Using Cellphone O-D Data for Regional Travel Model Validation

Delivered at 15th TRB Planning Applications Conference

May 19, 2015

Travel Model Coverage Area
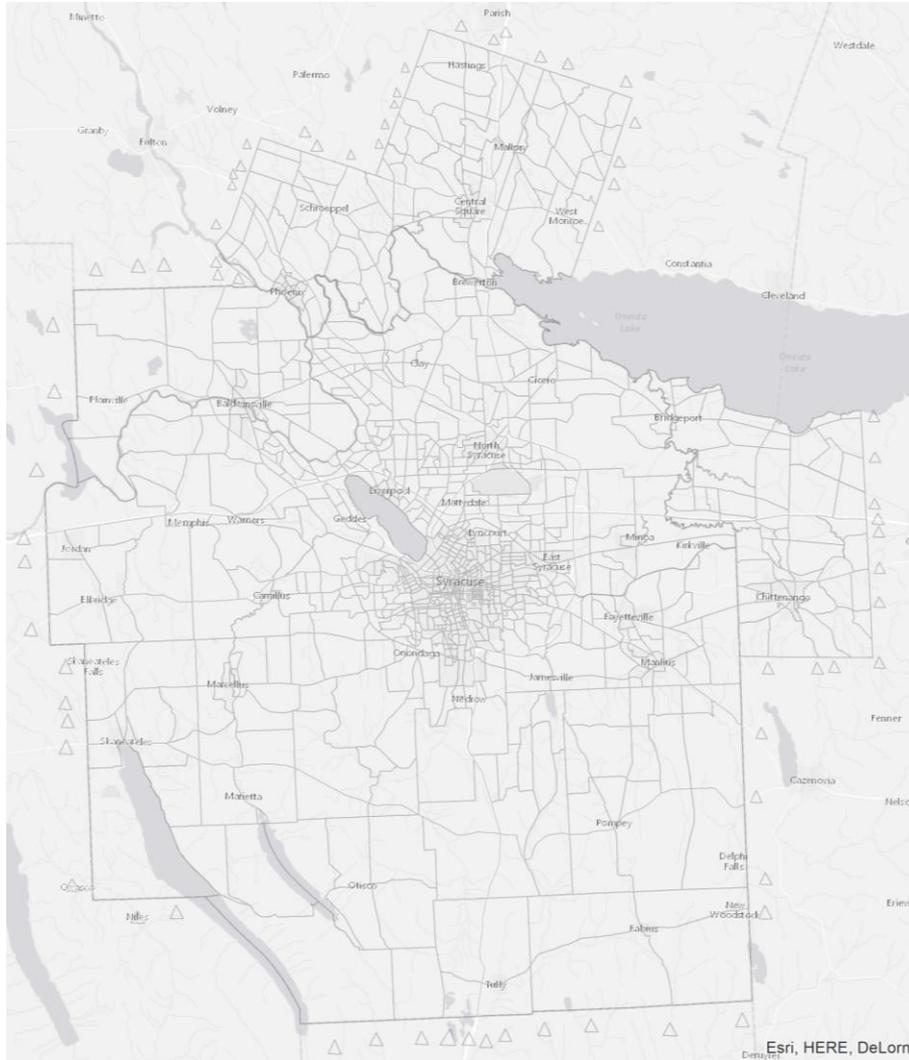
- 1,185 I & E Model Zones

- 788 Zones for Cellular Data

- 45 Miles North-South

- 35 Miles East-West

- Four-step Travel Model built in TransCAD by RSG

AirSage Methodology Overview

Activity pattern analysis/point generation: Assumed home and work locations are imputed based on 4 to 6 weeks of recorded data. Statistical *clustering* algorithms are used to identify home, work and intermediate stop locations.

- Home location: where mobile user clusters between 9pm – 6am
- Work location: where mobile user clusters between 9am – 5pm
- "Trip Legs" are formulated around home & work locations to arrive at a daily trip pattern

Device location processing: Time stamped locations coordinates from devices are continuously accessed and recorded. Trip movements identified by time & distance criteria.

- Trip O-Ds must be at least 1.2-1.5 km (0.75-0.93 miles) in distance
- If device movement stops for 5+ minutes, a destination is assumed

Population synthesis: Trip movements from the observed sample devices are expanded based on the ratio of observed devices to the 2010 Census population at the *Census Tract level*

Trip analysis: Trips are distinguished by traveler type, trip purpose, time of day and day of week

Data aggregation and packaging: Trip O/Ds are summarized by geography, typically zip codes or TAZs

Assumptions / Limitations / Strengths

Data Collection / Device Location: AirSage locations are collected only when the device interacts with the cellular network (*start/end a call, send text message, data transfer, etc*).

Linked Trips: HBW trips with a 5+ minute stop for day-care drop off or coffee could become HBO and NHB trip reducing the % of HBW trips in the model and increasing the % of NHB trips.

Non-modeled trips: Walk-the-dog trip could also increase the % of NHB trips when comparing AirSage to model results.

O/D Location Inference: There could be some issues in areas where a substantial % of smartphone users belong to a certain non-conventional classification (e.g. students or night-shift workers).

No mode share, vehicle classification, auto occupancy information: Thus no comparison can be made along those dimensions

Trip count, trip purpose, TOD information is available: Should be able to focus on those in great detail

Comparing Trip Purpose Shares

HBW

Model	Value
NCHRP	14%
MATS AirSage	11%
MATS Model	26%
TPB AirSage	27%
TPB Model	19%
SMTC AirSage	25%
SMTC Model	15%

HBO

Model	Value
NCHRP	54%
MATS AirSage	51%
MATS Model	53%
TPB AirSage	51%
TPB Model	54%
SMTC AirSage	56%
SMTC Model	52%

NHB

| Model	Value
NCHRP	30%
MATS AirSage	38%
MATS Model	21%
TPB AirSage	21%
TPB Model	27%
SMTC AirSage	19%
SMTC Model	33%
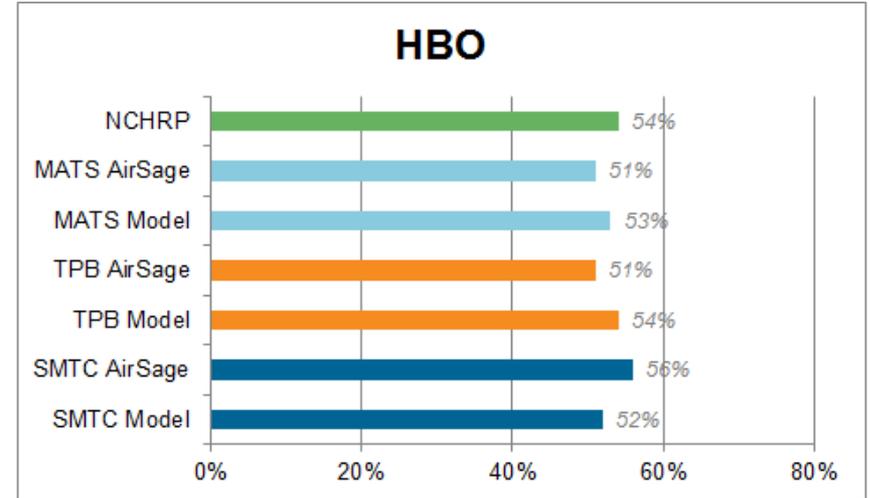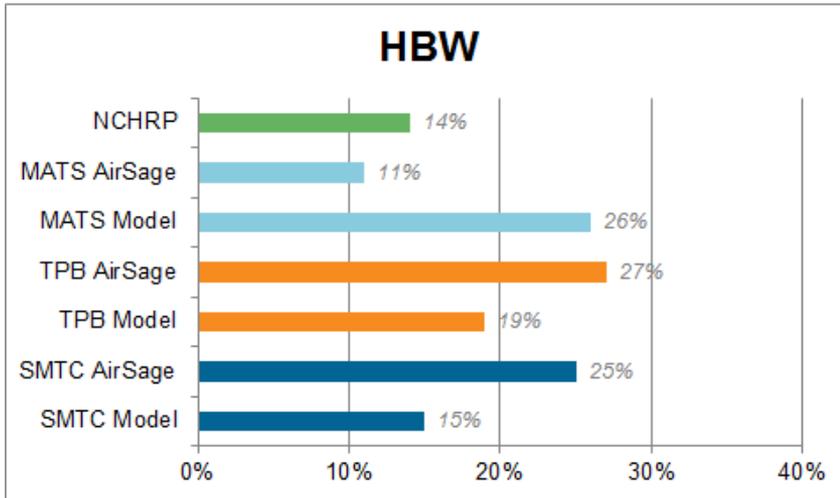
- AirSage appears capable of identifying Home-Based Other (HBO) trip purposes

- AirSage appears less capable of differentiating between HBW and NHB trip purposes
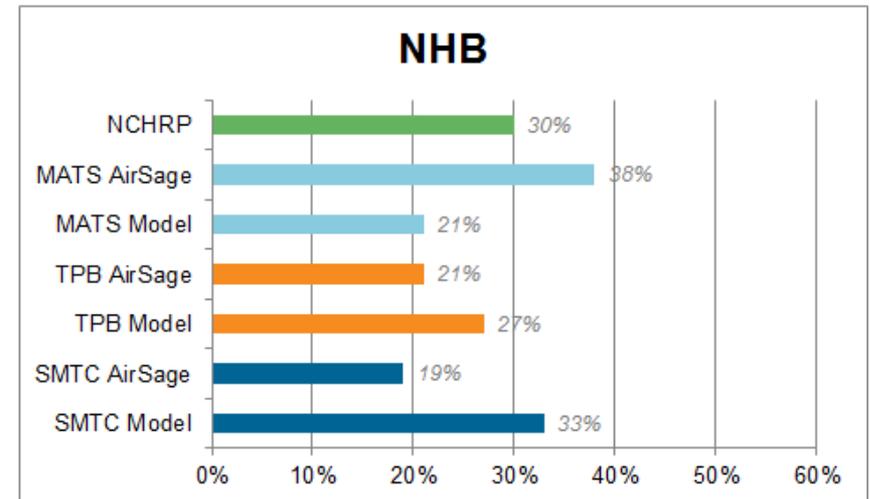
HBW Destinations vs. Employment Model

Model HBW Destinations

$$y = 0.6214x + 167.33$$
$$R^2 = 0.8433$$

Y-axis: HBW Trips (0 to 3,000)
X-axis: Employment (0 to 4,000)

Model work trip destinations are correlated with zone level employment locations
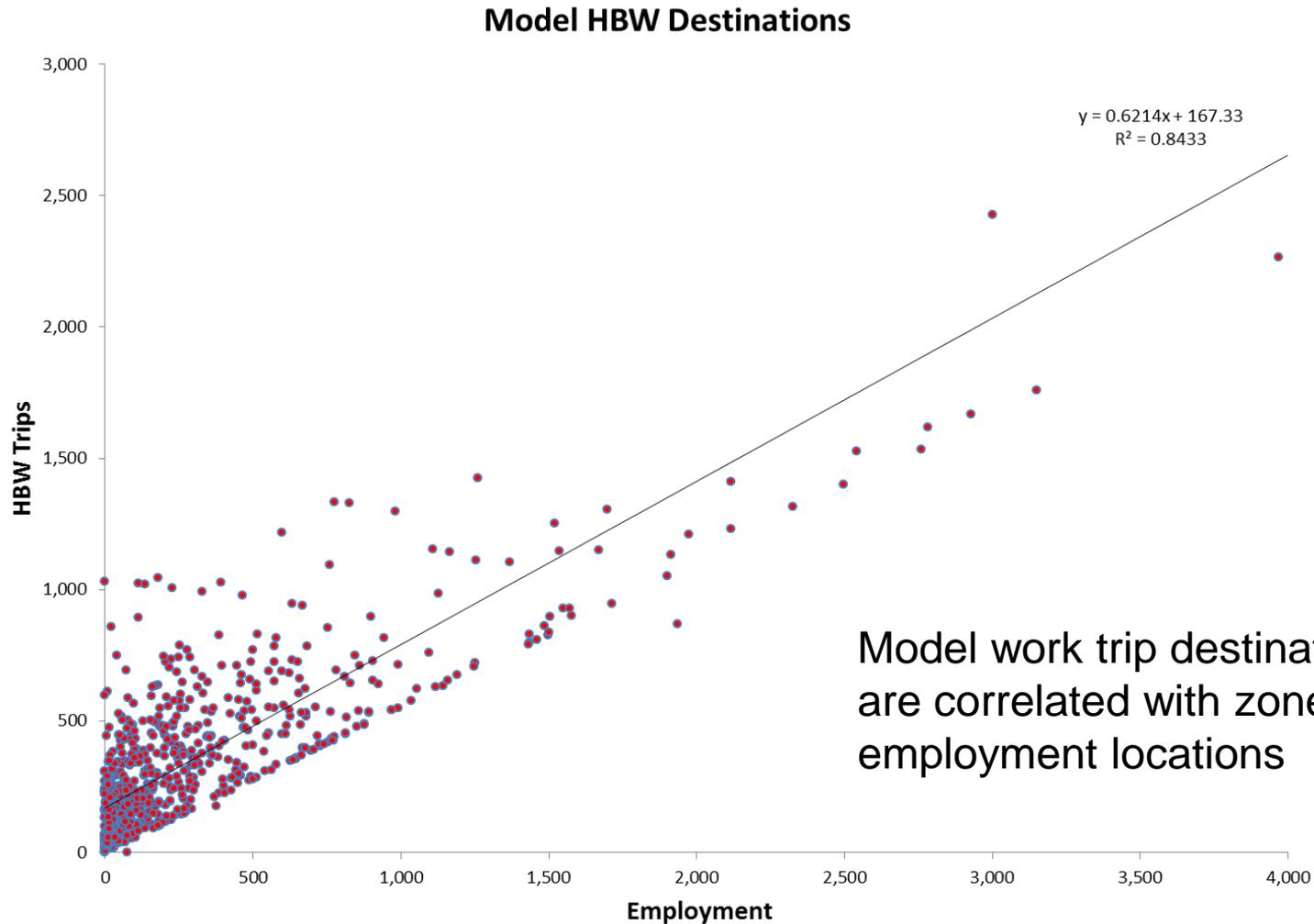
HBW Destination vs. Employment AirSage

AirSage HBW Destinations

$y = 0.6169x + 384.6$
$R^2 = 0.3576$

HBW Trips

Employment

AirSage work trip destinations are *not* correlated with zone level employment locations
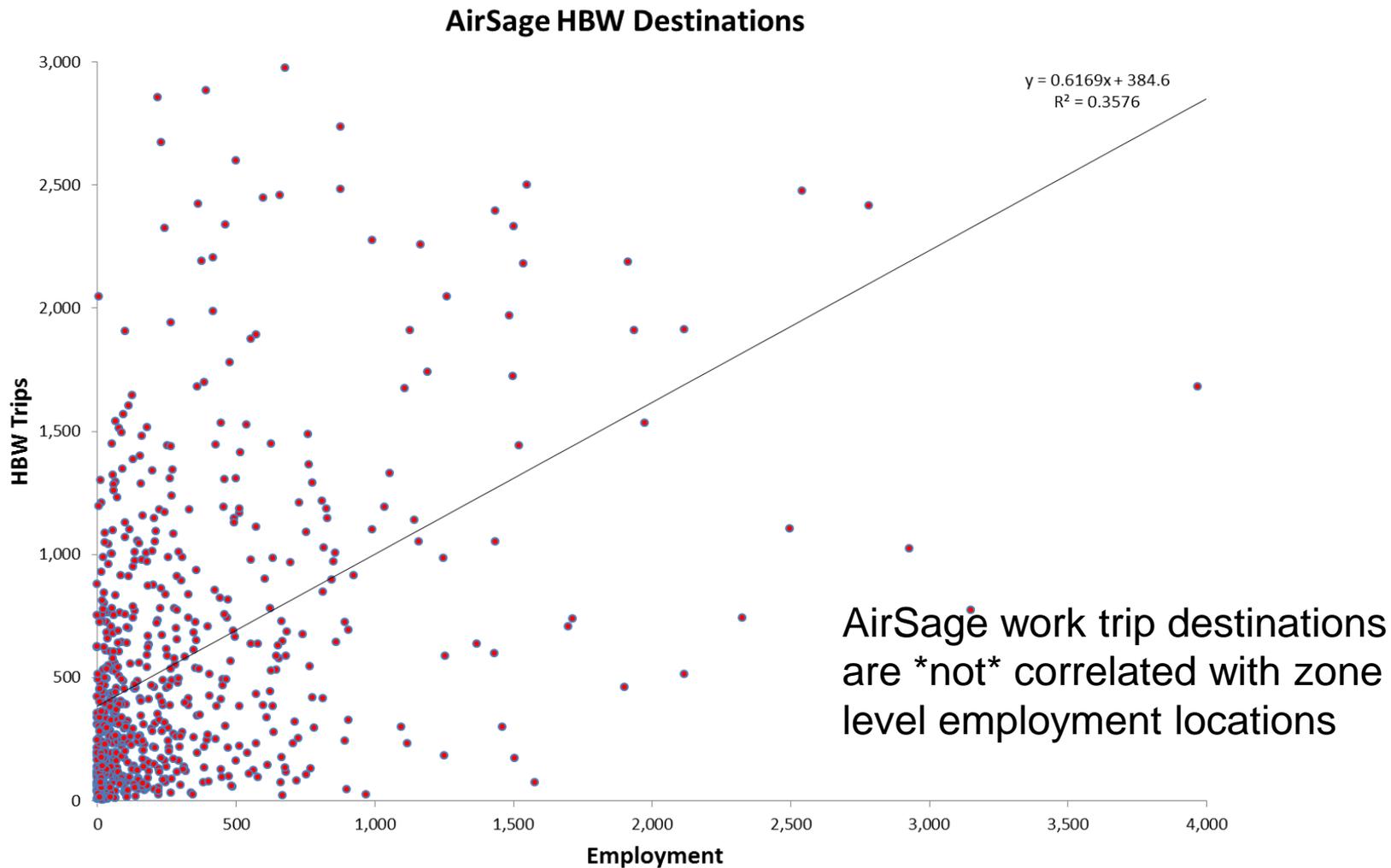
Disaggregate View of the Data
University of Syracuse as Test Case (3 zones)
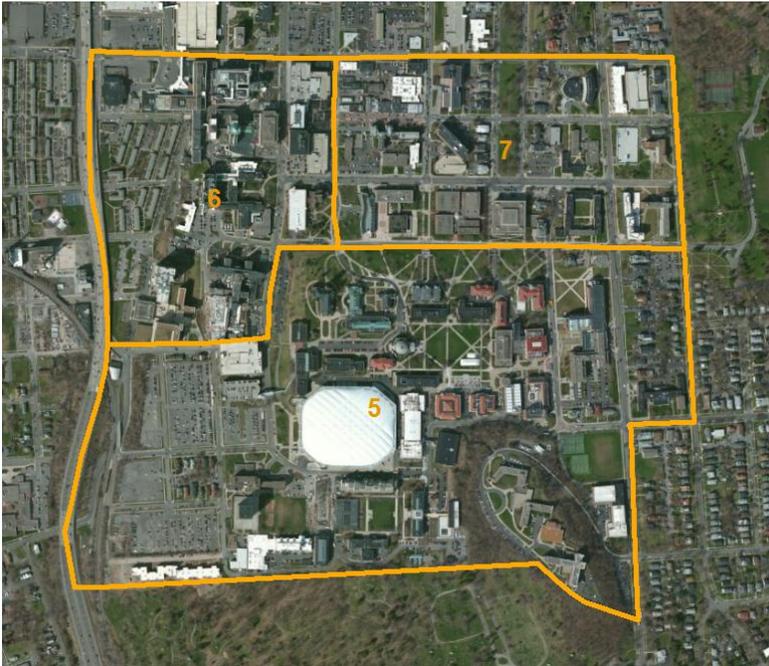

SU has ~20,000 students
SU has ~5,000 faculty/staff

AirSage trips seem much too low for this
small yet important subarea

Purpose	AirSage trips
HBW	9,300
HBO	9,400
NHB	3,900
TOTAL	22,500

| Purpose	Model trips
HBW	15,300
HBO	25,000
NHB	25,500
TOTAL	66,000

Dividing the Region into 26 Medium Districts Municipalities
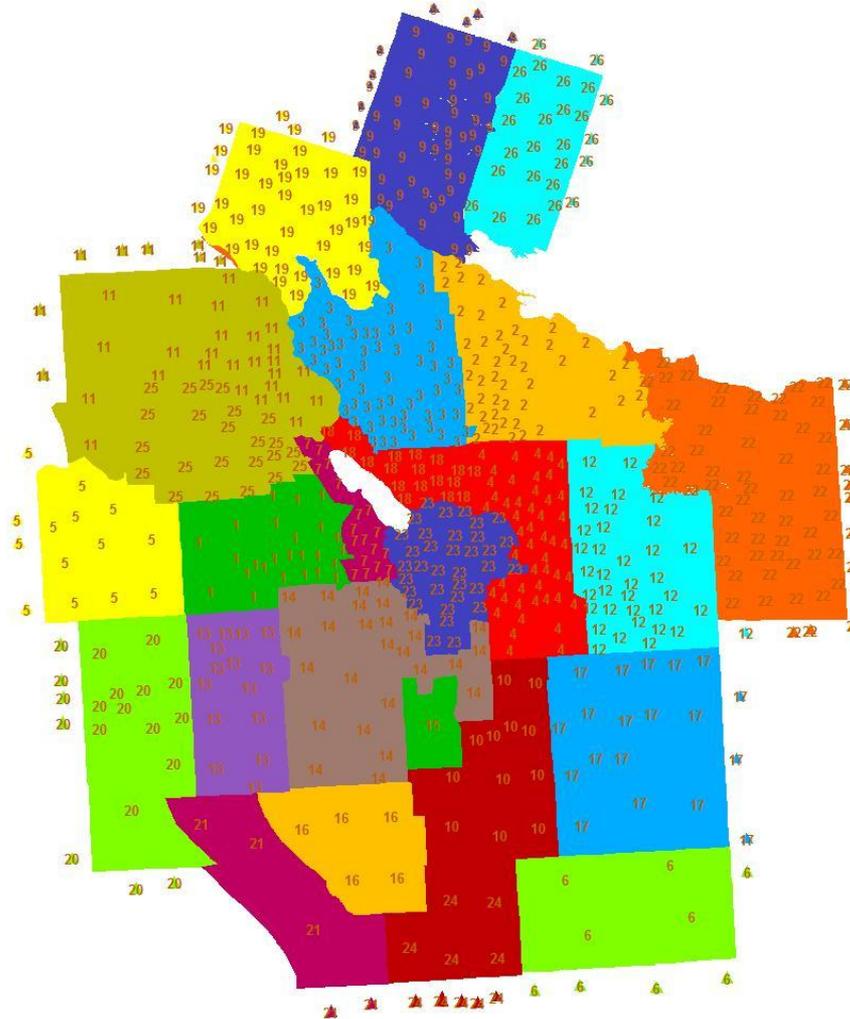
# Town-to-Town Trip Flows
[26 x 26] = 676 OD pairs plotted

Model vs. Airsage

Correlation = 0.945

y = x

Model Trips (Log)

AirSage Trips (Log)
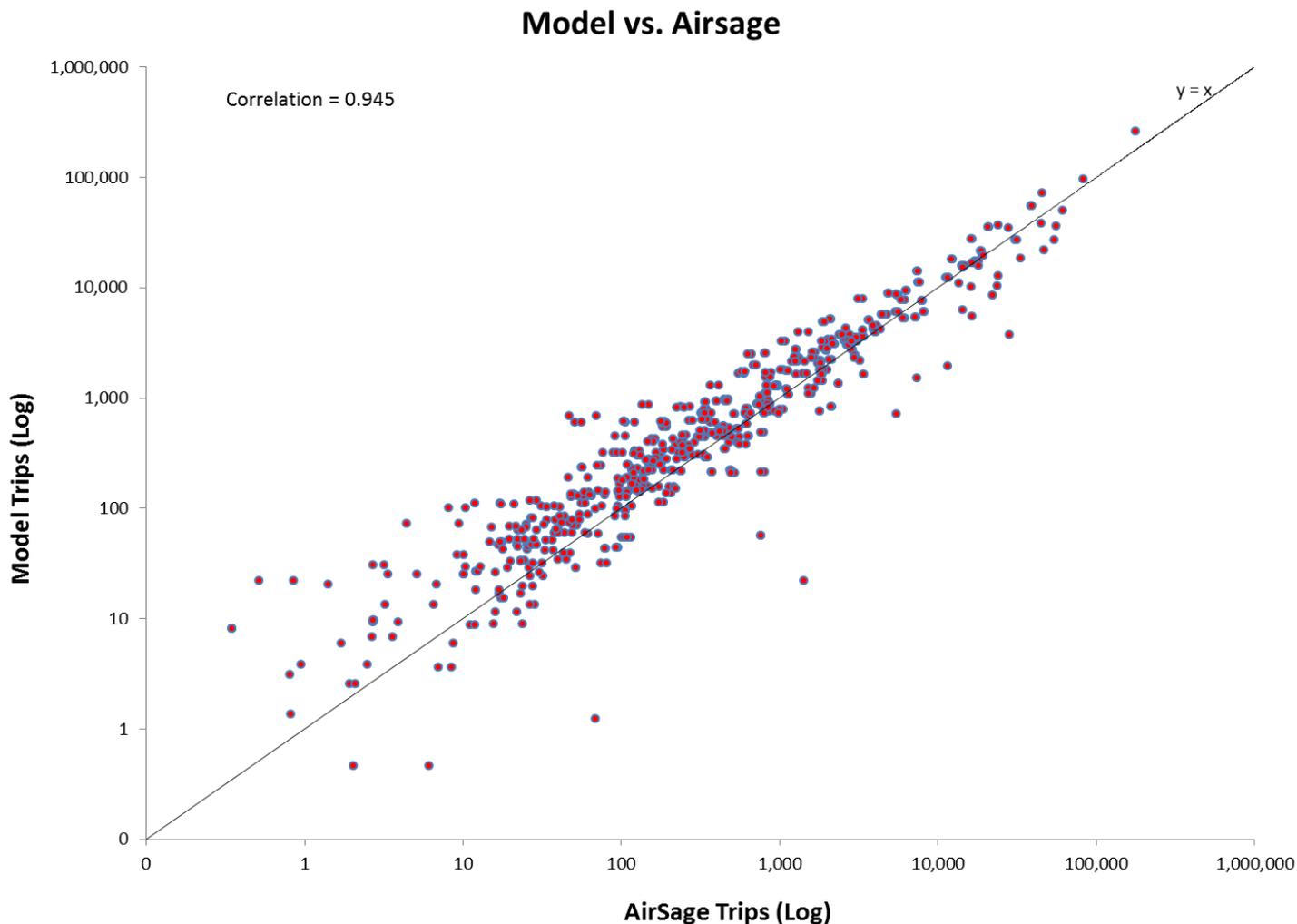
Town-to-Town HBW Trip Flows
[26 x 26] = 676 OD pairs plotted

Model vs. Airsage

Correlation = 0.947

● HBW

y = x

Model Trips (Log)

AirSage Trips (Log)
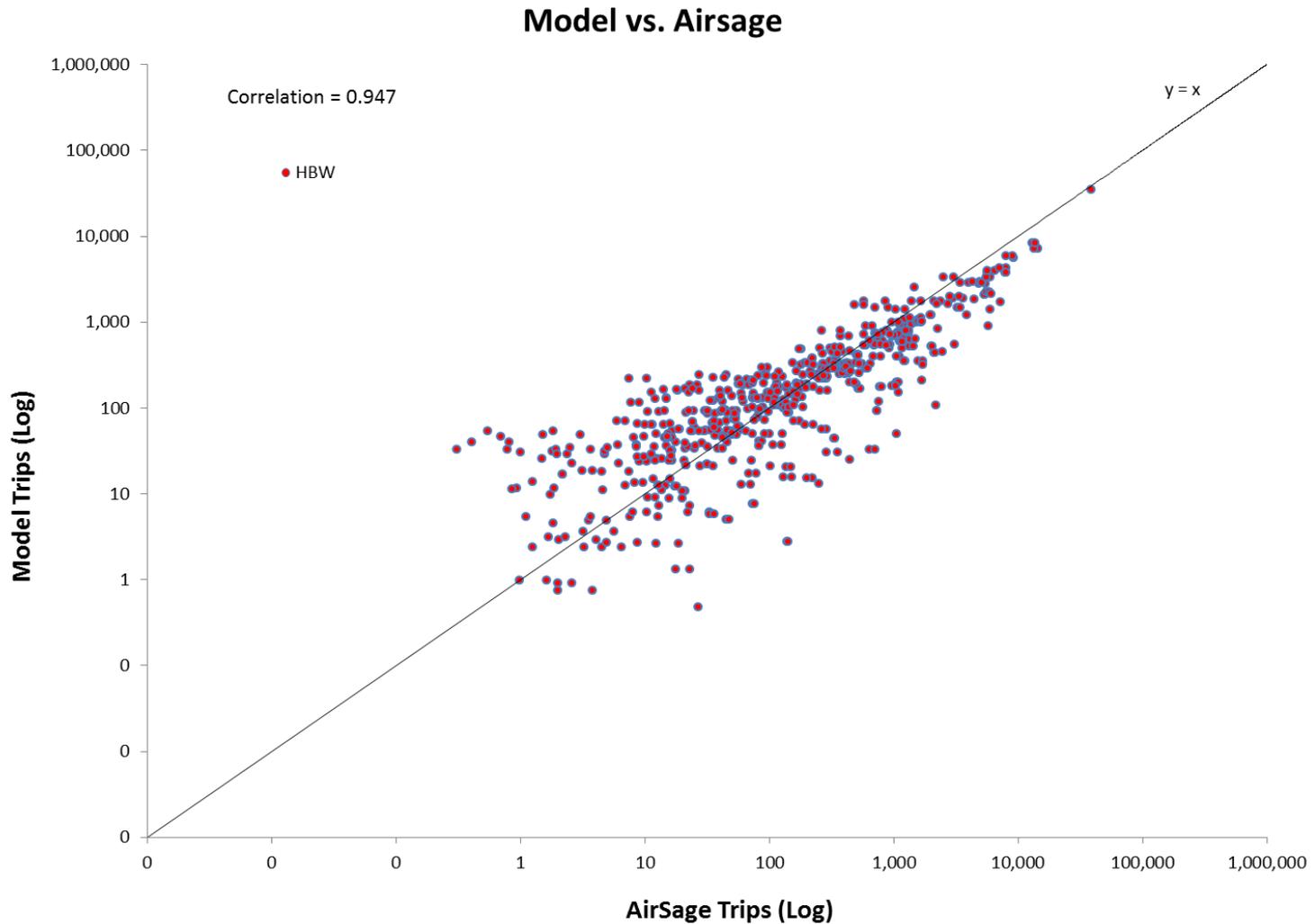
Town-to-Town HBO Trip Flows
[26 x 26] = 676 OD pairs plotted

Model vs. Airsage
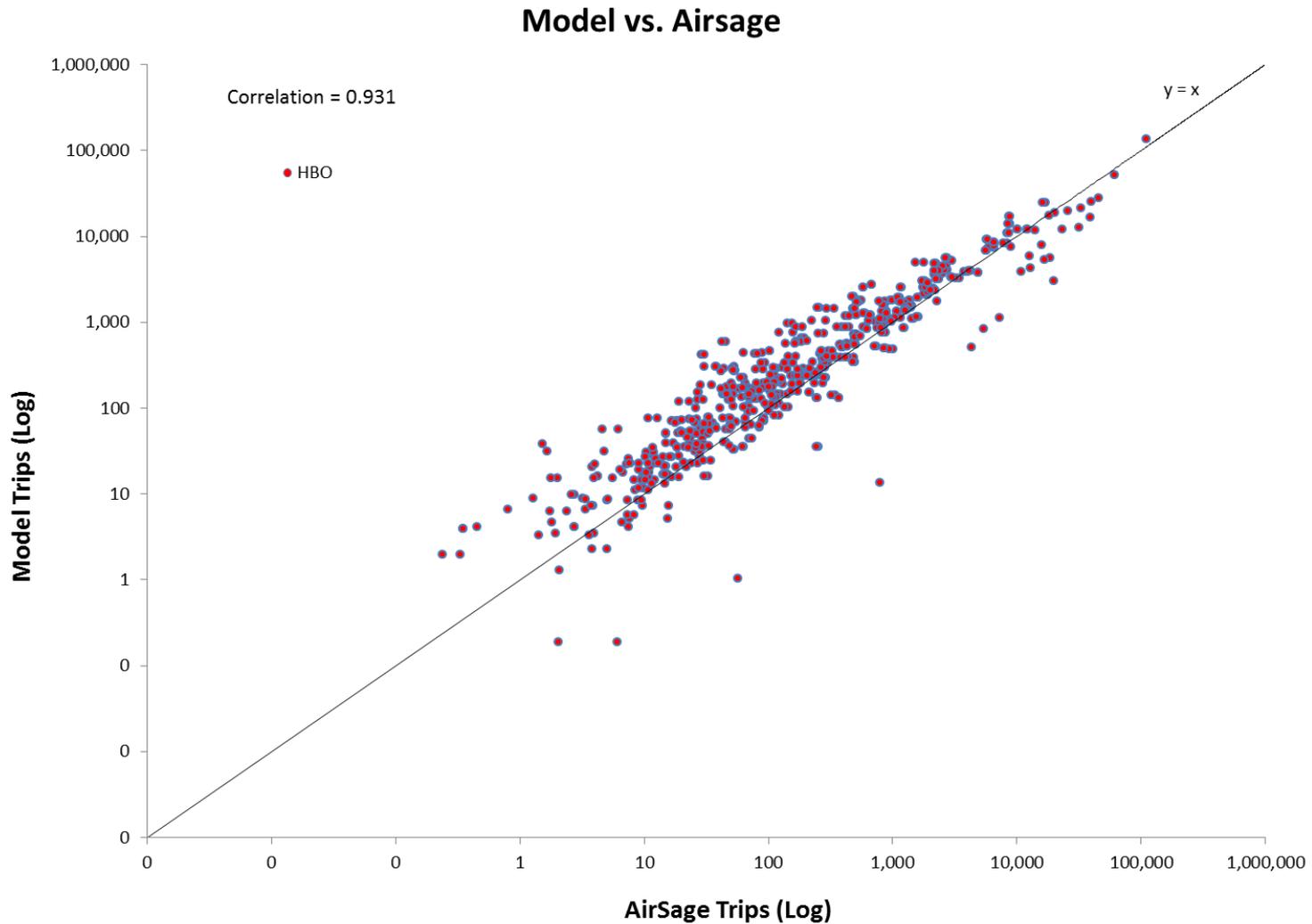
Town-to-Town NHB Trip Flows
[26 x 26] = 676 OD pairs plotted
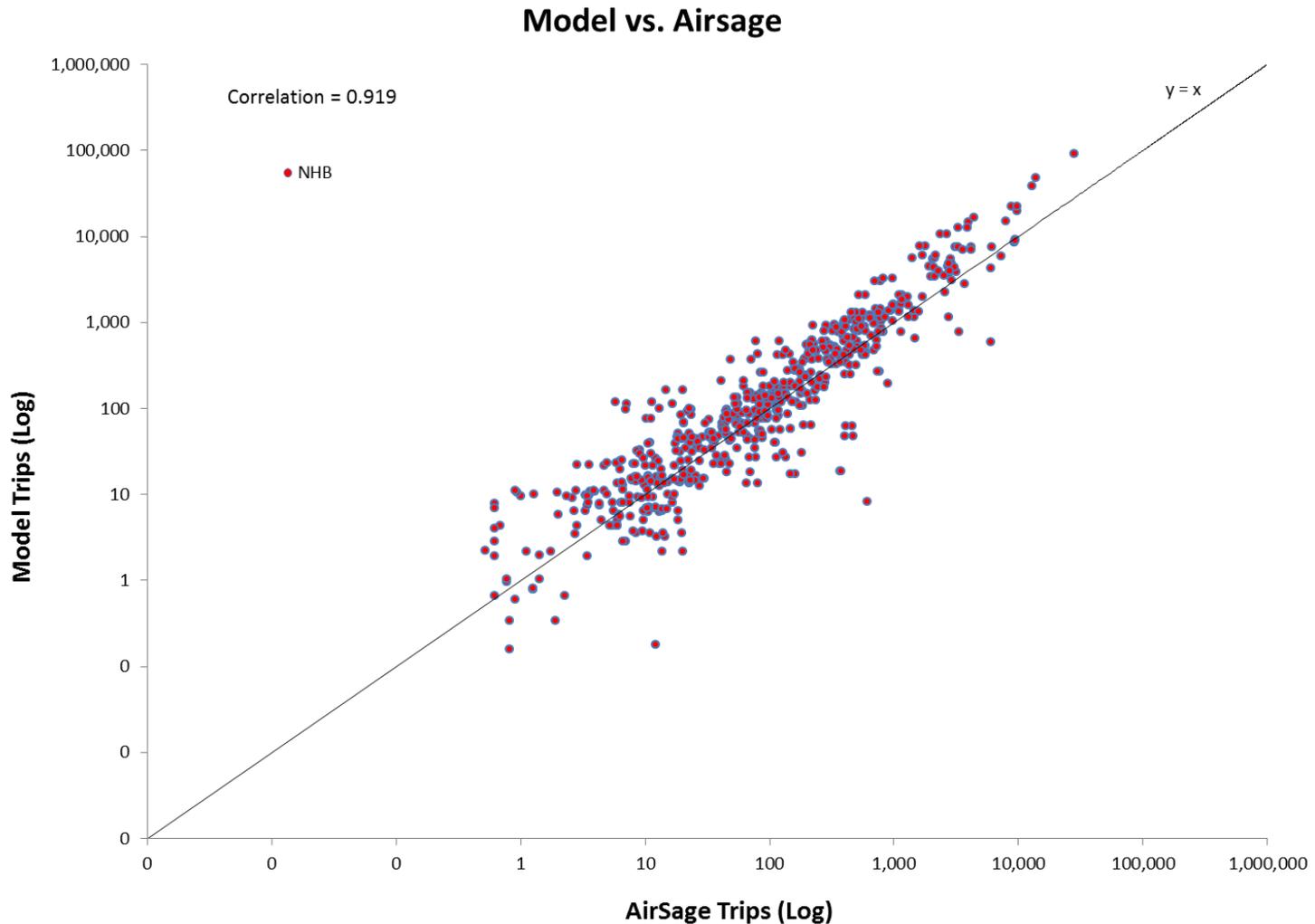
Model vs. Airsage

Talking Points

- Aggregation Levels

- Trip purposes

- Special zones of interest

- Size of external zones

- E-E trips

- Time of data collection

- Time of day partitions

- Trip length frequency distribution

- Select link / zone analysis

- Average trip length

Size of External Zones / E-E trips

- AirSage suggests that 30 to 45 minute travel time buffer is created around study area to form the external zones

- Small external zones will miss a lot of E-E, I-E and E-I trips.

- Large external zones will have add a lot of E-E trips, especially if the external zone contains mid or large size cities.

# Talking Points

- Aggregation Levels

- Trip purposes

- Special zones of interest

- Size of external zones

- E-E trips

- Time of data collection

- Time of day partitions

- Trip length frequency distribution

- Select link / zone analysis

- Average trip length

Time of Data Collection (Rapid City Model)
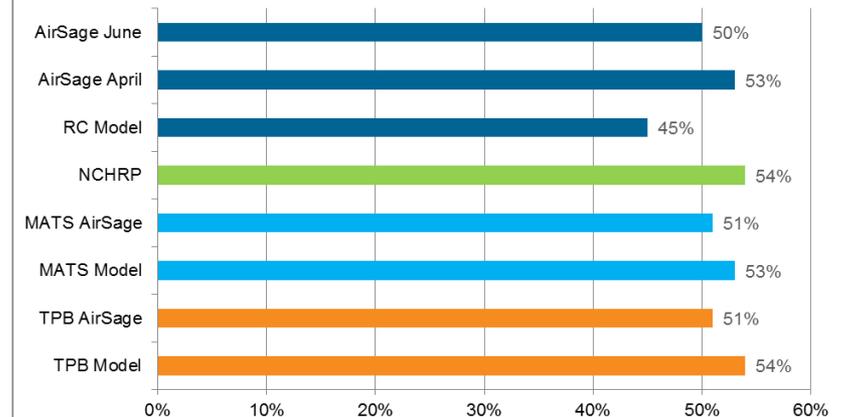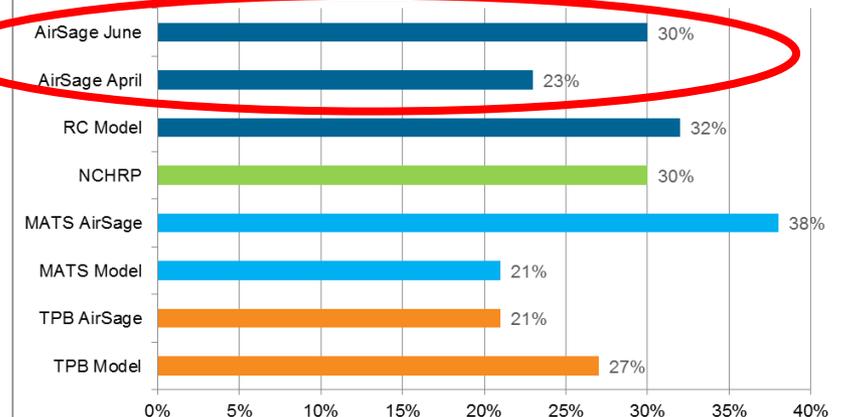
Talking Points

- Aggregation Levels

- Trip purposes

- Special zones of interest

- Size of external zones

- E-E trips

- Time of data collection

- Time of day partitions

- Trip length frequency distribution

- Select link / zone analysis

- Average trip length

Time of Day Distributions

AM period = 6am – 10am (*4hrs*)
MD period = 10am – 3pm (*5hrs*)
PM period = 3pm – 7pm (*4hrs*)
NT period = 7pm – 6am (*11hrs*)
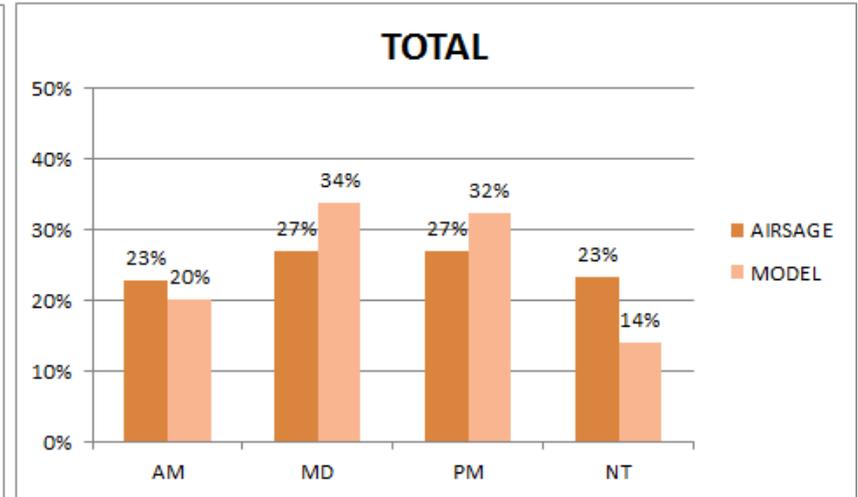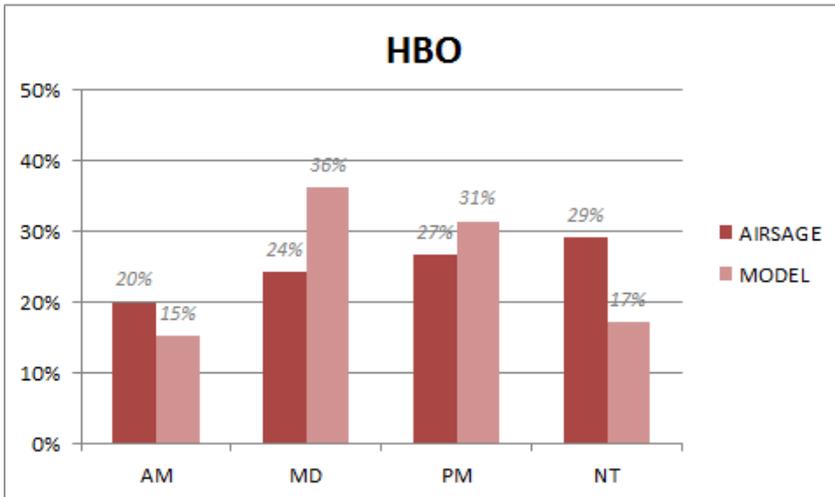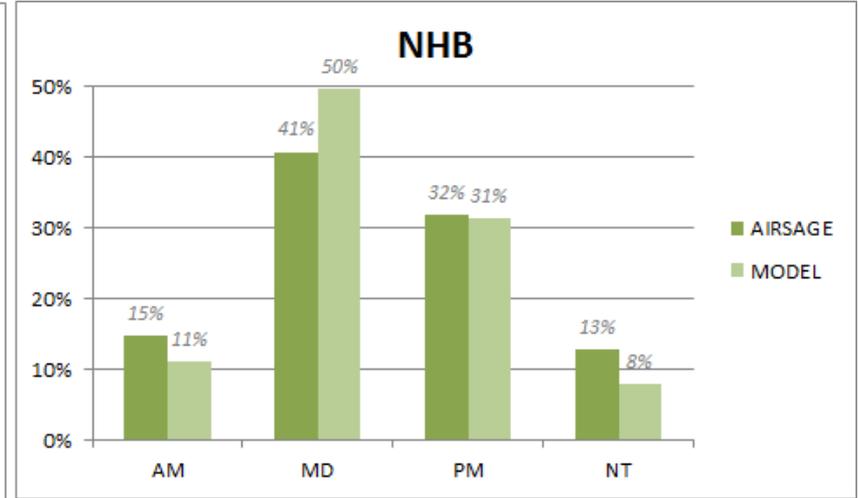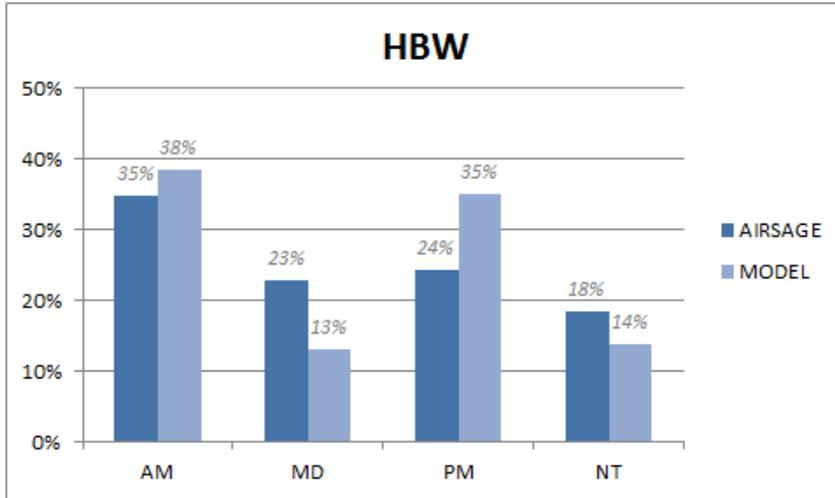
Talking Points

- Aggregation Levels

- Trip purposes

- Special zones of interest

- Size of external zones

- E-E trips

- Time of data collection

- Time of day partitions

- Trip length frequency distribution

- Select link / zone analysis

- Average trip length

Trip Length Frequency Distribution
Total Trips (2-mile bins)

AirSage and Model trip length distributions for ALL trips are consistent and compare favorably
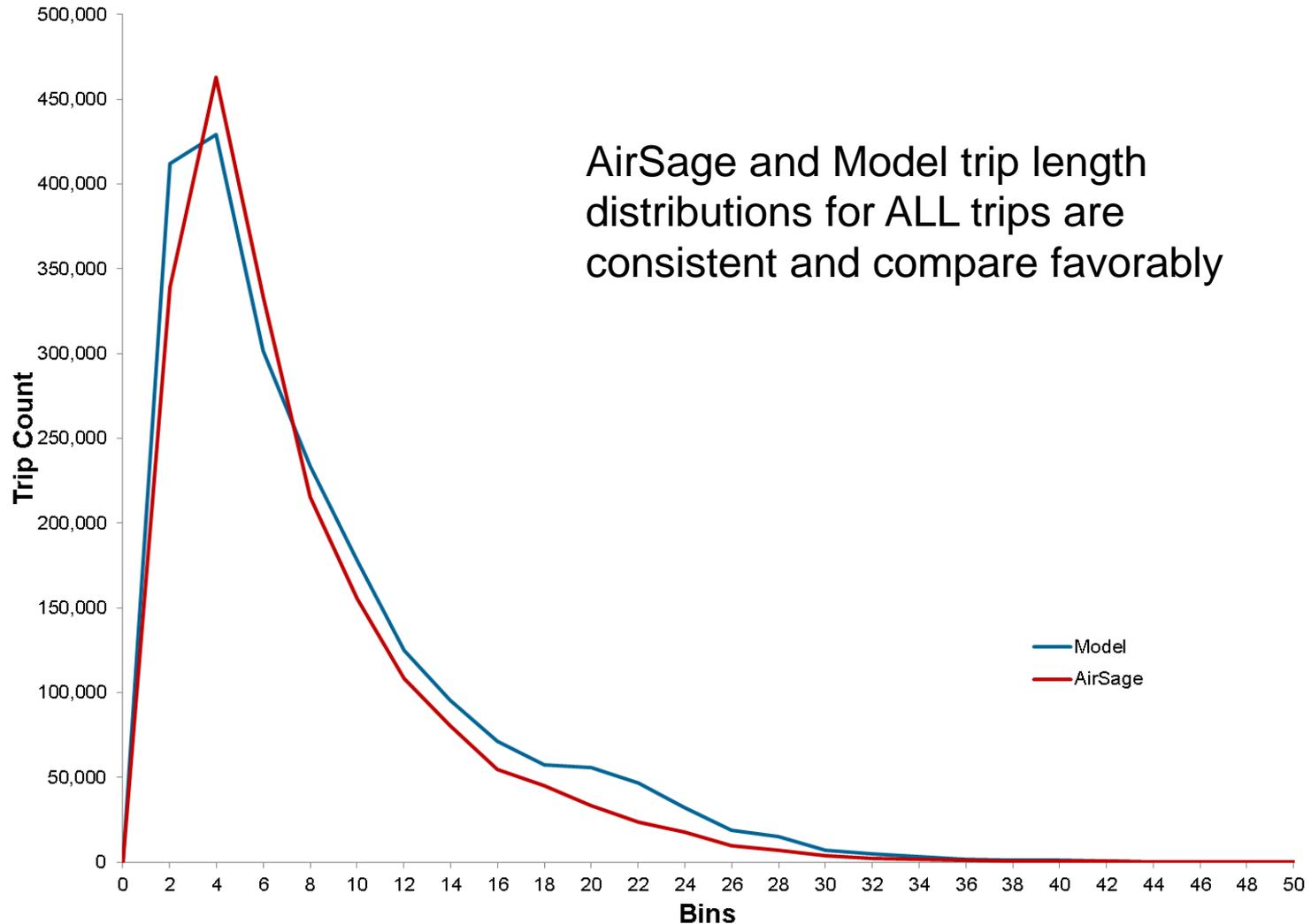
Trip Length Frequency Distribution HBW w LEHD (2-mile bins)
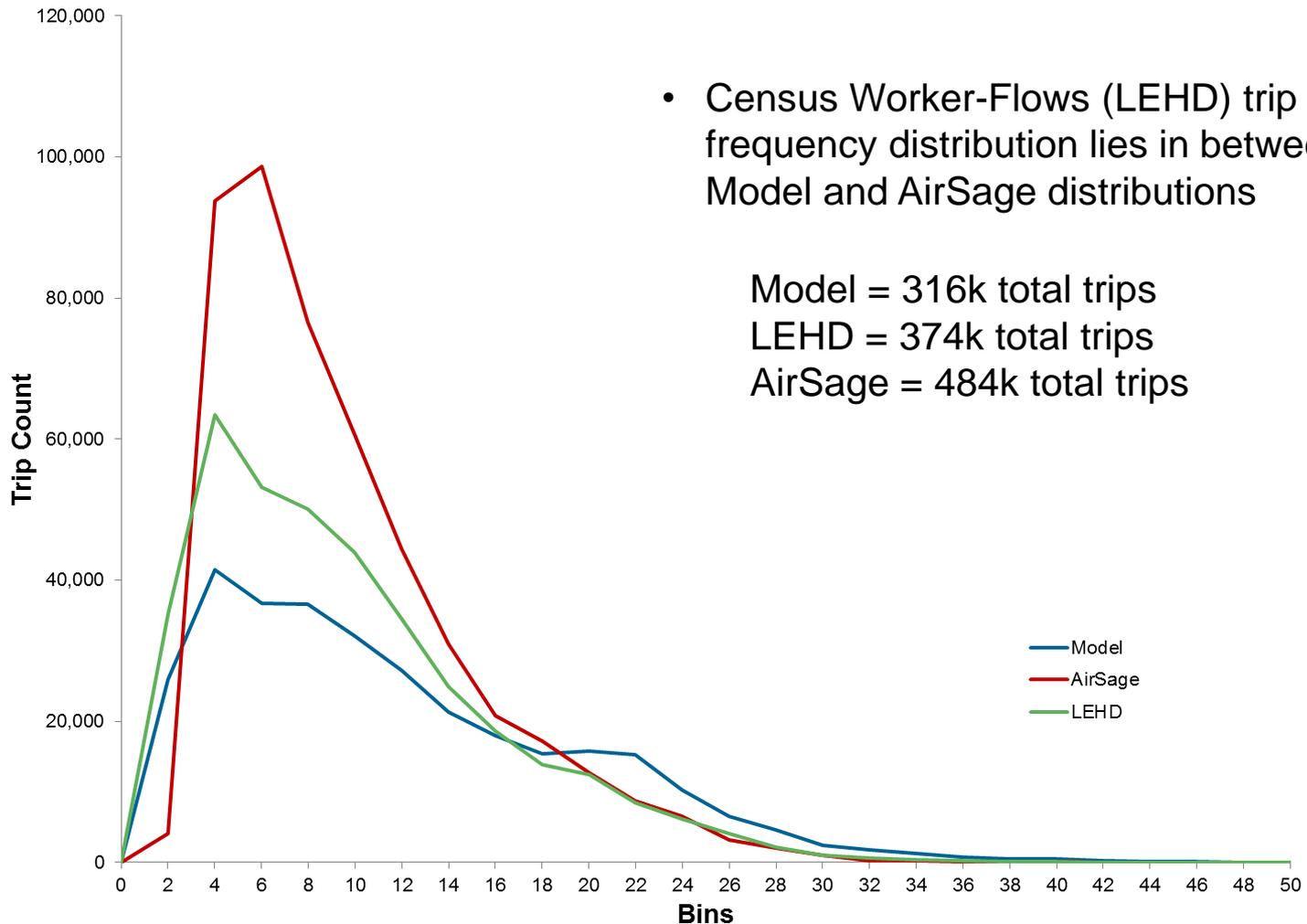
- Census Worker-Flows (LEHD) trip length frequency distribution lies in between the Model and AirSage distributions

  Model = 316k total trips
  LEHD = 374k total trips
  AirSage = 484k total trips

Trip Length Frequency Distribution NHB Trips (2-mile bins)

AirSage has <u>fewer</u> short non-home based trips which are less than 6-miles in length
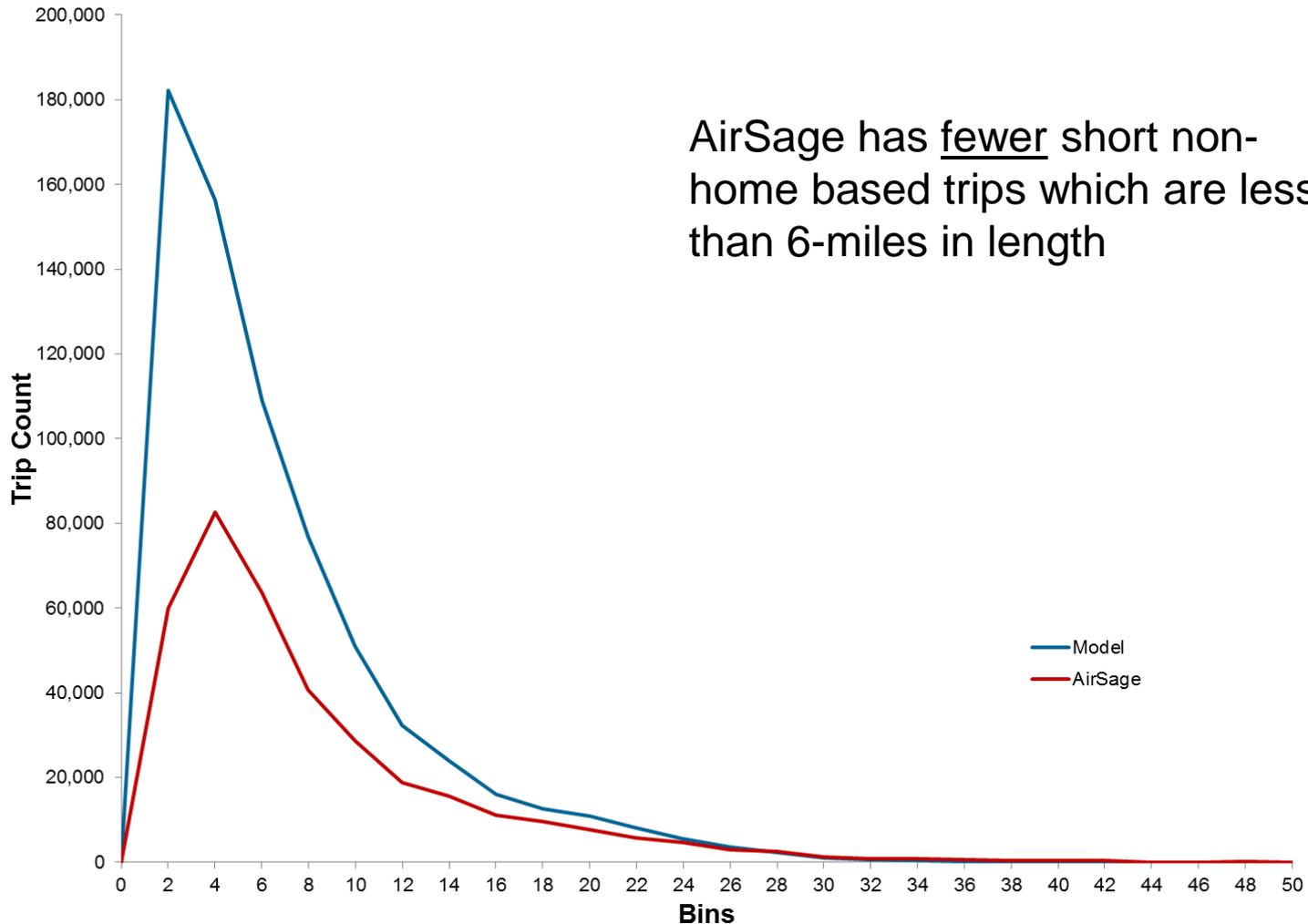
Trip Length Frequency Distribution HBW + NHB Trips (2-mile bins)
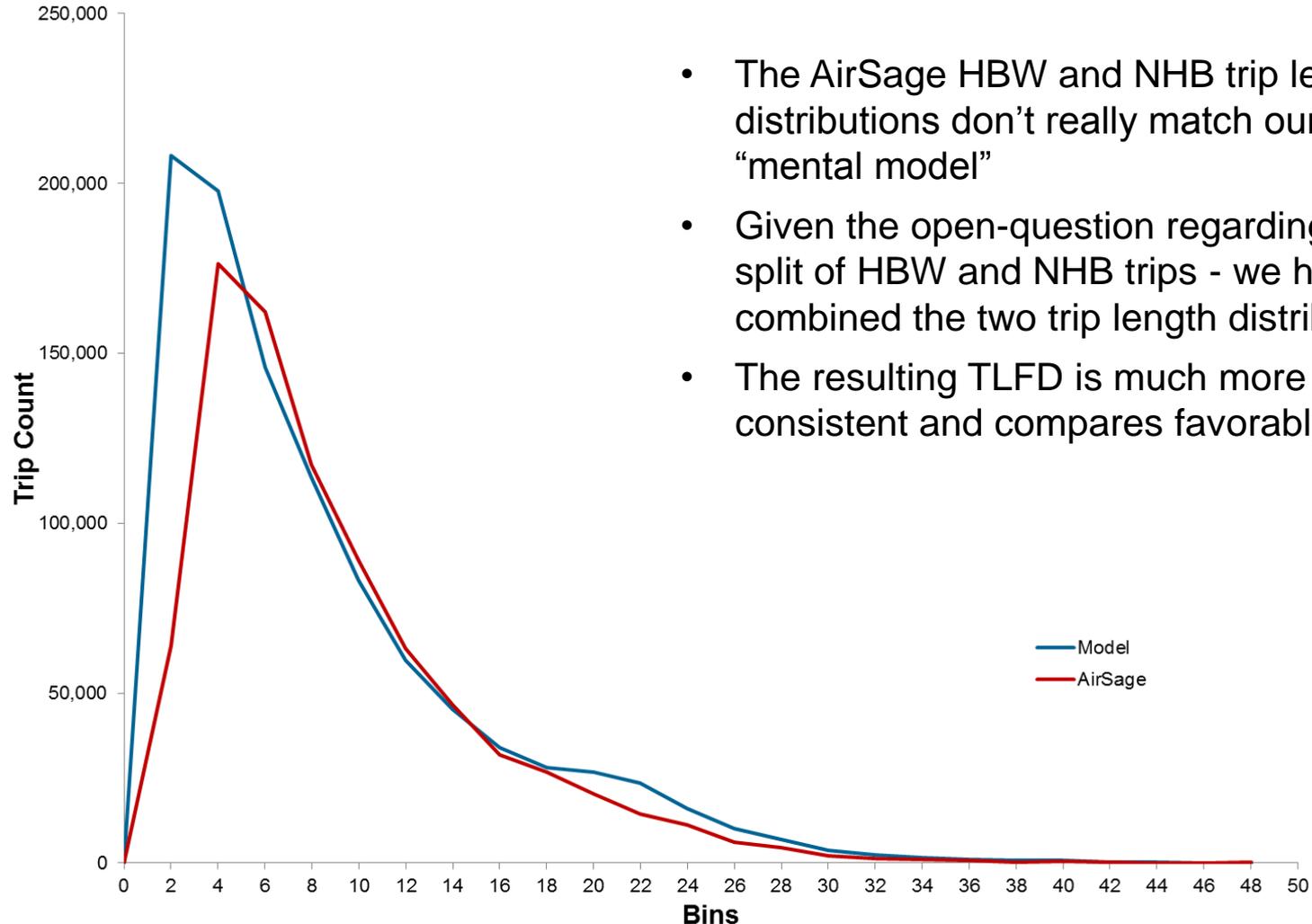
- The AirSage HBW and NHB trip length distributions don't really match our "mental model"

- Given the open-question regarding the % split of HBW and NHB trips - we have combined the two trip length distributions

- The resulting TLFD is much more consistent and compares favorably

Talking Points

- Aggregation Levels

- Trip purposes

- Special zones of interest

- Size of external zones

- E-E trips

- Time of data collection

- Time of day partitions

- Trip length frequency distribution

- Select link / zone analysis

- Average trip length

Select Link Analysis
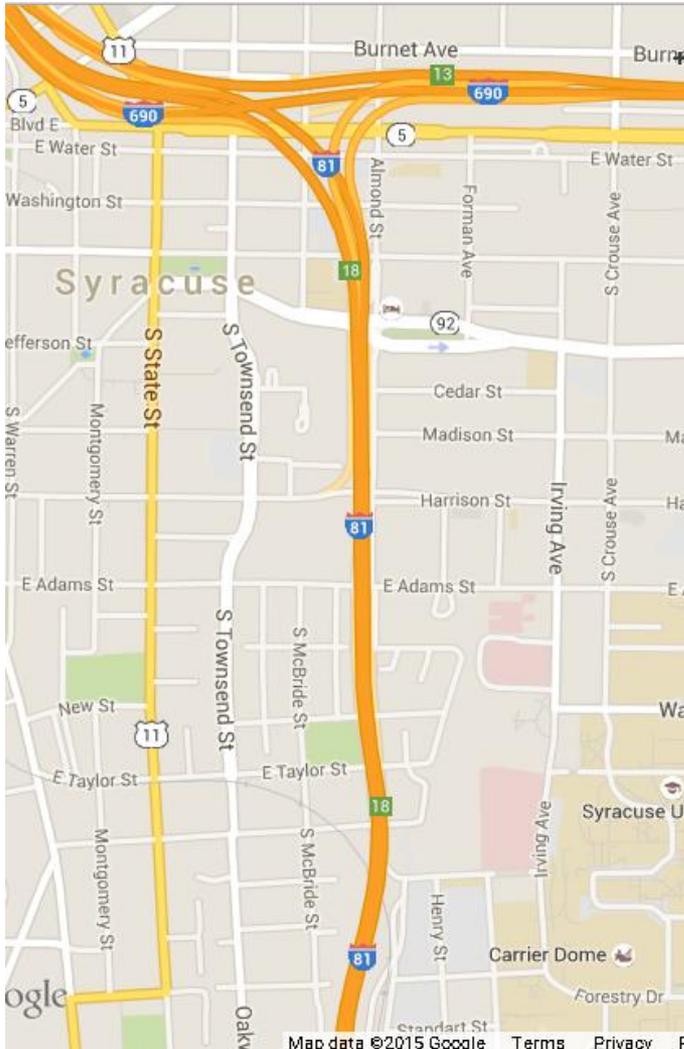
For Select Link analysis using the AirSage data, a ONE mile long section along I-81 just south of Route 92/Genesse Street was selected.

Since AirSage cannot differentiate between vehicles travelling Northbound or Southbound, the SMTC model was run to collect origin and destination zones of vehicles travelling along this section in either direction.

Also since AirSage only collects **sample of devices** travelling across the section, only a comparison of percentages between AirSage and the Model results is possible.

Select Link Analysis .. contd
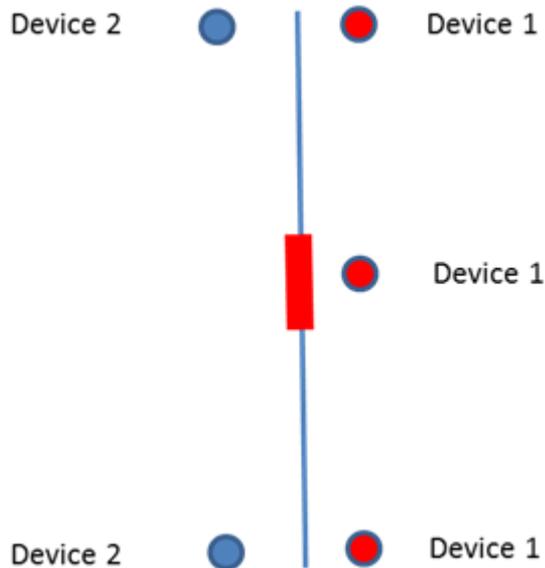
It is important to note that per AirSage, a transient point is registered when there is a device session (call/text/data) and/or when the device changes cell towers.

AirSage only reports the devices that are seen on the select link i.e. devices that ping while they are moving.

In this figure, **device 1 will be** reported and **device 2 will not be** reported (hence it's only observed sample and not the traffic count).

Select Link Analysis .. Town Level Comparison

Model vs. Airsage
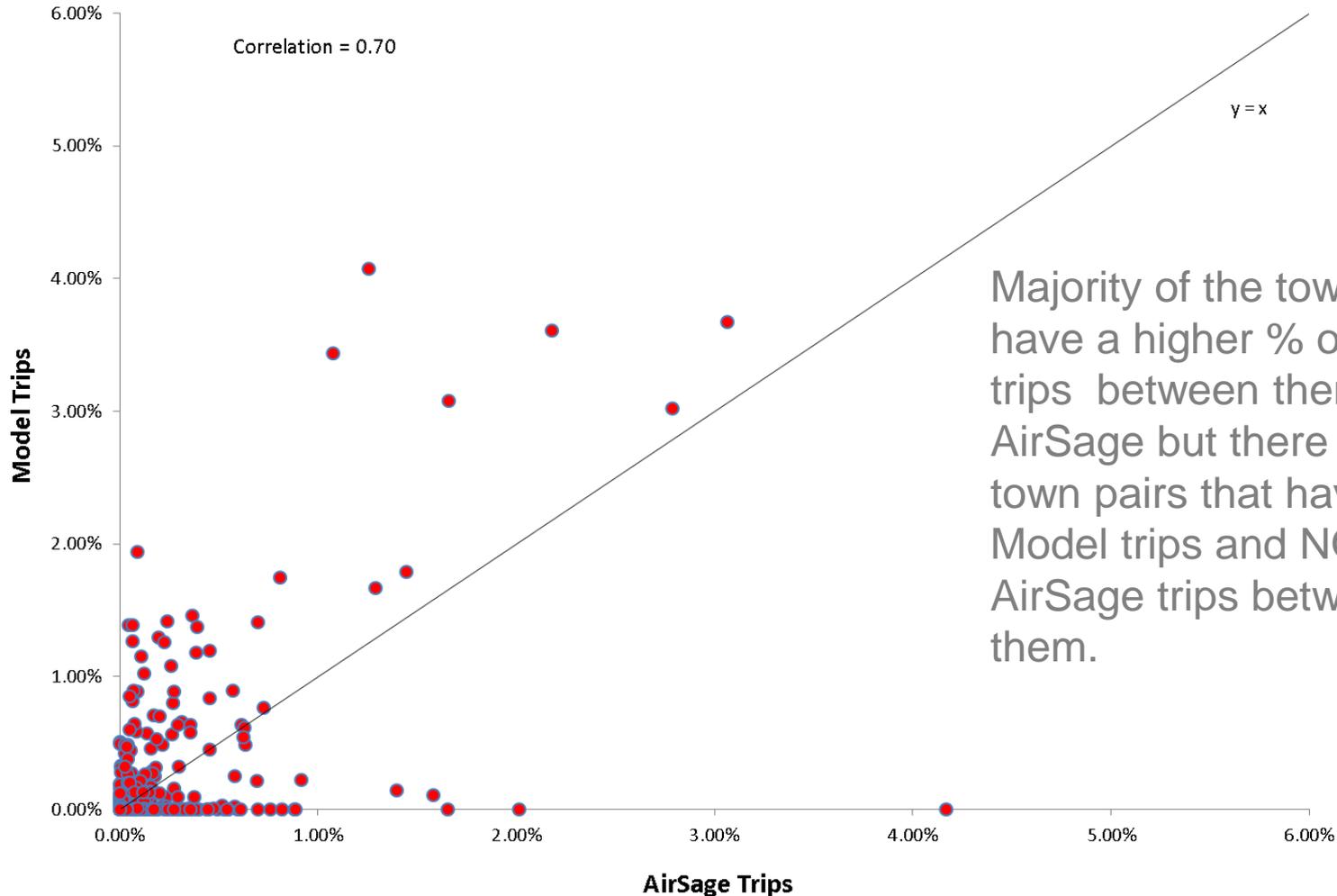
Correlation = 0.70

y = x

Model Trips

AirSage Trips

Majority of the town pairs have a higher % of Model trips between them than AirSage but there are a few town pairs that have ZERO Model trips and NONZERO AirSage trips between them.

Talking Points

- Aggregation Levels
- Trip purposes
- Special zones of interest
- Size of external zones
- E-E trips
- Time of data collection
- Time of day partitions
- Trip length frequency distribution
- Select link / zone analysis
- Average trip length

Average Trip Distances (Miles)

Purpose	Model (all)	Model (I-I only)	AirSage (all)	AirSage (I-I only)
HBW	10.9	9.2	8.7	8.1
LEHD worker-flows	8.9	8.3	*n/a*	*n/a*
HBO	7.6	6.2	5.6	6.8
NHB	6.0	5.5	7.4	7.9
TOTAL	**7.6**	**6.4**	**6.4**	**7.4**

Note:

AirSage data does not include trip distance. Model trip distances (*skims*) are appended to the AirSage OD data to derive averages and trip length distribution frequencies.

NHTS - Travel Day Purpose of Trip Work = 10.2 miles

Key Takeaways: Travel Model vs. AirSage

1 Aggregation is good; disaggregation is bad

- Agencies should think about the level of lowest level of resolution that they can be happy with and develop zones accordingly. Will be economical too.

2 Get creative with external zone boundaries

- Don't make them too small so you don't miss out on external trips.
- Don't make them too large so that you include mid to large cities that you don't care about.

3 Think about what trip purposes you really need and why

- Home and Work based trip purposes should be pretty good unless there are significant number of students or shift workers in your region

4 Select link analysis should only be done on long links and with care

Thank You!

www.rsginc.com

Sumit Bindra
Sumit.Bindra@rsginc.com