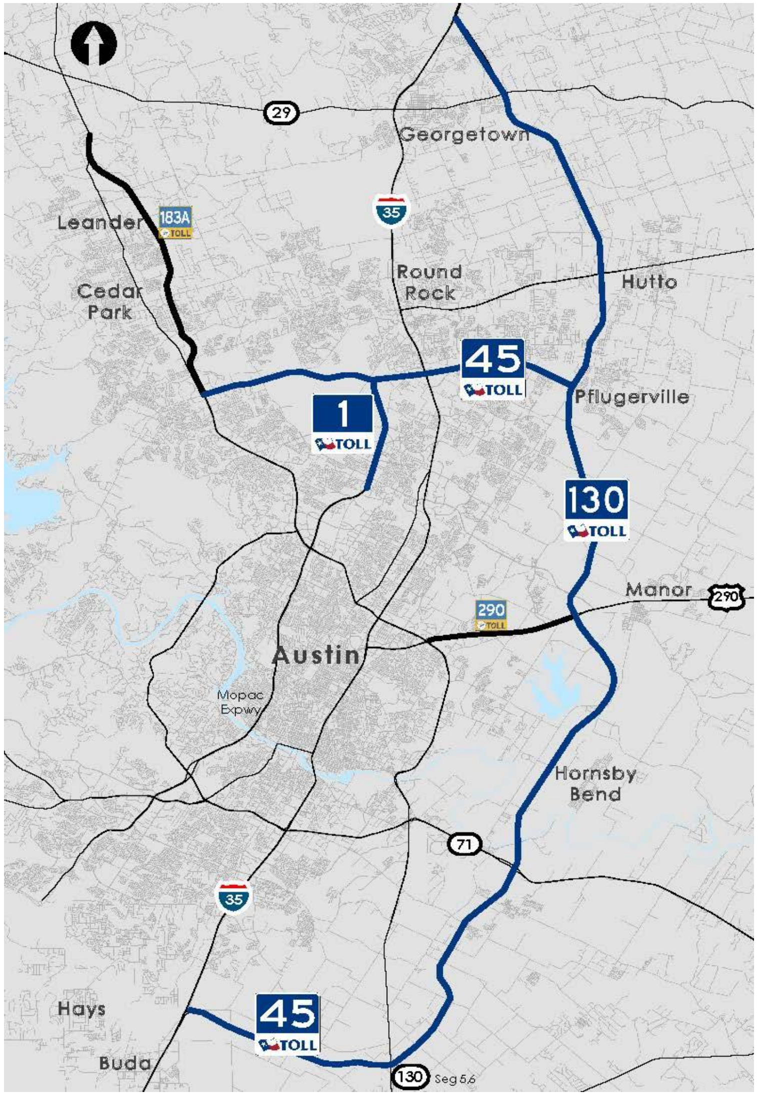
Markets using Causeway

OD Flow Between Districts			Pct Trips
11	and	1	37.6%
14	and	1	26.5%
11	and	2	6.0%
10	and	1	4.7%
14	and	2	3.6%
11	and	4	2.3%
15	and	1	2.3%
11	and	3	2.1%
11	and	8	1.9%
11	and	5	1.8%
Total			88.7%

Causeway Share of Key Markets

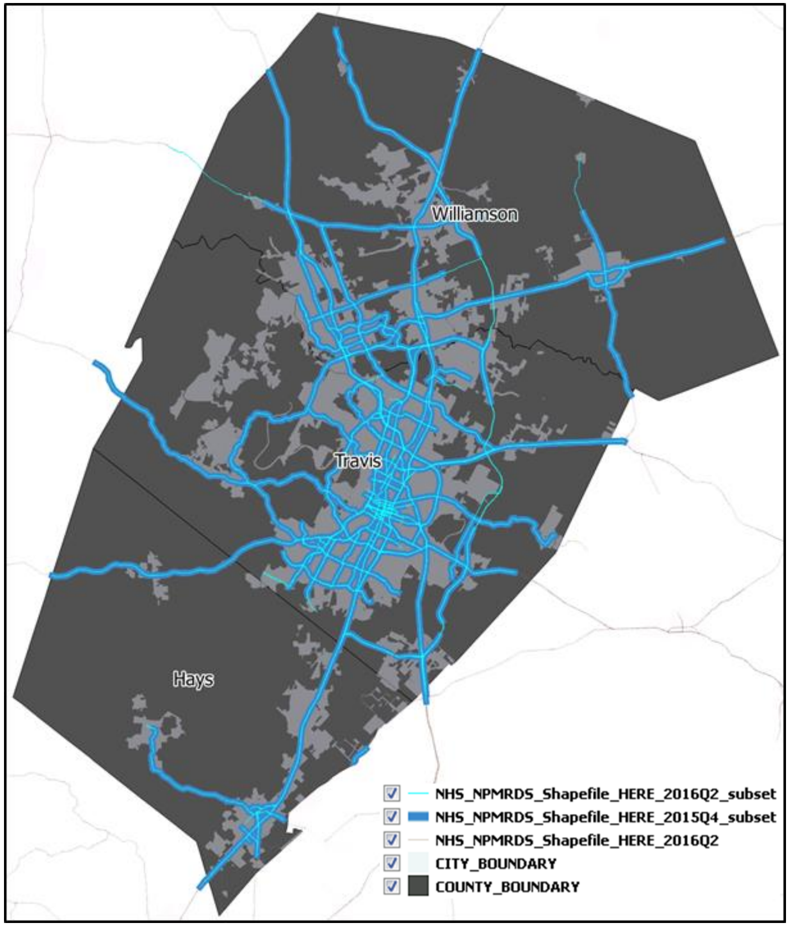
BRIDGE CROSSING DISTRIBUTION	
BRIDGE	% DIST
I-55	0.7%
Causeway	97.6%
US 11 / I-10	1.7%
Total	**100.0%**

Central Texas Turnpike System
2018 Traffic & Revenue Study



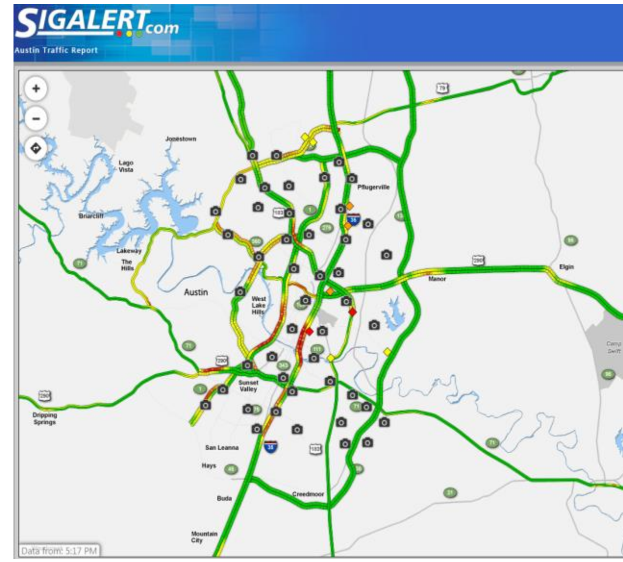
HERE Speed Data

- Provided as shape file, collection station information and travel time data
- **Features**
  - Collects data every 5 mins ,288 collection points per day
  - Data points include major highways and some local roads
- **Lessons Learned**
  - Provided shapefile does not align well with network, lots of manual work needed
  - Some link data appeared illogical and hard to locate
  - Temporary congestion or traffic signals could impact the overall average speed values



SigAlert Speed Data

- **Features**
  - Real-time traffic map
  - 24/7 speed/accident/construction coverage
  - Average lag time 3-5 minutes
- **Data Limitations for Modeling**
  - Data coverage on main highways only
  - Sparse data collection points
- **Lessons Learned**
  - Good open-source reference for general verification
  - Data are often too general for subtle speed variations study

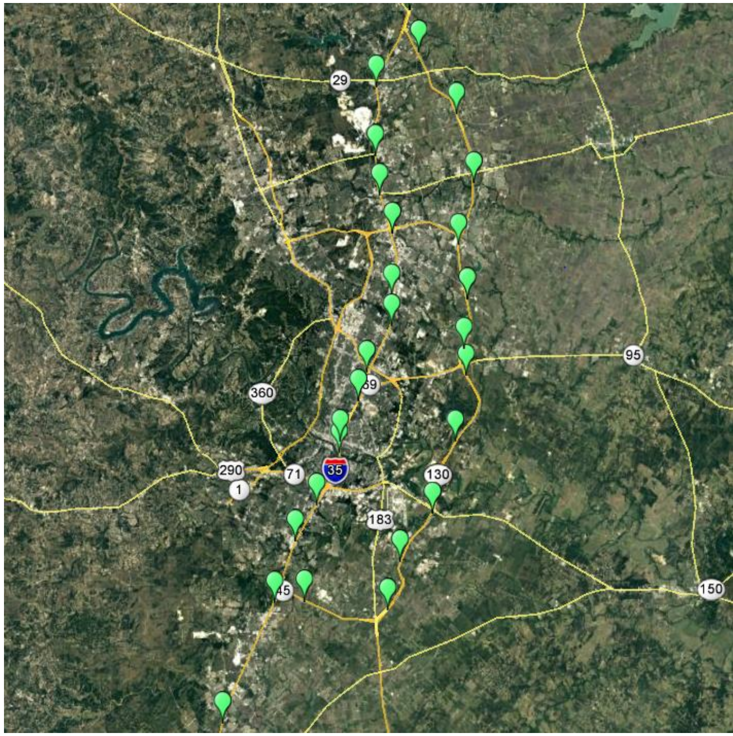


35 (Pan Am Expy) North North	
Location	MPH
Holly St	11
Riverside Dr	11
S I- 35 Svc Rd	25
Oltorf St	65
Ramp from US-290/TX-71	66
Teri Rd	66
US-290 West / TX-71 Johnson City Bastrop	66
S I-35 Service NB	66
Wm Cannon Rd	68
S I-35 Service NB	68
I-35 Frontage Rd	68
S I-35 Service NB	68
Slaughter Ln	69

Source: Sigalert.com

O-D Patterns (Bluetooth Data)

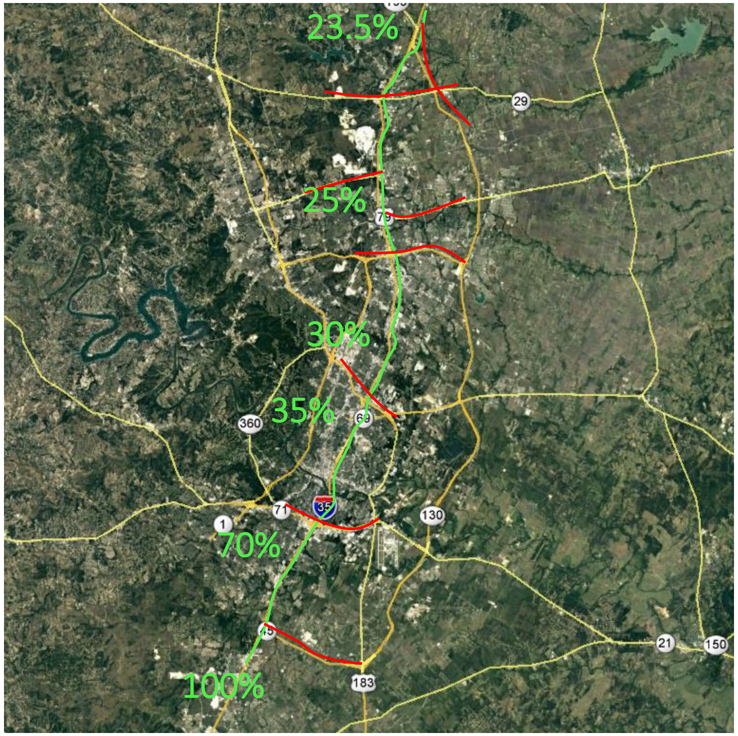
- Bluetooth instruments along major roadways
- Less points in a larger area
- **Features**
  - 24/7 data collection
  - Traces long-distance travel
- **Limitations**
  - Vehicle classification not available
  - Could miss some data points, trip may be split/merged
  - Inadvertent data capture processing



Bluetooth Data Usage

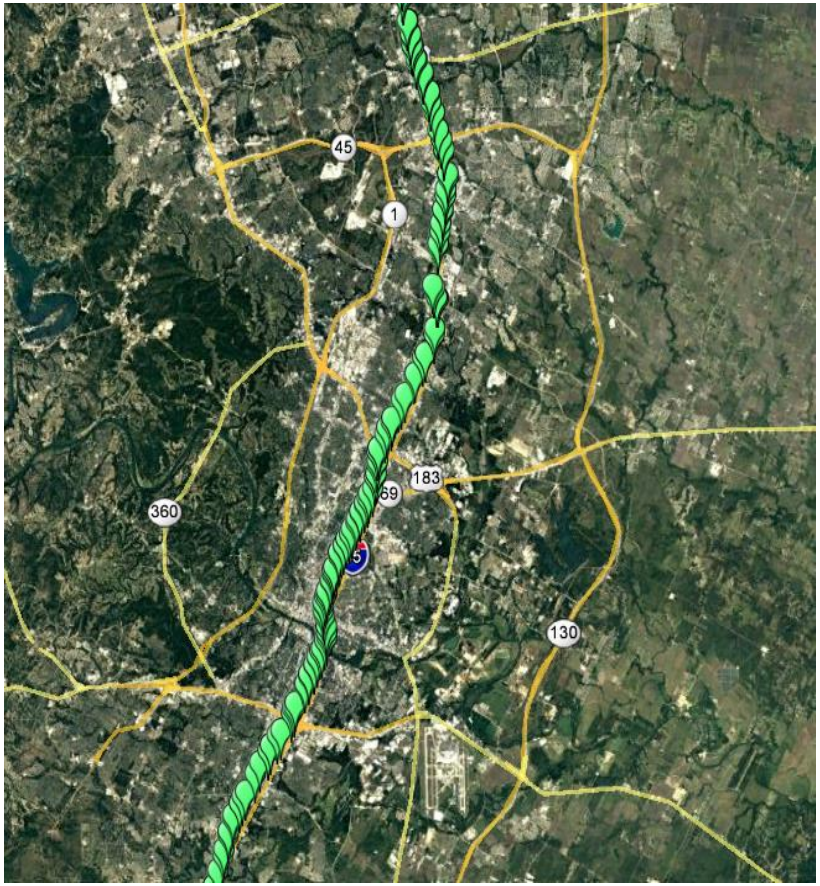
- Percentage of long-distance trips at each collection point by route
- Gives pattern of long-distance O-D between collection points

NB TRAFFIC (From IH-35 Buda)			
IH-35			
Destination Location	Obs.	Estimated	
	%Total	Trips	% Total
IH-35_SlaughterLn	86.8%	71,677	95.0%
IH-35_StassneyLn	73.3%	52,786	70.0%
IH-35_Riverside	55.3%	33,922	45.0%
IH-35_5thSt	53.1%	32,632	43.3%
IH-35_AirportBlvd	47.9%	26,286	34.8%
IH-35_US-183/US-290	43.5%	24,754	32.8%
IH-35_Braker	39.7%	22,937	30.4%
IH-35_Parmer	38.8%	22,606	30.0%
IH-35_SH-45Toll	36.7%	20,856	27.6%
IH-35_US-79	35.4%	18,844	25.0%
IH-35_FM-1431-RoundRock	32.3%	18,657	24.7%
IH-35_SH-29-Georgetown	30.4%	18,557	24.6%
IH-35_Georgetown	23.5%	17,634	23.4%

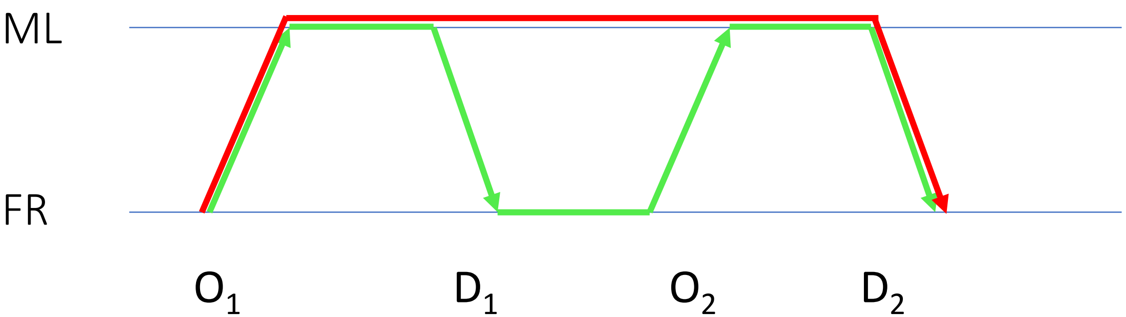


Skycomp Data

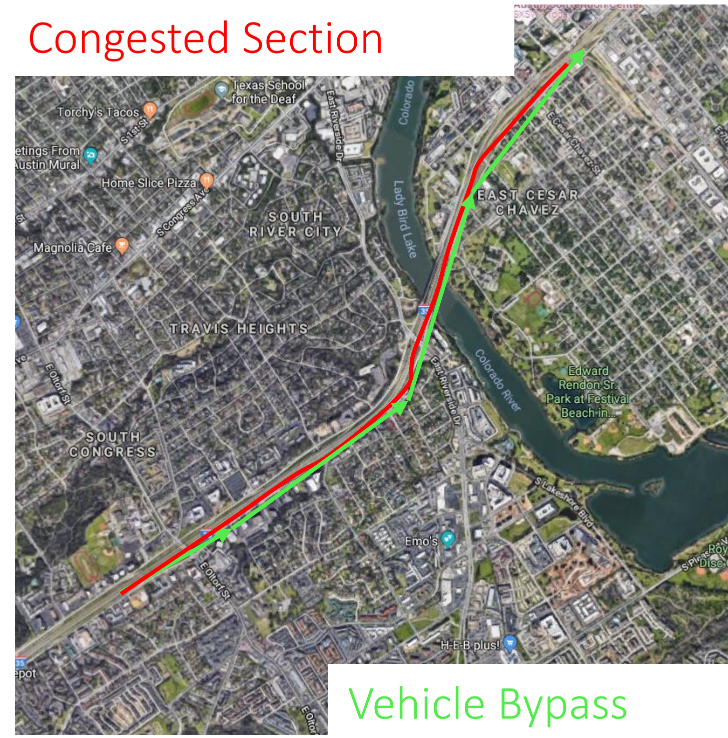
- INRIX GPS data
- **Features**
  - Tracks vehicle trajectory
  - Corridor entrance-exit pattern between I-35 & frontage roads
  - Traces bypass behavior
  - Extract true O-D for trips
- **Lessons Learned/Data Limitation**
  - Sample size very small
    - Some movements have minimal observations a month
  - Early samples biased towards trucks
    - More than 70% of samples are truck



Skycomp Data - True O-D vs. Observed O-D



- 'True O-D' is the actual O-D of a trip. Provides the first entrance and the last exit a vehicle take during one trip
- In this case, 2 O-Ds ($O_1 \rightarrow D_1$, $O_2 \rightarrow D_2$) observed, but true O-D is $O_1 \rightarrow D_2$
- 1 or more bypass movements can be made for each trip. Tracing complete trajectory generates true O-D



"Bypass" Behavior: Vehicles take frontage road for a short distance and then go back to mainline to avoid congested sections of highway

Off the Record….

Summarize Transaction and Trip Databases

Data:

200,000+ daily and 52+million annual records of transactions on the express lane facility. The input data was provided in individual daily excel files by direction. Each record has fields for time stamp, type of transaction, plaza location identifier, confidence of detection and corresponding toll.

Process:

Chaining transactions into trips to determine entry exit patterns, and full trip tolls.

Screening for low detection confidence, incomplete trip records, duplicate trips etc.

Product:

Generate weekly, monthly, yearly summaries of tolls and trips by Origin and Destination, and use this information to predict toll rates and expected utilization by time of day.

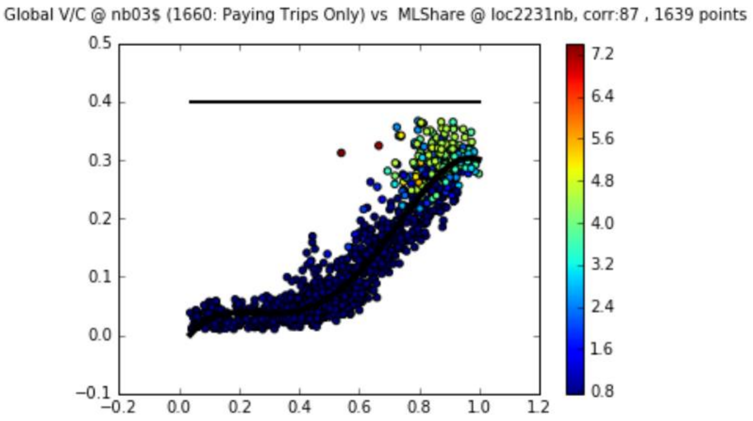
Software:

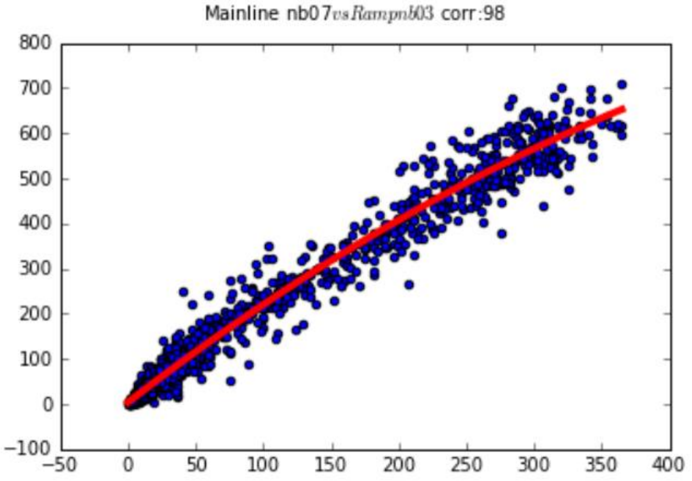
VBA in Excel and python pandas to clean up the input data files and re-format fields

PostgreSQL to create an aggregate database for all transactions

Develop Predictive Analytics

- Used python scripts to aggregate transaction and traffic into thousands of 15-minute periods over a several-month period.
- Identified trends between overall congestion indicators (global v/c) and the percent of traffic using the tolled express lanes (mainline market share)
- Apply this curve to determine future demand and toll rates in 15-minute increments based on projected traffic volumes from the travel demand model.
- The process also included an application of the actual toll algorithm simulator to forecast future tolls based on speed and flow rate metrics





Toll Algorithm Simulator

<u>Data:</u>

- Volume, Occupancy, and Speed data in 20 second increments for at least 2 weeks (~15.25 Million data points)
- All the above data from both Express Lanes and General Purpose Lanes
- Loop Assignment paths as well as a set of fuzzy parameters that control the toll calculation per O-D

<u>Process:</u>

Dynamic Toll is calculated based on fuzzy logic algorithm, a mathematical technique for handling data with many-valued logic solutions.

<u>Product:</u>

Toll Rates in 20 second intervals for each O-D pattern on the proposed facility

<u>Software:</u>

R-statistical programing language used to develop routine

Final routine implemented in CUBE