RETHINKING LONG DISTANCE AND VISITOR TRAVEL MODELS IN THE ERA OF BIG DATA

MODELING MOBILITY CONFERENCE

September 15, 2025

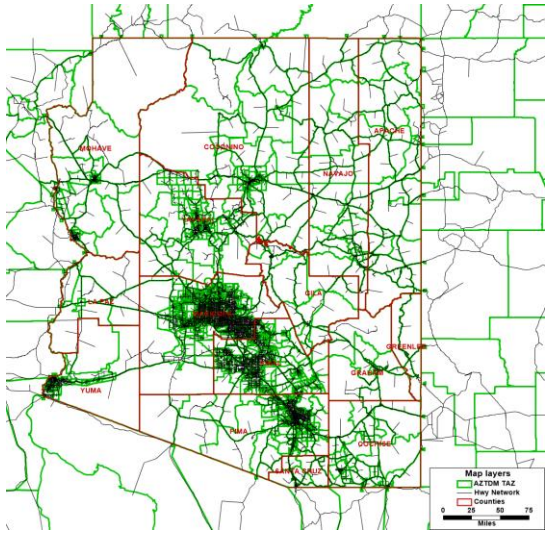
Current Challenges…

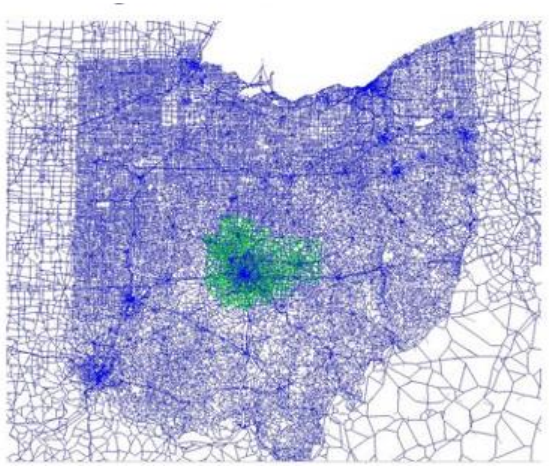
- Long distance travel important to statewide model but relative rare, particularly in HH travel surveys
- Non-resident visitor travel completely missing from HH based surveys

Fix It!

- Fill in with Big Data

Two Case Studies





Arizona

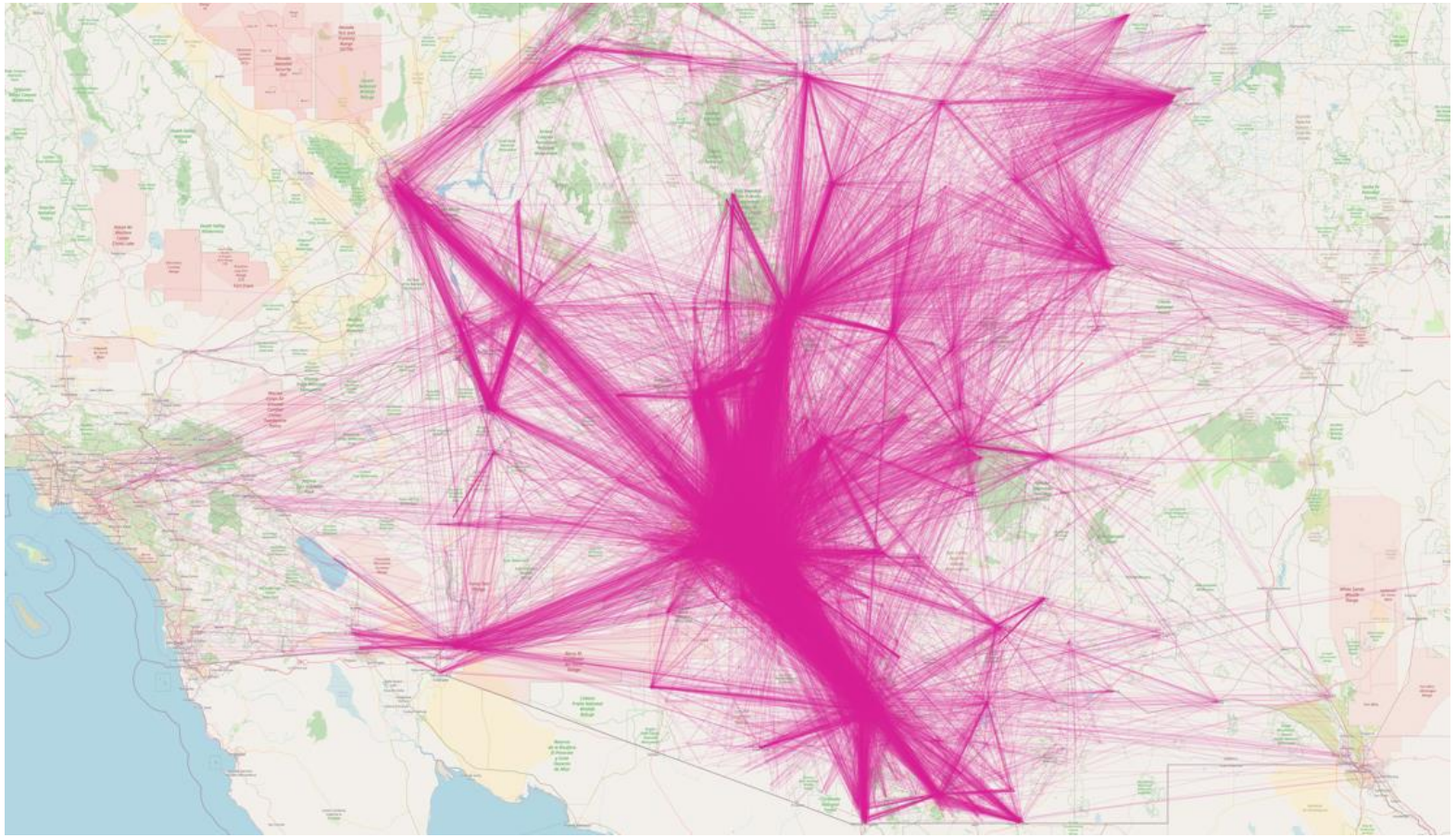
- Former model based on enumerating large sample 2001 NHTS
- No treatment of short distance trips made on long distance tours
- No custom survey data available
- No pre-specified 3POD vendor

Ohio

- Former model based on custom long-distance survey
- LD survey had no stops on tours, just ultimate destination
- Visitor model pieced together from resident models and a tourism survey
- No treatment of short distance trips made on long distance tours
- New LD survey and large sample HTS available
- StreetLight the specified 3POD vendor

Arizona

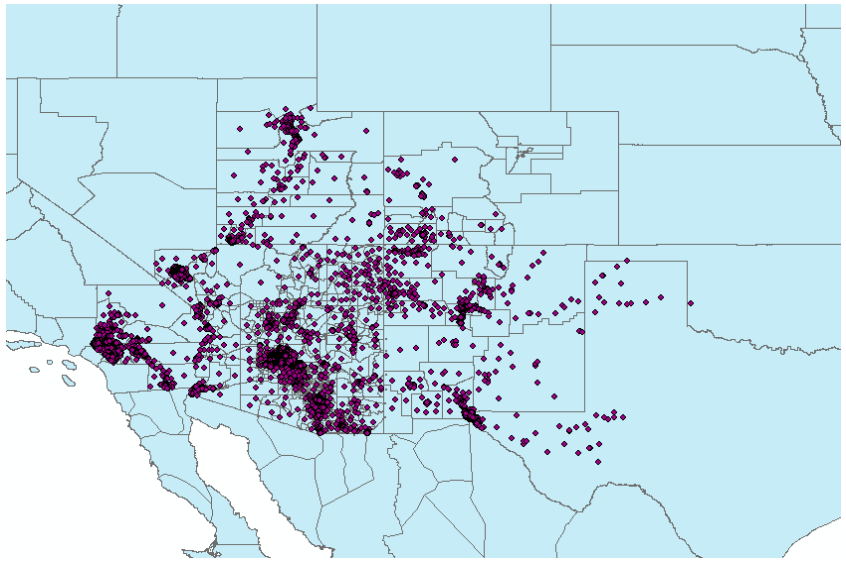
- Selected Replica Data
- Provides synthetic HHs and persons allowing HH based models
- Developed trip-based models with trip generation and purposes consistent with short distance models
- Allows for shared model components where the two have similar characteristics
- Replica allows easy identification of residents/visitors, short/long distance so that each travel market not captured in the NHTS based short-distance resident model can be modeled separately
- Various challenges to overcome…



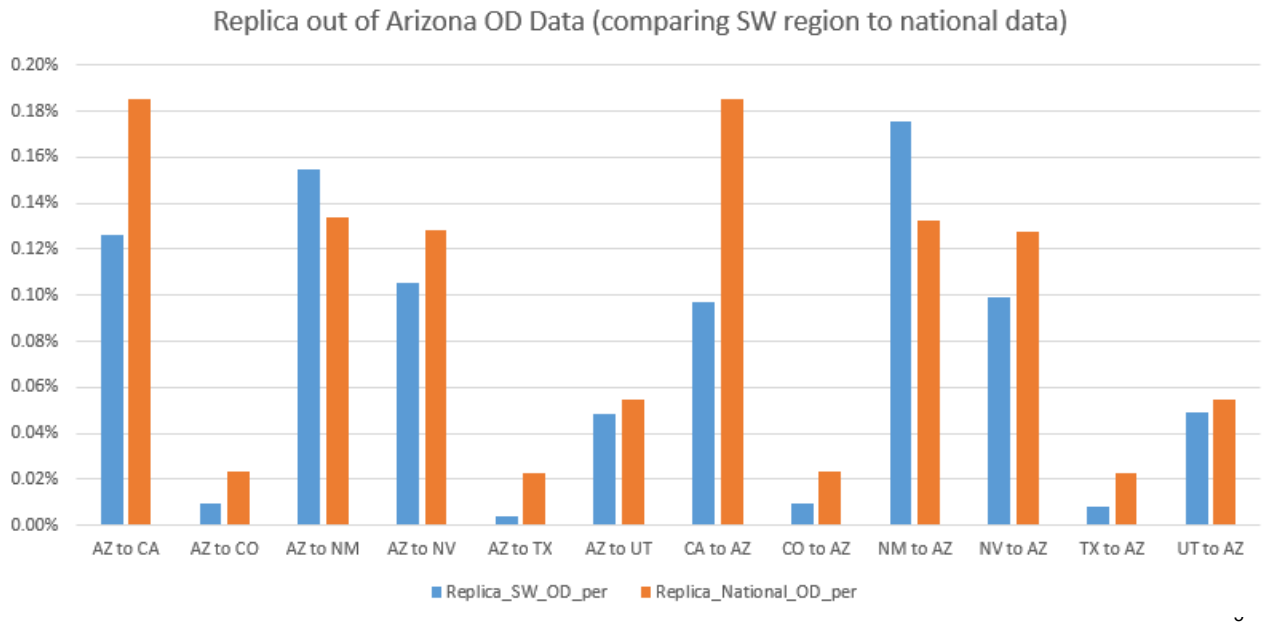
Long distance trips

Regional Boundary Issue

- Replica's data simulation is by region, we used the southwest region, but long-distance trips exceeding that region won't be available
- CA, CO, NV in different regions, however, there are buffer area trips
- Tested vs. Replica's national data product (which doesn't have all the disaggregate detail)
- Decided it was ok to proceed especially considering, 3POD trips don't tend to be exceedingly long due to the existence of incidental stops
- However, the Mexico border was another story…

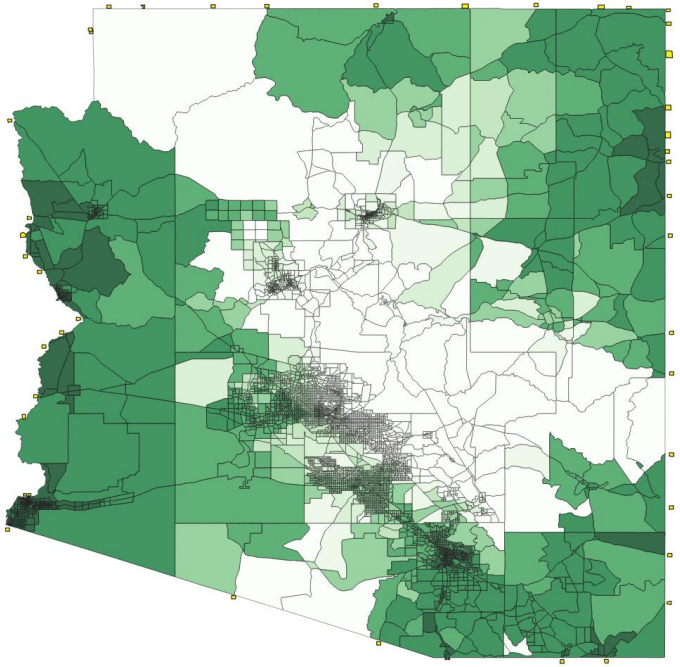


Distribution of trip ends available in SW Region

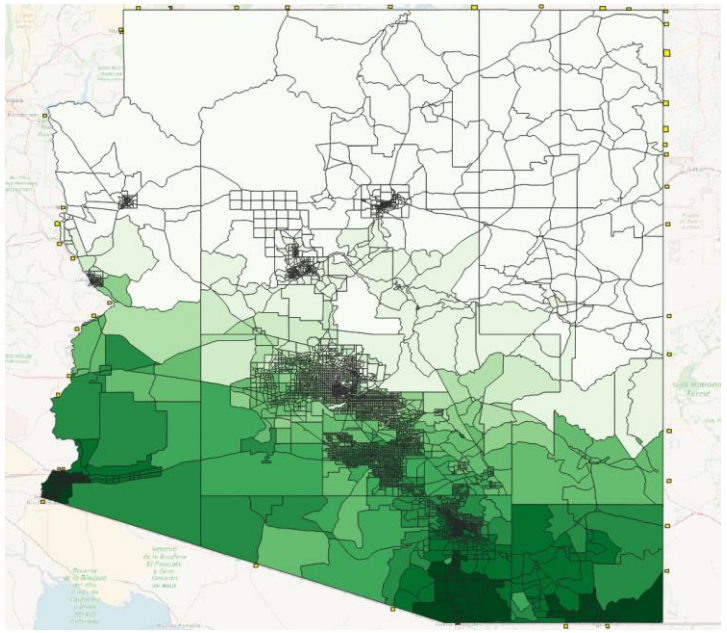


Synthesized Mexico IX/XI aggregate trips

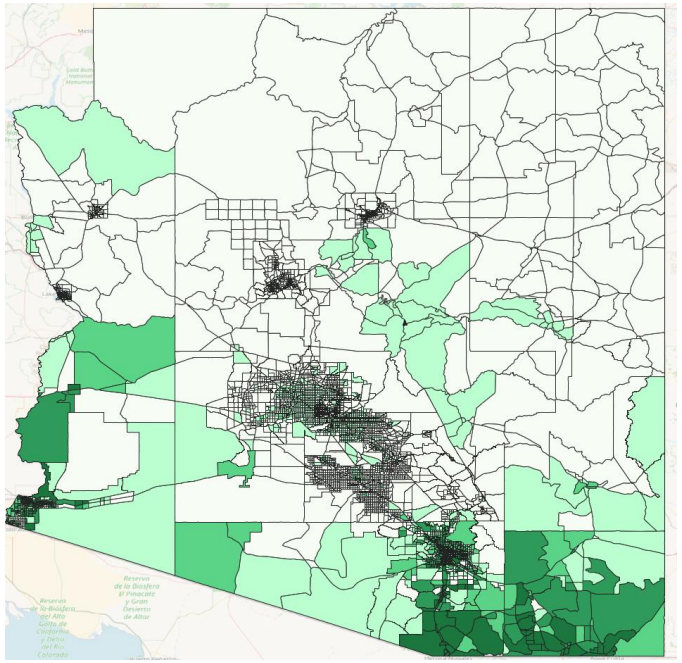
- The LD model will rely on a two level Fratar/Gravity Type Distribution common in LD models due to the secondary impact of travel impedance which results in TLFD that is not monotonically decreasing
- So, we need a base year complete aggregate distribution to start with for the Fratar component
- Created a temporary IE/EI model with accessibility to cordon in the trip generation on the US cordons
- Applied to Mexico cordon to generate synthetic trips to add to seed distribution



IX/XI Accessibility



Accessibility at Mexico cordon



Resultant trips to add

Other Issues

- We also had to be careful with the Replica truck data at the border crossing
- Trucks dray from the border to warehousing and then repackage before rest of US tour
- Needed to carefully compare traffic count locations
- Also had typical double and triple cordon crossing and external station gap issues found in all 3POD (and old-style cordon surveys)…

Note the lack of cars in the Replica totals

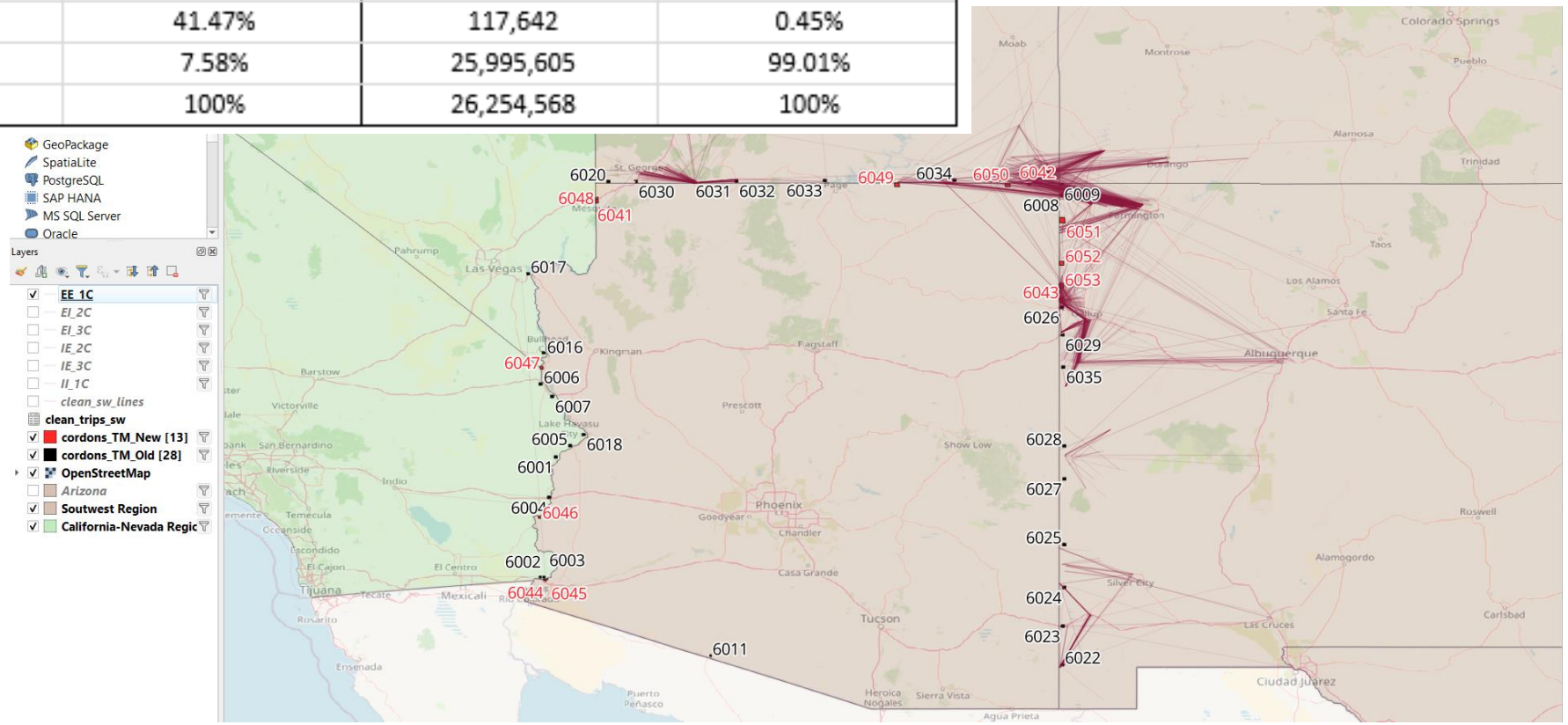
	Replica	Counts	Border Crossing Data	
Location	Total	Total	Total	Truck
US 95 San Luis (near Yuma)	281	19886	18780	268
Mariposa Rd (Nogales)	2206	10127	10127	2032
IR 19/SR 19B (Nogales)	935	27457	12113	0
US 191B (Douglas)	147	14806	8698	162
SR 85 Lukeville	24	2922	2426	4
SR 286 Sasabe	0	382	148	0
Naco	7	1507	1718	16

Warehousing at Nogales truck crossing

Cordon Based Model Issues

Trip Type	Cordon-based Query		State-based Query	
	Count	Percent	Count	Percent
EE	29,510	10.48%	26,510	0.10%
EI	113,917	40.47%	114,811	0.44%
IE	116,753	41.47%	117,642	0.45%
II	21,327	7.58%	25,995,605	99.01%
Total	281,507	100%	26,254,568	100%

- Queried Replica using both external zones and cordon zones
- Identified leakage caused by missing cordon zones and added them
- Identified trips crossing cordons multiple times



Trips crossing cordon in TAZ based query but not in cordon query

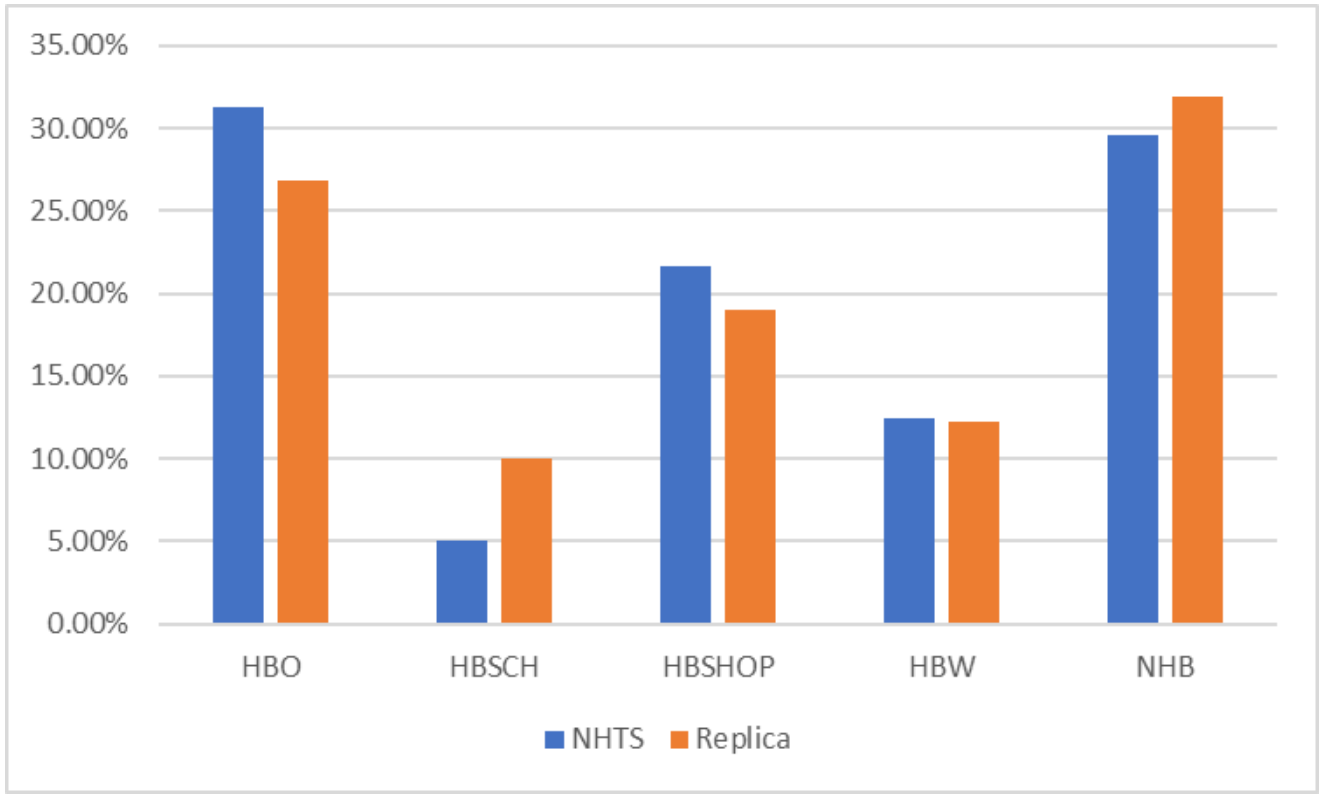
Compared to NHTS to Ensure Compatibility

- SD Resident model updated from NHTS, want the LD/Visitor models to be in scale and not have definitional difference
- Generated a Replica SD Resident model and compare to the NHTS model
- Compared well

Short Distance NHB Xclass Models

SD NHB - NHTS				
Income	1 person	2 persons	3 persons	4+ persons
0-25K	0.91 (264)	2.29 (144)	1.53 (42)	2.90 (90)
25K-50K	1.12 (158)	2.05 (182)	2.23 (84)	3.22 (97)
50K-75K	1.51 (86)	1.70 (145)	3.10 (58)	2.64 (78)
75K-150K	1.52 (79)	1.96 (253)	4.04 (101)	5.30 (192)
150K and more	0.51 (14)	3.02 (71)	1.70 (33)	5.21 (56)

SD NHB - Replica				
Income	1 person	2 persons	3 persons	4+ persons
0-25K	0.85	1.53	2.25	2.75
25K-50K	1.14	1.72	2.84	3.41
50K-75K	1.33	2.07	3.23	4.03
75K-150K	1.39	2.36	3.69	4.48
150K and more	1.25	2.48	4.03	4.97



Compared to NCHRP 735 Transferable Parameters

- **Compared well**
- **Thus, the rates coming out of the 3POD match expectation but with this data we get enough to do things such as…**

Table 4.2. Georgia long-distance internal and external trip rates by purpose, income, area, and persons per household.

Income	Area	Persons per Household	HBW-IE (GA Int-Ext)	HBW-II (GA Internal)	HBO-II (GA Internal)	NHB-II (GA Internal)
Low	Urban	1	0.008	0.001	0.036	0.005
		2	0.045	0.002	0.063	0.009
		3	0.025	0.003	0.083	0.020
		4	0.077	0.005	0.060	0.154
	Rural	1	0.045	0.045	0.016	0.010
		2	0.020	0.043	0.087	0.130
		3	0.091	0.003	0.045	0.040
		4	0.056	0.167	0.667	0.056
Non-Low	Urban	1	0.016	0.003	0.013	0.010
		2	0.046	0.005	0.041	0.017
		3	0.051	0.009	0.041	0.054
		4	0.051	0.015	0.127	0.036
	Rural	1	0.015	0.002	0.032	0.021
		2	0.035	0.022	0.104	0.042
		3	0.052	0.007	0.095	0.087
		4	0.070	0.022	0.081	0.059

Source: Atkins, *Development of Statewide Model Draft Report*, prepared for Georgia Department of Transportation, April 15, 2011.

Table 91: LD NHB Trip Rates

	1 person	2 persons	3 persons	4+ persons
0-25K	0.03	0.05	0.06	0.09
25K-50K	0.02	0.05	0.07	0.08
50K-75K	0.03	0.05	0.07	0.09
75K-150K	0.03	0.05	0.07	0.09
150K and more	0.02	0.05	0.07	0.09

Regression Attraction and Visitor Generation Models

- Regression on counties (TAZ too unstable)
- Attraction variables different from those seen in regional travel models
- Also get non-resident visitors including short and long distance (modeled separately) II trips

Table 98: LD NHB Visitor Trip Attraction Models

NHB Visitor	Short-distance		Long-distance	
	Coef.	t-Stat.	Coef.	t-Stat.
Total Population	0.14374	41.95357	0.00319	6.70568
Seasonal Dwelling Units	0.37326	3.17891	0.13158	8.08177
Park Visitors	0.00133	1.28803	0.00133	9.29583
Total Employment	0.13903	32.34401	0.00372	6.24292

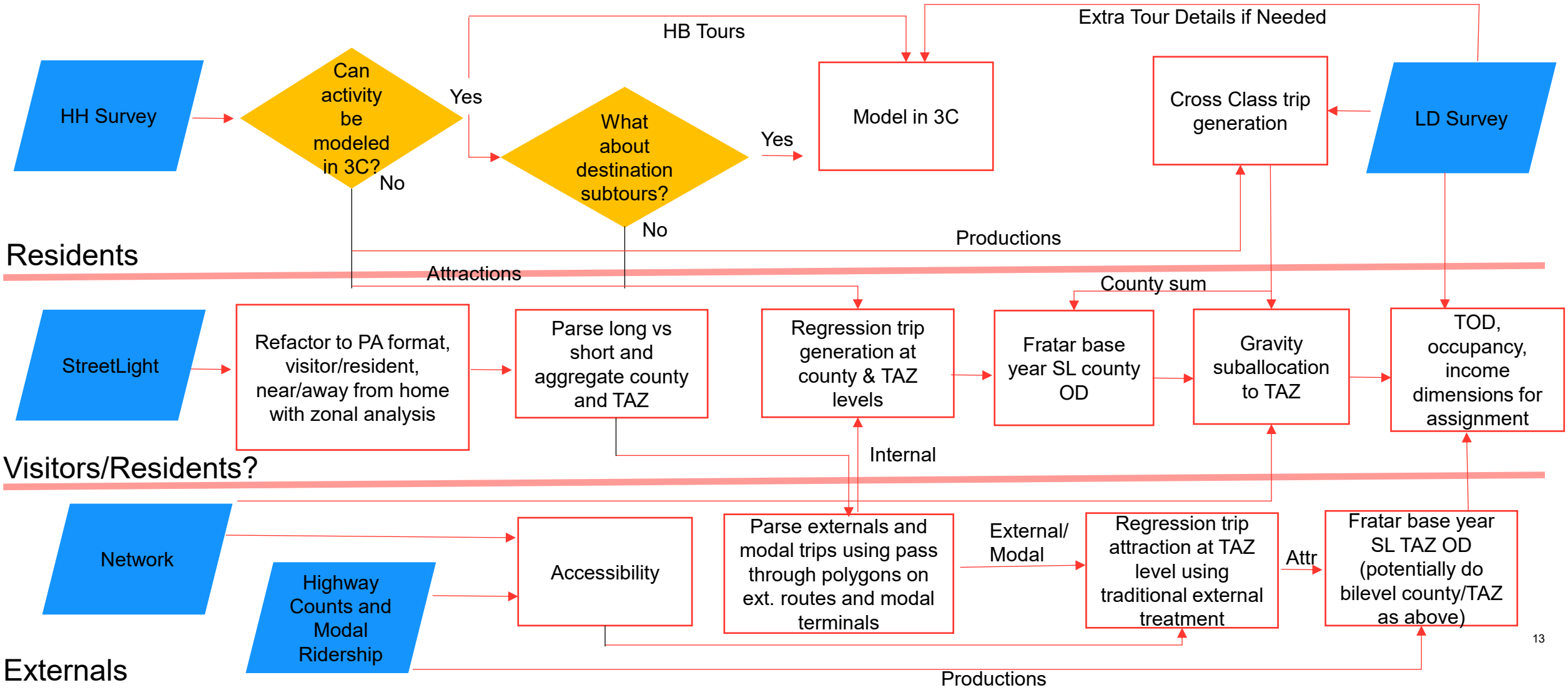
Table 95: LD NHB Resident Trip Attraction Models

NHB	Urban		Suburban		Rural	
	Coef.	t-Stat.	Coef.	t-Stat.	Coef.	t-Stat.
Total Population	0.00713	25.29358	0.00998	29.85649	0.03628	15.84586
Seasonal Dwelling Units	0.01345	0.70039	0.15362	14.59562	0.05114	1.58117
Park Visitors	0.00489	1.53264	0.00081	2.46701	0.0005	3.13352
Total Employment	0.00457	16.66534	0.01213	17.94426	0.05656	11.85876

Ohio

- **A new long-distance survey is available which was obtained via GPS capture**
- **Includes all trips over 50 miles (but not shorter trips on such long-distance tours)**
- **Also misses start/end of tours due to necessary GPS latency in 6-month long survey**
- **Only includes trips for those HH members participating by GPS device which introduces additional challenges**
- **Also have a very large (25K) traditional HTS which may contain substantial full detail LD trips**
- **SD resident models are ABM, desire to incorporate as much travel in the HH activity as possible**
- **StreetLight Prespecified as the 3POD**
  - Avoids the regional boundary issues of Replica
  - Lacks disaggregate HH level data, in particular, can't separate residents/visitors with standard StreetLight interface, need special help from StreetLight staff
- **Result…**

We are embarking on a fairly complex data analysis exercise to separate out the various market segments



THANK YOU

wsp.com