Using Big Data for Statewide Models: Learnings from Colorado

Joanathan Lim
StreetLight Data
joan.lim@streetlightdata.com
Outline

I. Big Data: What Is It?
   I. Types of Location Data
   II. Technical Characteristics

II. Colorado’s Statewide Model

III. Q&A
Big Data: What Is It?
Big Data is Just a Buzz Word...But Location Records Created by Mobile Devices are Valuable for Modeling

LBS Data Location
Circle radii vary: they accurately reflect the spatial precision of each unique data point

Navigation-GPS Location
Circle enlarged for visibility

Note: This image shows a filtered subset of data to improve visibility
Key Advantages of Big Data for Statewide Models

1. Comprehensive, Current, and **Empirical**
2. Study Travel Patterns in Very **Large** Areas
3. Study Travel Patterns in Very **Small** Areas
There Are Three Main Types of Location Data for Travel Demand Modeling Projects

- Navigation-GPS
- Location-Based Services
- Cellular Tower
Items to Consider When Evaluating Big Data Resources

1. Spatial Precision
2. Sample Size and Bias
3. Type of Trip
4. Pinging Frequency
The Impact of Spatial Precision is Significant
Sample Size Matters. How Do You Measure It?
Putting it All Together: Data Selection Should Depend on Your Priorities

- Spatial Precision
- Sample Size and Bias
- Trip Type
- Pinging Frequency

Combine Nav-GPS and LBS Data Sources for AADT

Cellular Data
Data Processing Engines Can Turn this Messy Data into Useful Metrics for Travel Demand Modeling

**Input:**
- **Big Data**
  - Anonymous and accurate Locational Big Data
  - Road network, land use, parcel, census and more Contextual Data

**Processing:**
- **RouteScience®**
  - Clean
  - Patternize
  - Contextualize
  - Aggregate

**Output:**
- **StreetLight InSight Metrics**
  - **Basic Metrics:**
    - Origin-Destination, Select Link, Zone Activity
  - **Premium Metrics:**
    - AADT, Trip Purpose, Demographics, Commercial Tours, Home/Work Analysis
  - **Customization:**
    - Day Parts, Day Types, Data Period

---

*Proprietary and Confidential*
Results in Colorado

Colorado DOT Used StreetLight Data’s Analytics to Build its Statewide Travel Demand Model
What Data was Important to Colorado DOT?

1. Regional Flows within Colorado and with neighboring states, deep dive into the rural area

2. Understanding traffic flow during various season, especially during winter season

3. I-25 corridor flow during peak hour
Regional Flows in Colorado and Neighboring States

- Origin-Destination Matrices
  - Full year
  - Summer vs. Winter
- All Colorado TAZs
- External Zones – neighboring states
Regional Trips Flows in Colorado – Full Year of Travel, All-Day, Average Day

One example output is this heat map. It shows the relative volume of trips originating in a CO TAZ or neighboring state that end in central Denver.
Comparing Origins of Weekend Trips to Aspen-Snowmass TAZ in Winter vs. Summer

Winter 2016 (All-Day Average)

The heat map above compares origin locations by relative volume of trips. The pie chart compares the volume of trips that begin in the Aspen/Snowmass TAZ to trips that begin elsewhere.

Summer 2016 (All-Day Average)

The heat map above compares origin locations by relative volume of trips. The pie chart compares the volume of trips that begin in the Aspen/Snowmass TAZ to trips that begin elsewhere.
Key Benefits of GPS Data for Regional Flows

- Summer vs. Winter
- Temporal Precision (Weekends vs. Weekdays, Day Parts)
- TAZ-Level Spatial Precision
- “True” Origins of External Trips
An Additional Modeling Application: Behavior on I-25 Corridor in Denver, Average Weekday, Peak AM

Relative volume of trips beginning in each Denver region ZIP, passing through I-25 at 84th St SB, and ending at Denver Tech Center ZIP.

Relative volume of trips beginning in each Denver region ZIP, passing through I-25 at 84th St SB, and ending north of Tech Center.
Key Benefits of GPS Data for Modeling Corridor Flows

- Corridor-Level Spatial Precision
- Temporal Precision (Weekends vs. Weekdays, Day Parts)
- “True” Origins and Destinations of Corridor Users
- Commercial Truck Trips
Questions?

Thank You!
Joanathan Lim
StreetLight Data
joan.lim@streetlightdata.com
Different Types of “Big” Geospatial Data Offer Different Benefits for Modelers

<table>
<thead>
<tr>
<th></th>
<th>Cellular</th>
<th>Navigation-Based GPS</th>
<th>Location-Based Services</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial Precision</strong></td>
<td>200-1000 meters</td>
<td>5 meters</td>
<td>5 meters – 50 meters</td>
</tr>
<tr>
<td><strong>Frequency of Data Pings</strong></td>
<td>Irregularly; every 15 min – hours</td>
<td>Regularly; every 1 sec – 1 min</td>
<td>Variable; sometimes triggered by location change</td>
</tr>
<tr>
<td><strong>Type of Trip</strong></td>
<td>Blends personal and commercial trips</td>
<td>Differentiates personal and commercial trips</td>
<td>Blends personal and commercial trips (for now)</td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td>Varies by telco: ~10% of population for small telcos and ~25-30% for large telcos</td>
<td>Varies by region; ~1-4% personal trips; ~10-12% of commercial trips (for INRIX)</td>
<td>Medium – 30M+ US devices in our database (&gt;10% of US adult population)</td>
</tr>
</tbody>
</table>
Bias Matters. How Do You Account for It?
This Is What You Get With 1000 meters