Estimating Transit Ridership Patterns Through Automated Data Collection Technology
A Case Study in San Luis Obispo, CA

Ashley Kim
ITE Western District Annual Meeting
San Diego, CA
June 20, 2017
Overview

• Objective
• Literature Review
• Research Design
• Data Filtering
• Results
• Conclusions
Objective
Objectives

• Evaluate the reliability of data collected with emerging automated data collection technologies
  – BlueMAC and Automatic Passenger Counter (APC)
• Estimate origin-destination matrices based on the data
• Analyze ridership patterns in a tactical urbanism setting: Downtown San Luis Obispo Farmer’s Market
Why assess origin-destination flow patterns through automated technologies?

- Transit route planning & stop location placement
- Identify travel patterns to provide conclusions for future planning, operational analysis, & service management
- Current data collection methods are labor-intensive and costly
  - Ex. On-board surveys, smart card transactions, video recording
- Automated technologies can provide continuous data at a low cost-per-datum rate
Literature Review
Current data collection techniques

• Automatic passenger counter (APC)
  – SLO Transit

• Automated fare collection system (AFC)
  – Clipper Card for Bay Area, CA
  – MetroCard for New York City, NY

• Onboard survey
  – LYNX Light Rail Blue Line Extension by the Charlotte Area Transit System
Bluetooth data collection has been used for various transportation engineering applications

- Mass movements
  - Evacuation procedures
  - Tourism patterns
  - Construction effects
- Movements through airport security
- Vehicular travel time
  - Los Osos Valley Road in San Luis Obispo, CA
Bluetooth devices use radio waves to wirelessly connect to a phone or computer.

Source: Libelium
Tactical Urbanism

• Low-cost, temporary changes to the built environment.
• PARK(ing) day
  – On-street parking to public space
  – Awareness towards allotted space for private vehicle storage
• Downtown San Luis Obispo Farmer’s Market
  – Thursdays, 6-9PM
  – Street/structure parking maximized
  – Higher transit demand
Research Design
Study Area

• City of San Luis Obispo
  – 2009-2013 Population: 58,684 including Cal Poly students
  – 99% of residents live within 0.25 mile from transit stop

• SLO Transit
  – Public bus service to City and Cal Poly San Luis Obispo
  – 7 weekday routes
  – 6 Saturday routes
  – 4 Sunday routes
Route 4/5

Route 6A

Route 6B

Route 1

Route 3

Route 2

Source: OTvia
Four data collection methods were utilized for this study.

- **Bluetooth Data**
  - 5 BlueMAC devices
  - February-March 2017
  - Detection time stamp and MAC ID

- **GPS Probe**
  - 2 Probe Days, 3 Probe Runs
  - March 21 and March 22, 2017
  - Latitude, longitude, elevation, time

- **Automatic Passenger Counter (APC)**
  - Bishop Peak Technology
  - Historic GPS data of buses

- **Passenger Survey**
  - 100 passengers @ Kennedy Library Bus stop
  - Trip frequency, wait times
  - Bluetooth enabled? Why?
  - 27 out of 100 passengers have Bluetooth enabled during bus trips
Bluetooth detection specifications

- Unique match media access control (MAC) ID per device
- Low-energy USB plug-ins for 30 ft detection radius
- Class I Bluetooth

Source: Digiwest
BlueMAC Data

- February 28 - March 30, 2017
  - 30 total days
- Five BlueMAC Devices
  - Each deployed on a SLO Transit bus
  - Daily Dispatch Logs to match routes

<table>
<thead>
<tr>
<th>BlueMAC ID</th>
<th>Bus ID Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-01</td>
<td>858</td>
</tr>
<tr>
<td>CP-02</td>
<td>860</td>
</tr>
<tr>
<td>CP-05</td>
<td>862</td>
</tr>
<tr>
<td>CP-04</td>
<td>1264</td>
</tr>
<tr>
<td>DIGI-150</td>
<td>1365</td>
</tr>
</tbody>
</table>
Bluetooth Sensitivity Check

• Do passengers stay within detection range during their trip?
• What if an adjacent vehicle on the road is detected?
• What if a bicyclist, pedestrian or non-passenger is detected?
A Bluetooth sensitivity check was conducted on April 28, 2017 on Route 4.
Data Filtering
The data was binned into projects by week and route, then downloaded as CSV files.

<table>
<thead>
<tr>
<th>Route 2 Week 1</th>
<th>Route 2 Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start to End Date/Time:</strong></td>
<td><strong>Start to End Date/Time:</strong></td>
</tr>
<tr>
<td>2/28/2017 12:00 AM to 3/2/2017 11:59 PM (UTC-8)</td>
<td>3/7/2017 12:00 AM to 3/8/2017 11:59 PM (UTC-8)</td>
</tr>
<tr>
<td><strong>Active Locations:</strong></td>
<td><strong>Active Locations:</strong></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Route 2 Week 3</th>
<th>Route 3 Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start to End Date/Time:</strong></td>
<td><strong>Start to End Date/Time:</strong></td>
</tr>
<tr>
<td>3/14/2017 12:00 AM to 3/16/2017 11:59 PM (UTC-7)</td>
<td>3/7/2017 12:00 AM to 3/9/2017 11:59 PM (UTC-8)</td>
</tr>
<tr>
<td><strong>Active Locations:</strong></td>
<td><strong>Active Locations:</strong></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Route 4A Week 1</th>
<th>Route 4A Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start to End Date/Time:</strong></td>
<td><strong>Start to End Date/Time:</strong></td>
</tr>
<tr>
<td>2/28/2017 12:00 AM to 3/1/2017 11:59 PM (UTC-8)</td>
<td>3/21/2017 12:00 AM to 3/22/2017 11:59 PM (UTC-7)</td>
</tr>
<tr>
<td><strong>Active Locations:</strong></td>
<td><strong>Active Locations:</strong></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
The BlueMAC website displayed the Capture Rate of each device: the number of devices detected per hour.
APC and Unfiltered BlueMAC Data

Raw Data from Route 2 on Tuesday, February 28, 2017

Hour of Day

APC Counts & Bluetooth devices detected

0 50 100 150 200 250 300 350 400 450 500

APC
Bluetooth
Compared to the APC passenger counts, there were significantly more Bluetooth devices detected due to noise and inconsistencies.

<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
<th>Route</th>
<th>Bluetooth</th>
<th>APC</th>
<th>Percent Difference (APC – BT)/APC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday</td>
<td>2/28/2017</td>
<td>2</td>
<td>4,244</td>
<td>282</td>
<td>-1404.96%</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3/1/2017</td>
<td>2</td>
<td>4,169</td>
<td>291</td>
<td>-1332.65%</td>
</tr>
<tr>
<td>Thursday</td>
<td>3/2/2017</td>
<td>2</td>
<td>4,317</td>
<td>280</td>
<td>-1441.79%</td>
</tr>
<tr>
<td>Tuesday</td>
<td>3/7/2017</td>
<td>3</td>
<td>1,613</td>
<td>426</td>
<td>-278.64%</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3/8/2017</td>
<td>3</td>
<td>1,580</td>
<td>417</td>
<td>-278.90%</td>
</tr>
<tr>
<td>Thursday</td>
<td>3/9/2017</td>
<td>3</td>
<td>1,712</td>
<td>390</td>
<td>-338.97%</td>
</tr>
<tr>
<td>Tuesday</td>
<td>3/14/2017</td>
<td>6A</td>
<td>5,521</td>
<td>970</td>
<td>-469.18%</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3/15/2017</td>
<td>6A</td>
<td>4,836</td>
<td>976</td>
<td>-395.49%</td>
</tr>
<tr>
<td>Thursday</td>
<td>3/16/2017</td>
<td>6A</td>
<td>5,283</td>
<td>921</td>
<td>-473.62%</td>
</tr>
</tbody>
</table>

Average Percent Difference: -712.69%
Raw BlueMAC Data Set

Filter 1
MAC ID ≠ 6 units

Filter 2
Bus service hours

Filter 3
Number of observations < 6 times

Filter 4
Grouping MAC IDs into multiple trips

Filter 5
3 Minutes < Detection Duration < Route Duration

Final Cleaned Data Set
After filtering the data, the weekly total detected devices were plotted with the raw data.
APC and BlueMAC hourly counts for Tuesday, Wednesday, and Thursday were plotted for Routes 3 and 6A.
For Route 2, hourly Bluetooth device counts were higher than APC hourly counts.
The data filter for trip duration was adjusted on statistical analysis software to different trip durations for Route 2.
Overall, Route 6A had the most consistent percent differences, and was chosen for the OD estimation.

<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
<th>Route</th>
<th>Bluetooth Raw Detection</th>
<th>Bluetooth Filtered Detection</th>
<th>APC</th>
<th>% Difference (APC-BT)/APC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday</td>
<td>2/28/2017</td>
<td>2</td>
<td>4,244</td>
<td>512</td>
<td>282</td>
<td>-81.56%</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3/1/2017</td>
<td>2</td>
<td>4,169</td>
<td>528</td>
<td>291</td>
<td>-81.44%</td>
</tr>
<tr>
<td>Thursday</td>
<td>3/2/2017</td>
<td>2</td>
<td>4,317</td>
<td>654</td>
<td>280</td>
<td>-133.57%</td>
</tr>
<tr>
<td>Tuesday</td>
<td>3/7/2017</td>
<td>3</td>
<td>1,481</td>
<td>200</td>
<td>407</td>
<td>50.86%</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3/8/2017</td>
<td>3</td>
<td>1,580</td>
<td>419</td>
<td>417</td>
<td>-0.48%</td>
</tr>
<tr>
<td>Thursday</td>
<td>3/9/2017</td>
<td>3</td>
<td>1,551</td>
<td>270</td>
<td>386</td>
<td>30.05%</td>
</tr>
<tr>
<td><strong>Tuesday</strong></td>
<td><strong>3/14/2017</strong></td>
<td><strong>6A</strong></td>
<td><strong>5,521</strong></td>
<td><strong>678</strong></td>
<td><strong>970</strong></td>
<td><strong>30.10%</strong></td>
</tr>
<tr>
<td><strong>Wednesday</strong></td>
<td><strong>3/15/2017</strong></td>
<td><strong>6A</strong></td>
<td><strong>5,577</strong></td>
<td><strong>766</strong></td>
<td><strong>1,083</strong></td>
<td><strong>29.27%</strong></td>
</tr>
<tr>
<td><strong>Thursday</strong></td>
<td><strong>3/16/2017</strong></td>
<td><strong>6A</strong></td>
<td><strong>5,283</strong></td>
<td><strong>678</strong></td>
<td><strong>921</strong></td>
<td><strong>26.38%</strong></td>
</tr>
</tbody>
</table>

Average Percent Difference: -14.49%
Results
Origin-destination matrices were generated for Route 6A.
APC data for Route 6A was used to generate an OD matrix.

<table>
<thead>
<tr>
<th>RouteID</th>
<th>Route</th>
<th>Stop ID</th>
<th>StopName</th>
<th>Observed</th>
<th>CountIn</th>
<th>CountOut</th>
</tr>
</thead>
<tbody>
<tr>
<td>957</td>
<td>Route 6A</td>
<td>63</td>
<td>Highland at Mt. Bishop</td>
<td>3/16/2017 9:06</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>957</td>
<td>Route 6A</td>
<td>63</td>
<td>Highland at Mt. Bishop</td>
<td>3/16/2017 9:06</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>957</td>
<td>Route 6A</td>
<td>63</td>
<td>Highland at Mt. Bishop</td>
<td>3/16/2017 9:06</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>957</td>
<td>Route 6A</td>
<td>63</td>
<td>Highland at Mt. Bishop</td>
<td>3/16/2017 9:06</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

3/16/17 5:10pm and 5:40pm Number of Rides by runs from APC data

<table>
<thead>
<tr>
<th>ROUTE 6A Origin</th>
<th>Cal Poly Kennedy Library</th>
<th>Highland at Mt. Bishop</th>
<th>Highland at Cuesta</th>
<th>Highland at Jeffrey</th>
<th>Highland at Patricia</th>
<th>Patricia at S. Tassajara</th>
<th>Patricia at Palomar</th>
<th>Patricia at Chorro</th>
<th>Casa at Murray</th>
<th>Casa at Deseret (NB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cal Poly Kennedy Library - S</td>
<td>-</td>
<td>0</td>
<td>21</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>13</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Highland at Mt. Bishop</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Highland at Cuesta</td>
<td>8</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Highland at Jeffrey</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Highland at Patricia</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patricia at Foothill</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Foothill at La Entrada</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramona at S. Tassajara</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramona at Palomar</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramona at Chorro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Casa at Murray</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Casa at Deseret (NB)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
APC data was used to compare ridership for Route 6B on Tuesday versus Thursday.
APC data for Route 6B connecting Cal Poly to Downtown SLO was analyzed for Thursday PM periods for Downtown San Luis Obispo Farmer’s Market.

![APC Data for Multiple Tuesdays and Thursdays on Route 6B](chart.png)
An OD matrix was generated using APC data for Route 6B comparing Tuesday and Thursday PM counts.

<table>
<thead>
<tr>
<th>Tues. &amp; Thurs, 3/14 &amp; 3/16 5:00pm</th>
<th>Number of Rides by Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cal Poly</td>
<td>-</td>
</tr>
<tr>
<td>Grand at McCollum</td>
<td>0</td>
</tr>
<tr>
<td>Grand at Wilson</td>
<td>0</td>
</tr>
<tr>
<td>Mill at Park</td>
<td>0</td>
</tr>
<tr>
<td>Mill at Pepper (WB)</td>
<td>0</td>
</tr>
<tr>
<td>Phillips at Pepper (WB)</td>
<td>0</td>
</tr>
<tr>
<td>Mill at Johnson (WB)</td>
<td>0</td>
</tr>
<tr>
<td>Mill at Santa Rosa (WB)</td>
<td>0</td>
</tr>
<tr>
<td>Downtown Transit Center</td>
<td>32, 103</td>
</tr>
<tr>
<td>Mill at Santa Rosa (EB)</td>
<td>6, 7</td>
</tr>
<tr>
<td>Mill at Johnson (EB)</td>
<td>5, 1</td>
</tr>
<tr>
<td>Phillips at Pepper (EB)</td>
<td>1, 2</td>
</tr>
<tr>
<td>Mill at Pepper (EB)</td>
<td>1, 1</td>
</tr>
<tr>
<td>California at Phillips</td>
<td>2, 0</td>
</tr>
<tr>
<td>California at Taft</td>
<td>1, 1</td>
</tr>
</tbody>
</table>
Conclusions
### BLUETOOTH
- Noise and inconsistencies
- Sample size: ~12%
- Probe run check
- Data filtering required
- Trip durations from detection times
- Passenger effort: enable Bluetooth
- Privacy implications from MAC IDs
- $3,200 per device

### APC
- Exact passenger counts at each bus stop
  - Sample size: ~100%
  - Specific passenger routes must be inferred
- Passenger effort: board and exit the bus
- Passenger identities are anonymous
  - $2,000 per bus door

---

**Objective** | **Literature Review** | **Research Design** | **Data Filtering** | **Results** | **Conclusions**
Downtown Farmer’s Market: the APC data showed higher ridership on Thursdays versus Tuesdays
Origin-destination estimation

- Time consuming for both APC and BlueMAC
- Similar trends for APC and BlueMAC
- Future studies
  - Longer data collection period for specified routes
  - Use statistical analysis software to count ODs
  - Group key locations into zones
Further Analysis and Research

• Data collection methodology
  – Optimal hardware settings and placement
  – Second detector on bus
  – Detector at bus stop
  – Data collection time period
  – Explore other methods: Wi-fi
  – Compare other cities’ data

• Data filtering procedures
  – Specific data filter for specific route
  – Test multiple iterations of filters
  – Explore different statistical analysis software

• OD Estimation
  – Pair bus GPS coordinates to BlueMAC time stamps onto GIS Map
Acknowledgements

• Thesis Committee
  – Dr. Anurag Pande, Advisor
  – Dr. Kimberley Mastako, Committee
  – Dr. Cornelius Nuworsoo, Committee

• Cal Poly Civil & Environmental Engineering Department

• City of San Luis Obispo Public Works
  – Transportation Engineering Team
    • Bryan Wheeler
  – SLO Transit & Bishop Peak Technology
    • Dee Lawson
    • Gamaliel Anguiano
    • John Osumi

• Cal Poly Statistics Department
  – Mitchell Collins
  – Professor Rebecca Ottesen

• Digiwest

• Cal Poly ITE friends & 2016-2017 Officer Team

• Transportation Engineering Student Project Area

• Friends & Family
Thank you for listening!

Ashley Kim, EIT
akim56@calpoly.edu