A Suite of Model Updating and Validation Procedures Using Third Party Origin-Destination Data

Innovations in Travel Modeling Conference
Session 5A
June 27, 2018

By
Gregory Giaimo
Ohio Department of Transportation

Background

- ODOT's Traffic Operations Center has purchased INRIX speed data for years, planning/modeling has benefited from their interest
- The ODOT Modeling & Forecasting Section proceeded cautiously for the past 5 years funding research through Ohio State University on using Origin-Destination data from third party vendors
- ODOT's Roadway Engineering Office became interested in Origin-Destination data to estimate weaves and feed OD based traffic microsimulation and purchased "full access" to such data from INRIX/StreetLight, again planning/modeling benefits with sudden access to mounds of data
- Lesson: planning people move a lot slower than operations/design people

Background

- While this is excellent data, it isn’t perfect
  - Ironically the “big” data is “small” sample sizes
  - Not random samples either
- Must be used carefully to avoid misuse
- ODOT developed a set of procedures to extract and analyze third party data
- Have also updated model validation procedures to incorporate

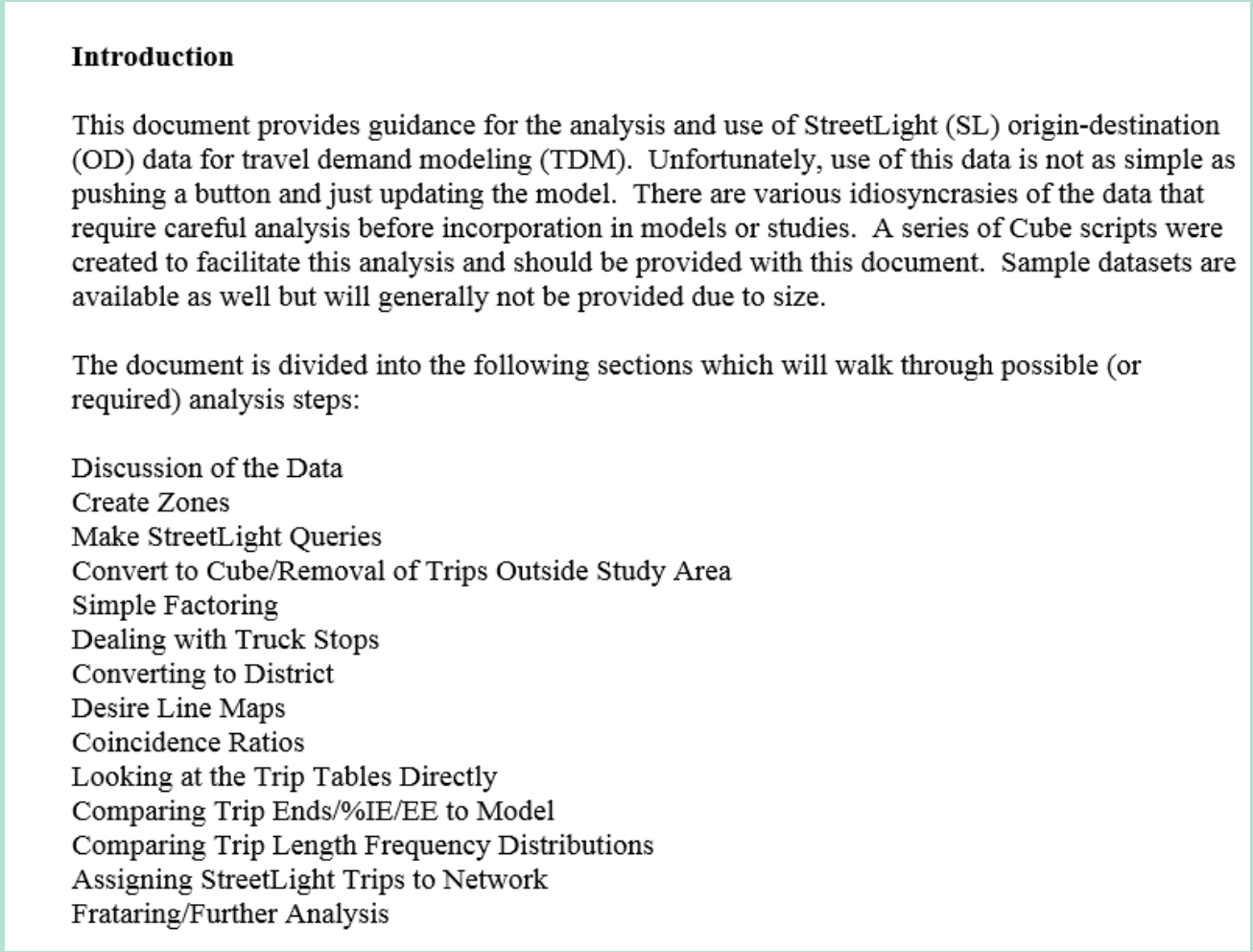
Data Analysis Procedures

- ODOT's "full access" license gives us, our consultants and other public agencies access, so standardized procedures made sense

- 40 page manual available on request

- Covers common processing/data conversion steps but also creating:

- Trip end summaries
- Trip length frequency distributions
- Coincidence ratios
- Assignment of raw trips to network
- Factoring



Introduction

This document provides guidance for the analysis and use of StreetLight (SL) origin-destination (OD) data for travel demand modeling (TDM). Unfortunately, use of this data is not as simple as pushing a button and just updating the model. There are various idiosyncrasies of the data that require careful analysis before incorporation in models or studies. A series of Cube scripts were created to facilitate this analysis and should be provided with this document. Sample datasets are available as well but will generally not be provided due to size.

The document is divided into the following sections which will walk through possible (or required) analysis steps:

Discussion of the Data
Create Zones
Make StreetLight Queries
Convert to Cube/Removal of Trips Outside Study Area
Simple Factoring
Dealing with Truck Stops
Converting to District
Desire Line Maps
Coincidence Ratios
Looking at the Trip Tables Directly
Comparing Trip Ends/%IE/EE to Model
Comparing Trip Length Frequency Distributions
Assigning StreetLight Trips to Network
Frataring/Further Analysis

Data Analysis Procedures

- 14 scripted procedures
- Besides facilitating analyses mentioned on previous slides, 3 common needs:

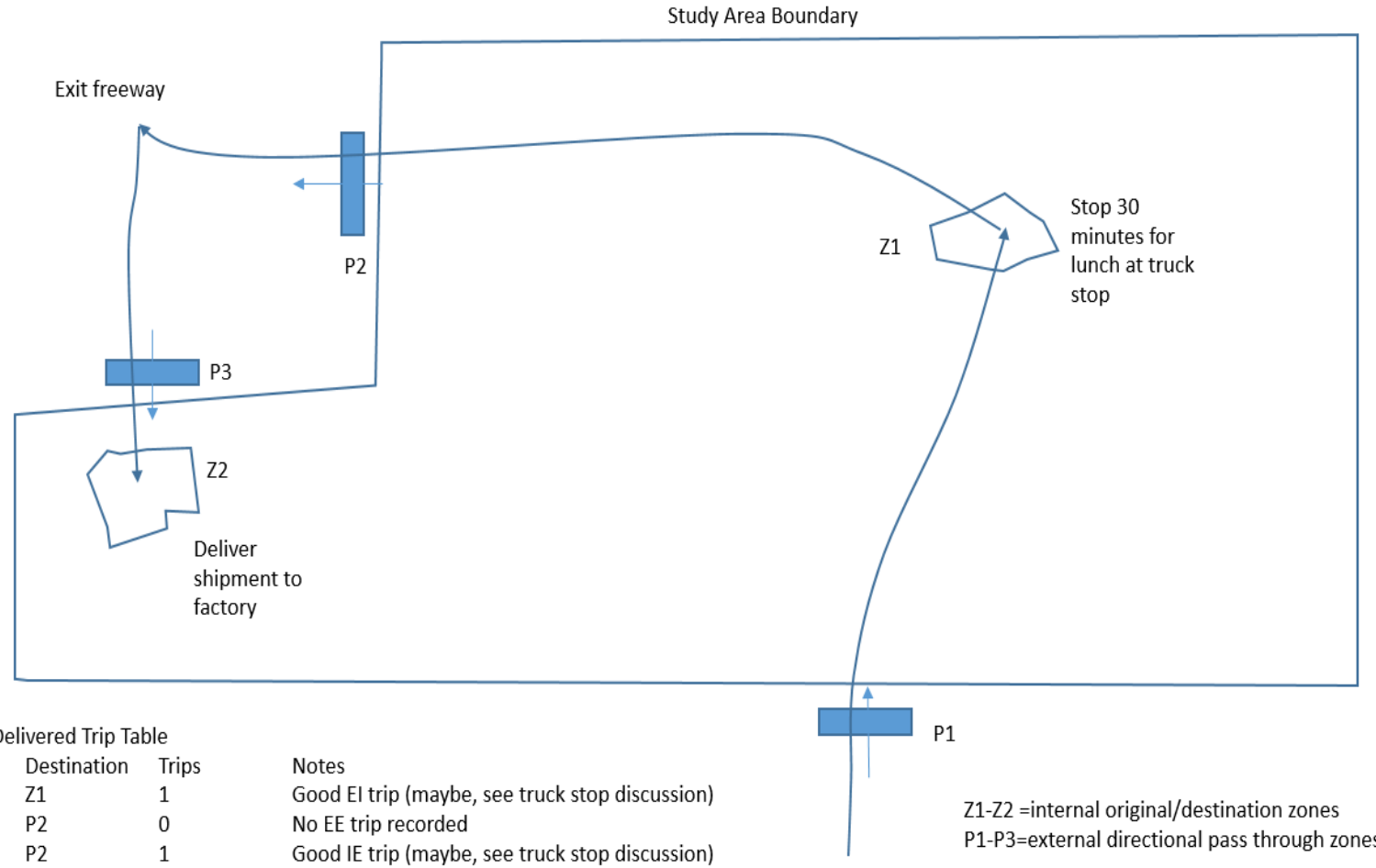
1. Conversion between TAZ and District, for many uses the data is simply too course to use at TAZ level
2. Removal of incidental stop locations such as rest areas and truck stops
3. Separation of Internal and External portions of trip matrices as those 2 domains have greatly different characteristics in the data and utility in the models

Summary of Provided Cube Processing Scripts

asciiin_GPS.s	Convert SL GPS data car/truck trip tables to Cube and remove trips study area
asciiin_LBSv2.s	Convert SL LBS data trip table and optional purpose trip tables to Cube and remove trips study area
asciiout.s	Convert Cube trip table to ASCII
coincidence_ratio.s	Calculate coincidence ratio between 2 trip tables
compareSL2OMSmodel.s	Create separate internal and external trip end summary comparisons betweem OMS model and SL including segmentation of optional special zones.
factor_StreetLight.s	Factor SL trip table to match totals trips in another (model) trip table
linkattributetonode.s	Copy link District code to nodes of Cube network
remove_truckstop_allzones.s	Move trips to/from specified zones to all other zones proportional to the other trips ends at those zones
remove_truckstops_COMPLEXMETHOD.S	Move trips from specified zones based on detailed user input of where to move them.
renumbertodist.s	Compress a TAZ trip table to districts
replace_omstrips_w_Sl.s	Factor OMS TOD Trip Tables to match cell by cell daily trip totals in SL
SL_tlfd.s	Create standard Cube TLFD report from 2 trip tables in a matrix file and 1 impedance matrix in a skim file
SL_tlfd_by_purp4OMS.s	Create standard Cube TLFD report from an OMS purpose specific trip table and SL LBS purpose specific trip table and an impedance matrix
splitext_int.S	Separate II from External portions of trip table into 2 separate Cube matrix files

Understanding the Data

- Concave model boundaries (or other viable routing options that exit and re-enter the area) are a real pain when extracting this data



Streetlight Delivered Trip Table

Origin	Destination	Trips	Notes
P1	Z1	1	Good EI trip (maybe, see truck stop discussion)
P1	P2	0	No EE trip recorded
Z1	P2	1	Good IE trip (maybe, see truck stop discussion)
Z1	P3	1	Bad IE trip, but easy to filter out if externals directional
Z1	Z2	1	Double counted II trip, need middle filter to ID
P2	P3	1	Bad EE trip, but easy to filter out if externals directional
P2	Z2	1	Bad EI trip, but easy to filter out if externals directional
P3	Z2	1	Good EI trip

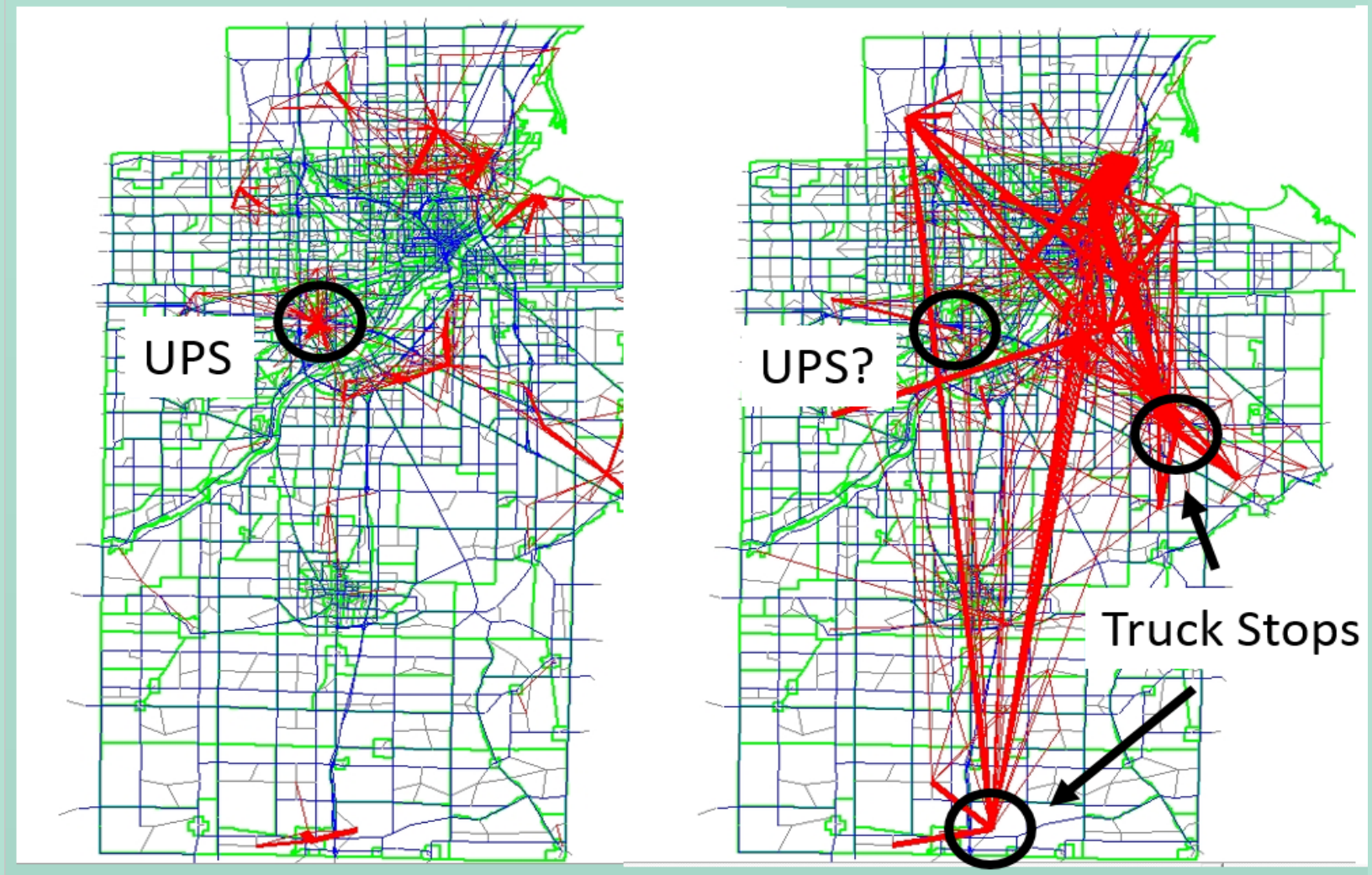
Understanding the Data

- You have to decide how to treat incidental stops
- They make a particularly huge difference for trucks

Desire Lines at TAZ Level (Truck Trips)

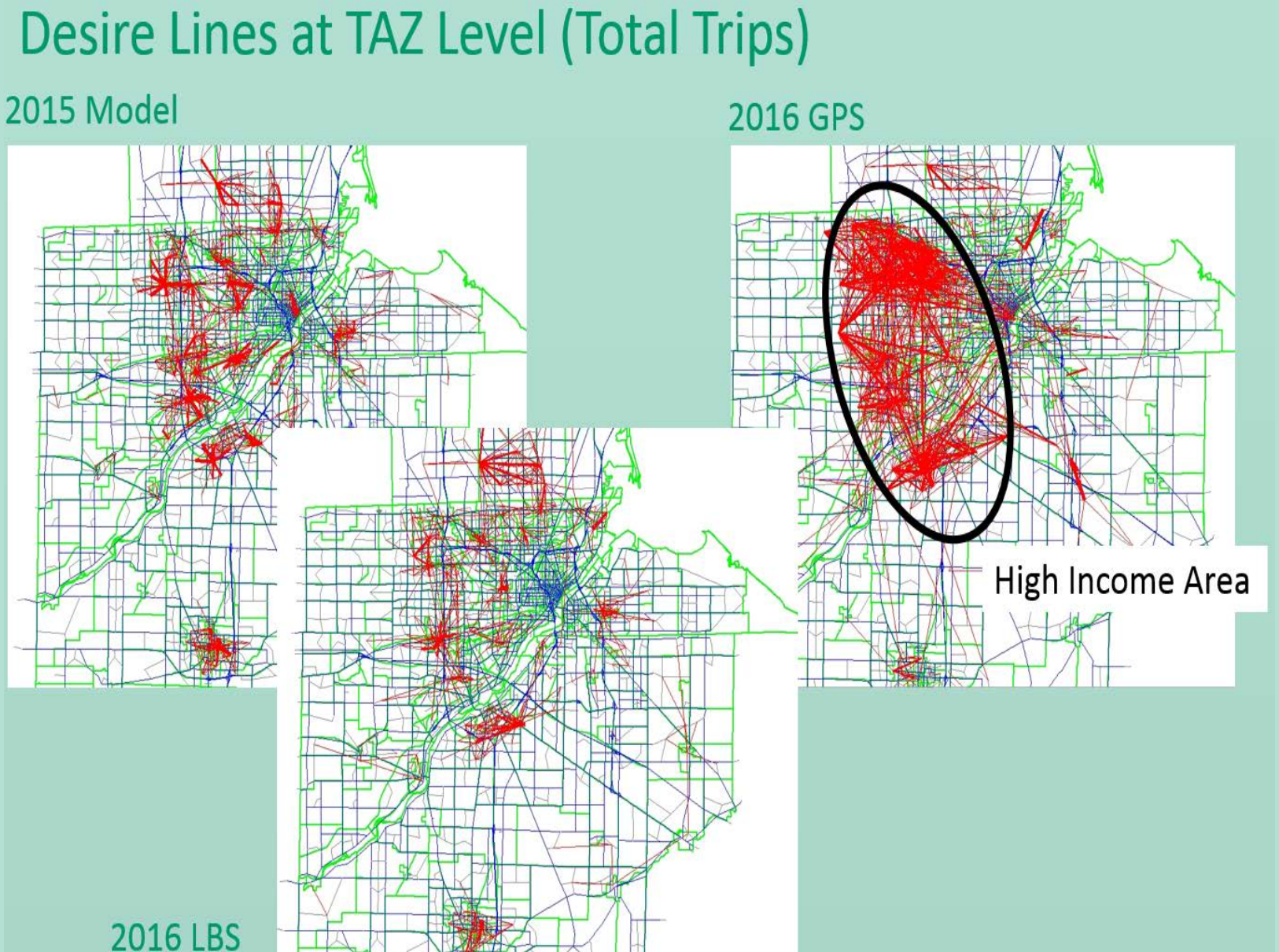
2015 Model

2016 GPS



Understanding the Data

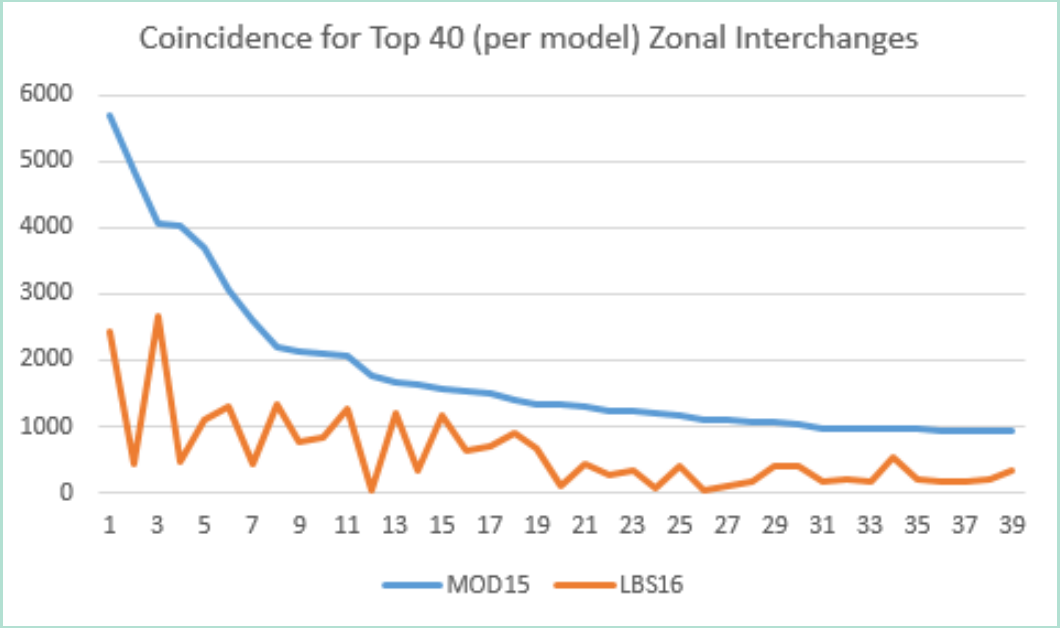
- There can be income biases in the car data because the data is captured from consumer devices



Understanding the Data

- Coincidence ratios found to be an excellent way to judge the similarity of OD matrices

- Different sources of 3rd Party data are no more similar to one another than to the model (and contain far more error than the changes we attempt to model) so there is no conclusive benefit to swapping in these observed values for model generated ones at TAZ level (except externals)



Trip Table TAZ Level Coincidence Ratios							
Test 1	2010 Mod	2010 Mod	2015 Mod	2015 Mod	2015 Mod	2016 GPS	2016 GPS
Test 2	2015 Mod	2045 Mod	2016 GPS	14-17 GPS	2016 LBS	14-17 GPS	2016 LBS
Car	0.96	0.85	0.28	0.29	na	0.63	na
Truck	0.97	0.87	0.24	0.24	na	0.63	na
Total	0.96	0.86	0.30	0.30	0.38	0.64	0.32

Understanding the Data

- However, you can swap this data in and assign it to your model networks and see what happens
- Some trip length bias in the 3rd Party data will result in higher VMT's but can also point out model problems (in this case model was generating too many short trips)

VMT Summary						
FACTYPE	Criteria	2010 Final	2015 Mode	GPS16	GPS1417	LBS16
Freeway	0.93-1.07	0.98	1.01	1.19	1.26	1.15
Expressway		0.90	1.00	1.08	1.13	1.06
Ramp		1.06	0.98	1.19	1.31	1.23
Arterial	0.90-1.10	1.01	1.01	1.14	1.22	1.25
Collector	0.85-1.15	0.98	0.97	1.13	1.23	1.23
Rural	0.85-1.15	0.97	0.97	1.13	1.24	1.30
AREATYPE						
Rural		0.97	0.99	1.02	1.10	1.07
Suburban		0.97	0.99	1.28	1.37	1.24
Urban		1.02	1.03	1.13	1.23	1.32
CBD		0.98	0.98	1.20	1.31	1.28
Outlying BD		0.98	0.96	1.13	1.17	1.14
Total		0.98	1.00	1.15	1.24	1.20

%RMSE Total				
2010 final	2015 mode	GPS16	GPS1417	LBS16
116.36	140.79	200.3	224.02	189.85
81.15	91.86	132.44	141.48	127.16
48.98	49.59	90.63	96.63	70.48
38.23	39.05	72.02	76.61	66.26
34.52	32.93	70.45	74.38	51.97
30.75	32.67	76.52	79.04	46.86
33.72	36.04	70.14	73.71	56.88
28.75	38.42	64.44	64.5	47.43
26.42	27.18	56.39	59.51	47.97
25.35	27.69	65.47	67.98	48.57
20.27	21.7	41.18	44.66	32.13
16.64	18.66	39.81	45.49	41.57
21.2	18.88	33.51	37.53	25.26
8.25	29.16	36.51	38.06	31.29
8.42	8.91	31.3	39.73	28.67
12.03	16.84	35.41	43.32	35.11
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
37.9	43.81	86.53	94.87	72.46

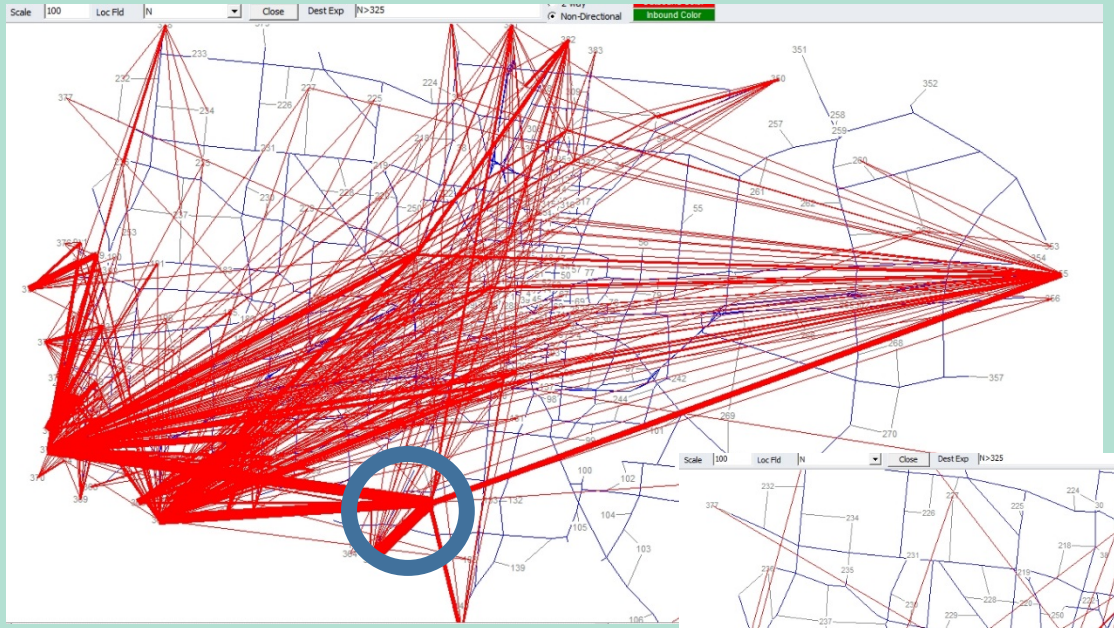
Updated Model Validation Procedures

- Special Generators
- Trip Length Frequency Distributions
- %IE/%EE Cordon Volume Splits
- Seed EE Trip Table

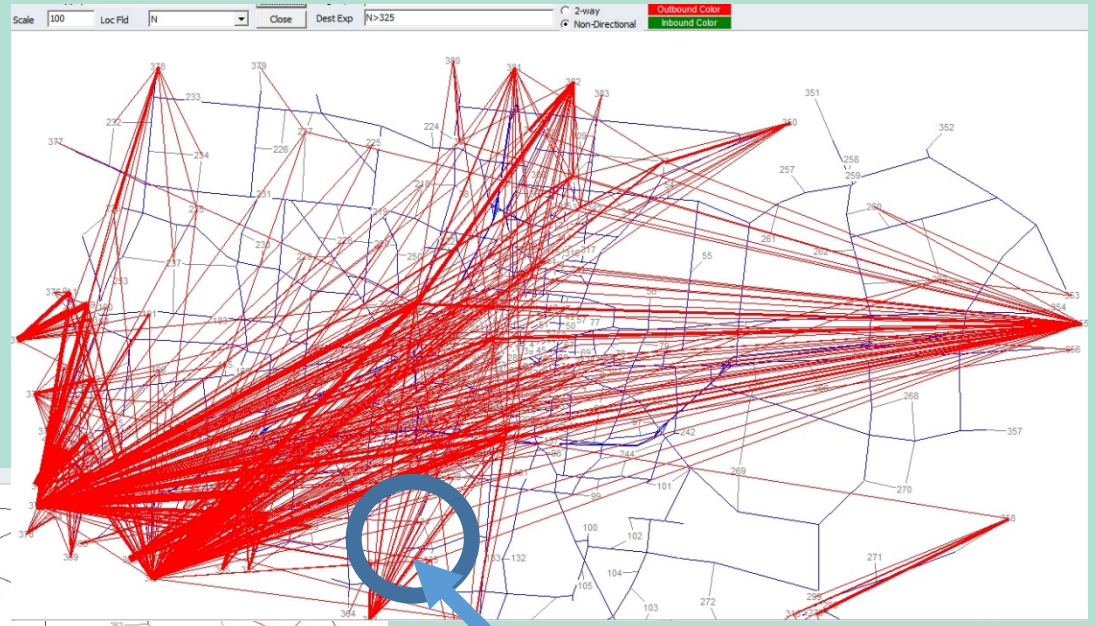
Updated Model Validation Procedures-Special Generators

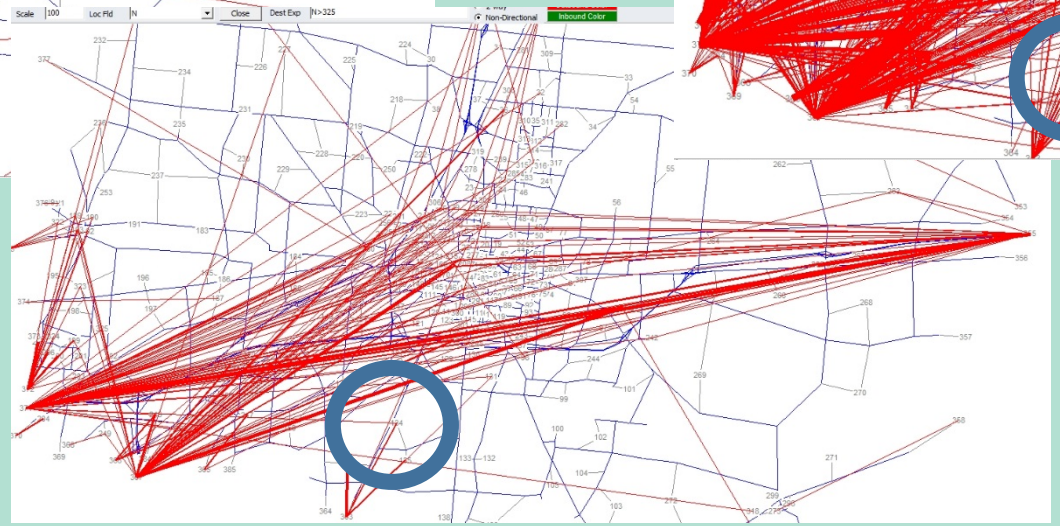
- Desire Line Maps help diagnose special generator problems

Original Model



Final Model



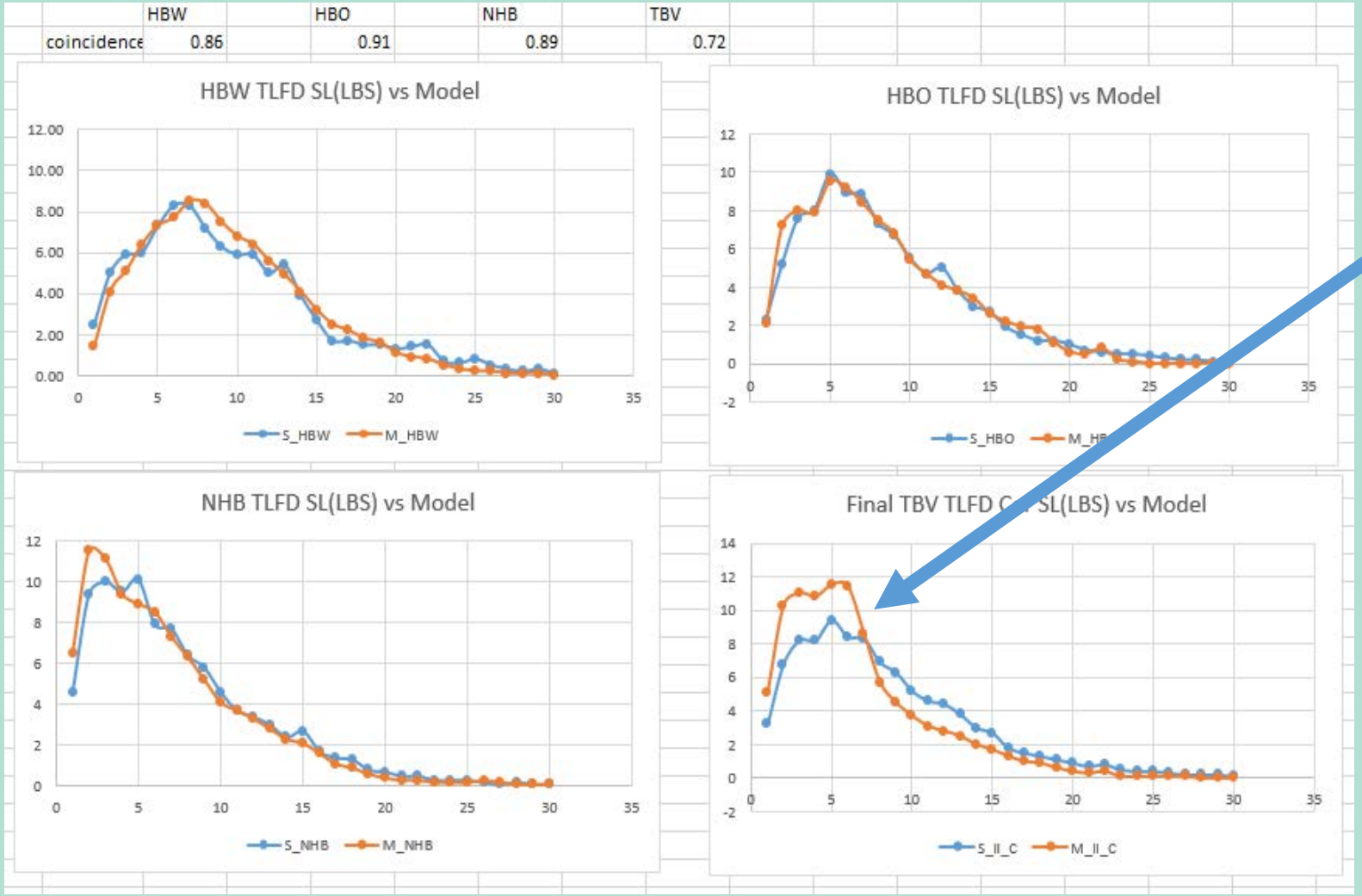


StreetLight Data

Erroneous Special Generator Removed

Updated Model Validation Procedures-TLFD

- While individual TAZ level OD could be suspect, aggregated statistics like TLFD very helpful



- Model compares well by trip purpose but not total
- A factoring process between initial purpose specific trip tables and final assigned trip table to blame and was adjusted

Updated Model Validation Procedures-%IE/%EE

- Generate a special report for updating these
- Can’t bulk update, instead targeted
- 3rd Party Data is generally too sparse at low volume roads, good on freeways
- This is fortunate in that old cordon roadside surveys were good on low volume roads where direct interviews were used and not so good on freeways which used license plate methods

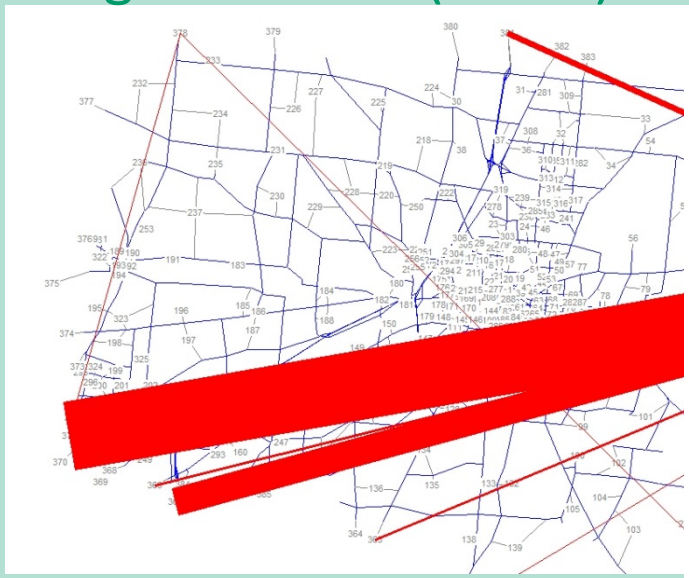
ZONE	VEH CLASS	SL NUM >0	SL IN	SL SPC	SL EXT	SL %IN	SL %SPC	SL %EXT	MOD IN	MOD SPC	MOD EXT	MOD %IN	MOD %SPC	MOD %EXT	%SL OVER MOD
350	CARS	414	780	0	631	0.55	0	0.45	2853	0	881	0.76	0	0.24	0.38
351	CARS	68	120	0	18	0.87	0	0.13	712	0	123	0.85	0	0.15	0.16
352	CARS	30	45	0	307	0.13	0	0.87	451	0	1590	0.22	0	0.78	0.17
353	CARS	70	142	0	314	0.31	0	0.69	774	0	1309	0.37	0	0.63	0.22
354	CARS	76	158	0	12	0.93	0	0.07	714	0	5	0.99	0	0.01	0.24
355	CARS	672	6882	0	30041	0.19	0	0.81	10064	0	21964	0.31	0	0.69	1.15

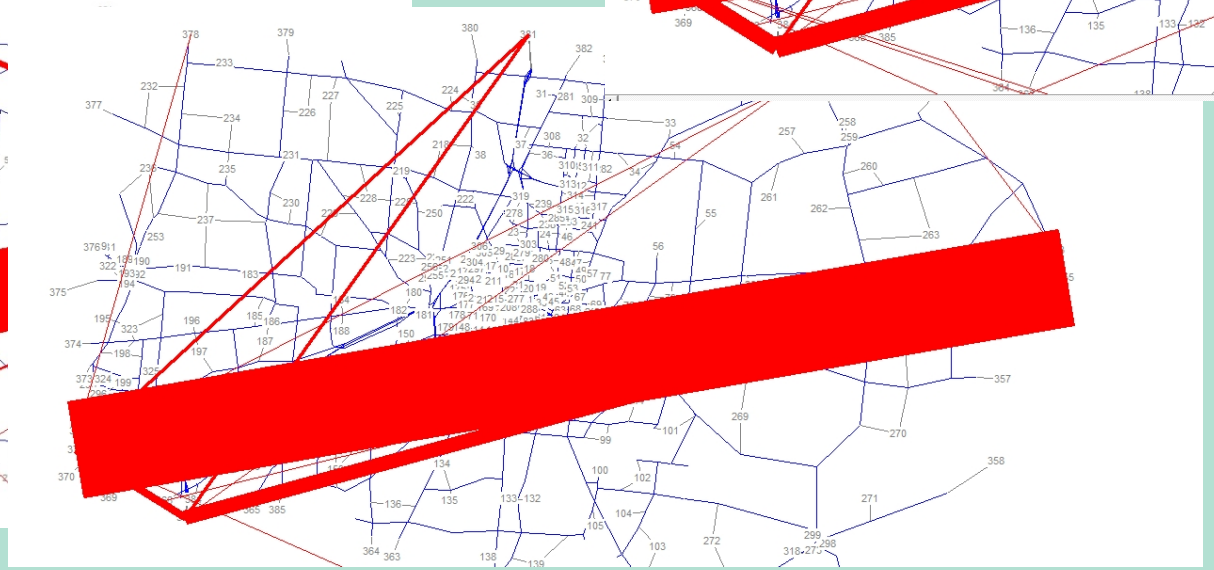
Updated Model Validation Procedures-EE Seed Table

- Same caveats as with %IE/%EE update

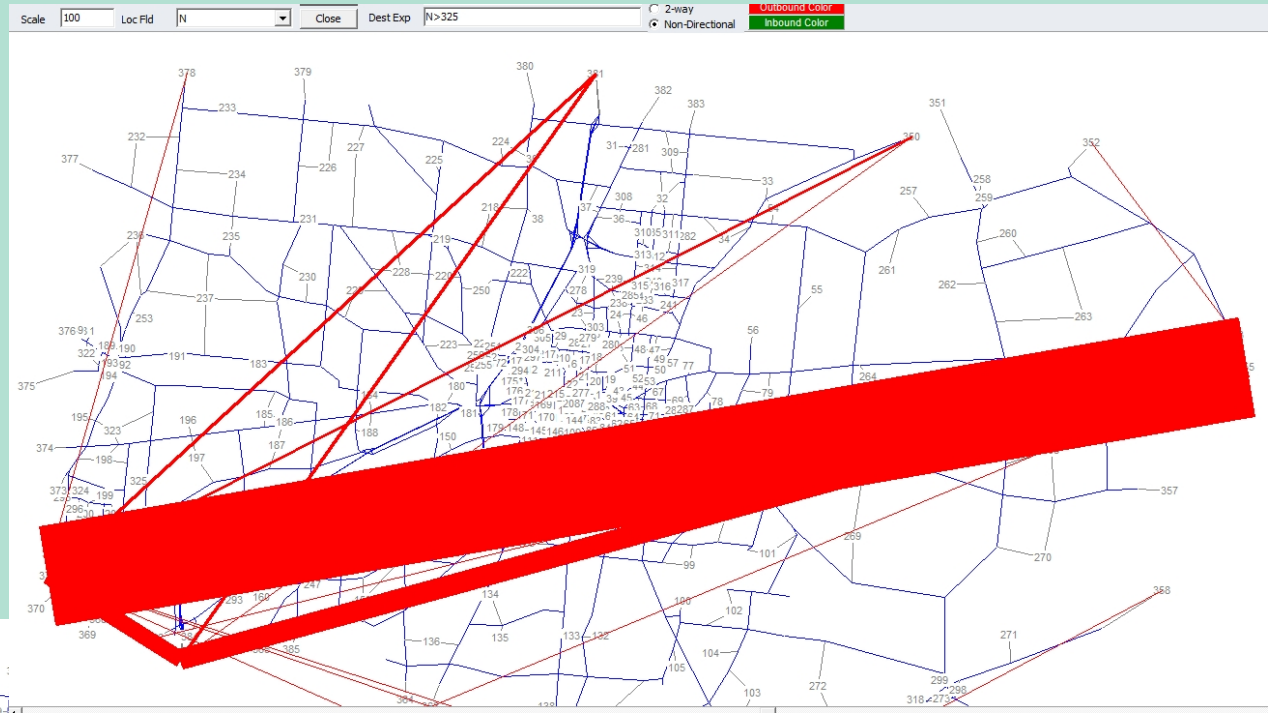
- Apply targeted updates to:
  - Freeway-Freeway interchanges
  - Roads not part of last roadside survey
  - Roads with drastically changed volumes

Original Model (trk EE)





StreetLight Data



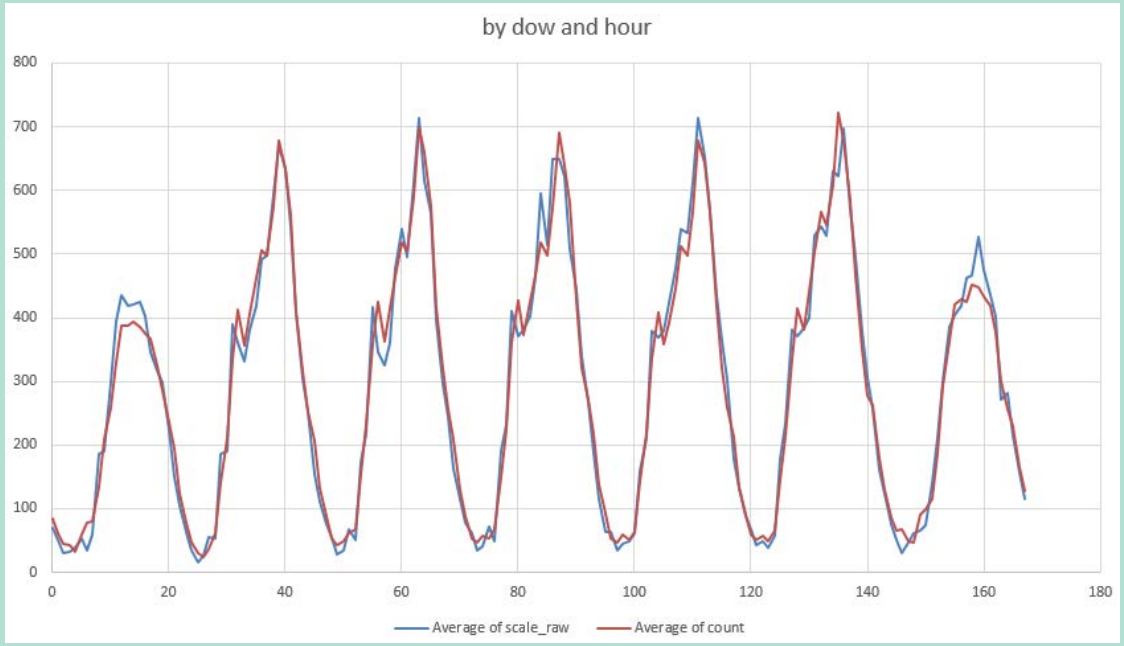
Final Model

The Future!

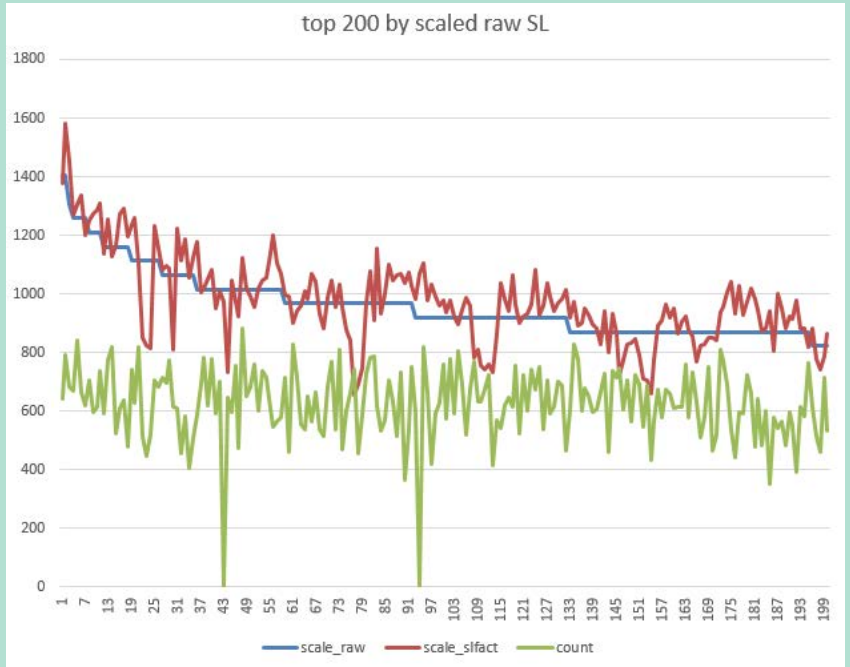
- Temporal data rather than snapshots from home interview surveys and tube counts gives many opportunities

- Currently trying to devise a method to estimate 30th Highest Hour volumes for design rather than relying on look up tables from a limited number of permanent traffic recorders

Good match to permanent counter on average



But currently has issues on the extremes due to greater variance in small sample of 3rd party data



The Future!

- Time varying OD implies a new way of near term forecasting in the future more akin to weather forecasting
- In that paradigm, behavioral models will be relied upon more for long term forecasting, with initial conditions (and gradient) defined by the observed data
- Need to figure out how to blend those rather than ad hoc validation/calibration of behavioral models
- But the observed data needs to be more robust first